



**ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ  
ΜΑΘΗΜΑΤΙΚΩΝ  
ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ**

**ΤΟΜΕΑΣ ΦΥΣΙΚΗΣ**

**ΑΝΑΛΥΣΗ ΑΛΓΟΡΙΘΜΩΝ ΑΥΤΟΜΑΤΗΣ  
ΚΑΤΑΤΜΗΣΗΣ ΕΙΚΟΝΑΣ ΚΑΙ ΕΦΑΡΜΟΓΕΣ ΤΟΥΣ  
ΣΕ ΑΡΧΑΙΑ ΚΕΙΜΕΝΑ**

**ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ  
Φοιβος Νικολαος Γαλατουλας**

**Επιβλέπων : Κωνσταντίνος Παπαοδυσσεύς  
αναπληρωτής καθηγητής**

**αθηνά, δεκεμβριος 2013**





**ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ  
ΜΑΘΗΜΑΤΙΚΩΝ  
ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ**

**ΤΟΜΕΑΣ ΦΥΣΙΚΗΣ**

**ΑΝΑΛΥΣΗ ΑΛΓΟΡΙΘΜΩΝ ΑΥΤΟΜΑΤΗΣ  
ΚΑΤΑΤΜΗΣΗΣ ΕΙΚΟΝΑΣ ΚΑΙ ΕΦΑΡΜΟΓΕΣ ΤΟΥΣ  
ΣΕ ΑΡΧΑΙΑ ΚΕΙΜΕΝΑ**

**ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ  
Φοιβος Νικολαος Γαλατουλας**

**Επιβλέπων : Κωνσταντίνος Παπαοδυσσεύς  
αναπληρωτης καθηγητης**

**Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 2014.**

.....  
**Κωνσταντίνος Παπαοδυσσεύς Ηλίας Κουκούτσης Κωνσταντίνος  
αναγνωστοπουλος .Αν.  
Καθηγητής Ε.Μ.Π. Επ.Καθηγητής Ε.Μ.Π Αν. Καθηγητής Ε.Μ.Π.**

**αθηνα,δεκεμβριος 2013**

## ΑΡΧΕΣ ΜΕΘΟΔΩΝ ΚΑΤΑΤΜΗΣΗΣ ΕΙΚΟΝΑΣ

### 1. Το αντικείμενο της κατάτμησης εικόνας

Ένα επιστημονικό πεδίο που γνωρίζει ανάπτυξη τα τελευταία χρόνια στον τομέα της Τεχνητής Νοημοσύνης είναι η μηχανική όραση, υπολογιστική όραση ή τεχνητή όραση (computer vision). η υπολογιστική όραση αναδύθηκε μετά το 1980 ως αποτέλεσμα επέκτασης του πεδίου της ψηφιακής επεξεργασίας εικόνας με αλγορίθμους ανάλυσης και κατανόησης εικόνων. Η μηχανική όραση σχετίζεται με τη θεωρία και την τεχνολογία που εμπλέκονται στη σχεδίαση και κατασκευή συστημάτων που λαμβάνουν και αναλύουν δεδομένα από ψηφιακές εικόνες. Τέτοιου είδους δεδομένα μπορούν να είναι φωτογραφίες, βίντεο, όψεις από πολλαπλές κάμερες, πολυδιάστατες εικόνες από ιατρικό σαρωτή κλπ. Στην παρούσα εργασία τα δεδομένα που εισάγουμε προς επεξεργασία είναι χαρακτήρες από ψηφιακές φωτογραφίες αρχαίων κειμένων.

Η κατάτμηση εικόνας, υποπεδίο της υπολογιστικής όρασης, είναι η διαδικασία κατακερματισμού μιας ψηφιακής εικόνας σε πολλαπλά χωρία (σύνολα εικονοστοιχείων τα οποία καλούνται και superpixels). Σκοπός τις περισσότερες φορές είναι τα χωρία να αναπαριστούν ένα χρήσιμο μέρος της εικόνας όπως τα όρια των οικοπέδων μιας αεροφωτογραφίας ή στην περίπτωση μας το περίγραμμα ενός χαρακτήρα από ένα αρχαίο κείμενο. Οι περιοχές της εικόνας που κόβουμε μπορούν να οριστούν και ως σύνολα εικονοστοιχείων που ομαδοποιούνται σε κυκλικά, ελλειπτικά ή πολυγωνικά. Η κατάτμηση δεν προϋποθέτει τα αλληλεπικαλυπτόμενα χωρία να καταλαμβάνουν όλη την εικόνα, δηλαδή μπορούμε να αγνοήσουμε το φόντο της εικόνας.

Με τη χρήση αλγορίθμων κατάτμησης επιδιώκουμε να πετύχουμε δύο στόχους.

Πρωταρχικώς να αποσυνθέσουμε την εικόνα σε χωρία προς περαιτέρω ανάλυση. Σε απλές περιπτώσεις, το περιβάλλον της εικόνας μπορεί εύκολα να αγνοηθεί από τον αλγόριθμο αφήνοντας μόνο τα κομμάτια που μας ενδιαφέρουν προς ανάλυση. Στην

κατηγορία αυτή εμπίπτουν και οι εικόνες των χαρακτήρων από τις περγαμηνές που μελετάμε, αφού το περιβάλλον της περγαμηνής μπορεί να θεωρηθεί ένα ευκόλως διαχωρίσιμο υπόβαθρο. Σε τέτοιες περιπτώσεις έχουμε αξιόπιστη κατάτμηση εν αντιθέσει με σύνθετες περιπτώσεις όπως η εξαγωγή του περιγράμματος ενός οδικού δικτύου σε μια αεροφωτογραφία σε κλίμακα του γκρι.

Δεύτερος στόχος των μεθόδων κατάτμησης είναι να επιτύχουν μια τροποποίηση της αναπαράστασης. Τα εικονοστοιχεία πρέπει να ομαδοποιηθούν σε κάποια δομή ανώτερης τάξης που έχει σημασία για το εκάστοτε πεδίο που μελετά την εικόνα. Φθάνουμε λοιπόν στο ερώτημα: υπάρχει μια μέθοδος κατάτμησης που να εξυπηρετεί όλα τα επιστημονικά πεδία; Η απάντηση είναι προφανώς όχι, η εμπειρία μας διδάσκει πως όταν εφαρμόζουμε τέτοιες μεθόδους θα πρέπει να διαθέτουμε μια ποικιλόμορφη εργαλειοθήκη από την οποία να μπορούμε να ανασύρουμε το κατάλληλο εργαλείο για την κατάλληλη περίπτωση.

## 2.1 Αναγνώριση περιοχών

Πριν ξεκινήσουμε την περιγραφή των μεθόδων θα περιγράψουμε τα χαρακτηριστικά που πρέπει να έχουν οι περιοχές που θα διαχωρίσει ο εκάστοτε αλγόριθμος:

- τα χωρία που προκύπτουν από την κατάτμηση εικόνας θα πρέπει να παρουσιάζουν μια ομοιομορφία ως προς κάποια παράμετρο πχ. το επίπεδο γκρι, το χρώμα κ.α
- Τα εσωτερικά των χωρίων θα πρέπει να είναι απλά χωρίς μικρές νησίδες ή οπές
- παρακείμενες περιοχές θα πρέπει να εμφανίζουν μεγάλες διαφορές στην τιμή της παραμέτρου ως προς την οποία παρουσιάζουν ομοιομορφία
- τα όρια κάθε χωρίου πρέπει να είναι ομαλά, όχι τραχεία, και διακριτά

Τα κριτήρια αυτά σπάνια ικανοποιούνται αφού σε χωρία που παρουσιάζονται ομοιομορφίες υπάρχουν οπές και τραχείες γραμμές. Θεωρώντας κάποιες διαφορές στην τιμή μιας παραμέτρου μπορεί να παρουσιαστεί συγχώνευση παρακείμενων χωρίων. Επίσης ομοιότητες που παρατηρεί ένας άνθρωπος σε χωρία μιας εικόνας μπορεί να μην είναι ξεκάθαρες λόγω των τιμών που παίρνει η μέθοδος κατάτμησης σε μικρή κλίμακα. Για αυτό χρειάζεται προσοχή στο σχεδιασμό των μεθόδων αναφορικά με την κλίμακα που επεξεργάζονται την εικόνα.

### 2.1.1 Μέθοδοι Clustering

Με τον όρο clustering, ομαδοποίηση ή συσταδοποίηση, στην αναγνώριση προτύπων αναφερόμαστε στην διαδικασία διαμερισμού ενός συνόλου στοιχείων σε μικρότερες ομάδες, υποσύνολα, που τα

ονομάζουμε clusters με βάση ένα μοτίβο. Η καταχώρηση αντικειμένων σε ίδια ομάδα μεταφράζεται ως *ομοιότητα* των αντικειμένων αυτών και αντίστροφα (αντικείμενα που ανήκουν σε διαφορετικές ομάδες είναι ανόμοια). Η ομοιότητα ή μη, μεταξύ των αντικειμένων, ουσιαστικά εξαρτάται από το συγκεκριμένο πρόβλημα και τη μορφή των «αντικειμένων».

Η χρήση αλγορίθμων clustering είναι αρκετά διαδεδομένη σε αρκετές μεθόδους κατάτμησης εικόνας. Ακολουθούν μερικά χαρακτηριστικά παραδείγματα.

#### 2.1.1.1 Κλασικοί Αλγόριθμοι Clustering

Στην Ανάλυση Εικόνας τα διανύσματα αναπαριστούν εικονοστοιχεία ή μερικές φορές μικρές γειτονιές σημείων γύρω από τα pixels. Τα διανύσματα αυτά περιλαμβάνουν ως παραμέτρους:

- τιμές έντασης (intensity)
- τιμές RGB και ιδιότητες χρωμάτων που εξάγονται από αυτές
- υπολογισθείσες ιδιότητες
- μετρήσεις υφής (texture)

Οποιαδήποτε παράμετρος ενός εικονοστοιχείου μπορεί να χρησιμοποιηθεί στην ομαδοποίηση των pixel. Μόλις τα pixels ομαδοποιηθούν σε clusters με βάση τιμές του μετρητικού-χώρου απλοποιείται η εύρεση συνδεδεμένων περιοχών μέσω των ετικετών στις τιμές των συνδεδεμένων συνιστωσών.

Έστω  $K$  το πλήθος των clusters  $C_1, C_2, \dots, C_K$ , με μέσους  $m_1, m_2, \dots, m_K$ ,

Ορίζουμε τη μέτρηση σφάλματος με τη μέθοδο των ελαχίστων τετραγώνων ως:

$$D = \sum_{k=1}^K \sum_{x_i \in C_k} (x_i - m_k)^2$$

Ουσιαστικά με τη χρήση της πάνω σχέσης υπολογίζουμε την απόσταση των δεδομένων από τα αντίστοιχα τους clusters. Μια μέθοδος υπολογισμού ελαχίστων τετραγώνων θα μπορούσε να τρέξει όλες τις πιθανές διαμερίσεις για την τιμή του  $K$  και να επιλέγει εκείνη που ελαχιστοποιεί το  $D$ . Η επίλυση αυτού του προβλήματος είναι υπολογιστικά ακατόρθωτη γι'αυτό όλες οι μέθοδοι κατάτμησης αποτελούν προσεγγίσεις αυτού του προβλήματος. Ένα βασικό ζήτημα είναι αν η τιμή του  $K$  είναι γνωστή εκ των προτέρων. Σε πολλούς αλγορίθμους η τιμή του  $K$  εισάγεται σαν παράμετρος από τον χρήστη ενώ σε άλλους αλγορίθμους το βέλτιστο  $K$  επιχειρείται να ευρεθεί σύμφωνα με κάποιο κριτήριο πχ. διατηρώντας τη διασπορά των clusters μικρότερη από μια προκαθορισμένη τιμή.

#### 2.1.1.2 Η μέθοδος K-Means

Ο αλγόριθμος K-Means είναι μια απλή επαναληπτική μέθοδος η οποία διαλέγει αρχικά τυχαία  $k$  περιοχές και στη συνέχεια αποδίδει σε κάθε μία από αυτές τα σημεία της αρχικής περιοχής με βάση ένα κριτήριο.

1. Θέτουμε το βήμα επανάληψης  $ic=1$
2. Επιλέγει τυχαία ένα σύνολο από  $K$  μέσες τιμές  $m_1(1), m_2(1), \dots, m_K(1)$

3. Για κάθε διάνυσμα  $x_i$  υπολόγισε το  $D(x_i, m_{kic})$  για κάθε  $k=1, \dots, K$  και απόδωσε το διάνυσμα  $x_i$  στο cluster  $C_j$  με την πλησιέστερη μέση τιμή.
4. Αύξησε κατά 1 το  $ic$  και υπολόγισε τις νέες τιμές του συνόλου των μέσων τιμών  $m_{1ic}, m_{2(ic)} \dots, m_{K(ic)}$
5. Επανάλαβε τα βήματα 3 και 4 μέχρις ότου  $C_{kic} = C_k(ic+1)$  για κάθε τιμή του  $k$

Ο αλγόριθμος αν και τερματίζει εγγυημένα έχει το μειονέκτημα ότι πιθανόν δεν θα βρεί το ολικό βέλτιστο, με την έννοια της προσέγγισης των ελαχίστων τετραγώνων. Το βήμα 2 μπορεί να τροποποιηθεί ώστε να διαμεριστεί το σύνολο των διανυσμάτων σε  $K$  τυχαία clusters και στη συνέχεια να υπολογίζει τις μέσες τιμές τους. Το βήμα 5 μπορεί να τροποποιηθεί ούτως ώστε να τερματίζεται εφόσον το ποσοστό των διανυσμάτων τα οποία αλλάζουν cluster σε μια δεδομένη επανάληψη είναι μικρό.

### 2.1.2 Ανάπτυξη χωρίων

Πέρα από την διαμέριση της εικόνας σε ομοειδή χωρία μπορούμε να ξεκινήσουμε και αντιστρόφως δηλαδή να χρησιμοποιήσουμε έναν αλγόριθμο ανάπτυξης χωρίου. Τέτοιοι αλγόριθμοι ξεκινάνε από ένα σημείο της εικόνας και επιχειρούν να επεκταθούν έως ότου τα προς σύγκριση εικονοστοιχεία παρουσιάσουν ανομοιογένειες που δεν τους επιτρέπουν να προσχωρήσουν στο χωρίο.

Για να εξακριβωθεί αν ένα εικονοστοιχείο είναι όμοιο ώστε να συμπεριληφθεί στην υπό ανάπτυξη περιοχή χρησιμοποιείται ένας στατιστικός έλεγχος. Ο Haralick πρότεινε την ακόλουθη τεχνική υποθέτοντας πως κάθε περιοχή αποτελείται από συνδεδεμένα pixels που έχουν ίδιο πληθυσμιακό μέσο και διασπορά.

Συγκεκριμένα, ορίζουμε τον μέσο περιοχής  $X$  :

$$X = \frac{1}{N} \sum_{[r,c] \in R} I[r,c]$$

Όπου  $N$  πλήθος γειτονικών εικονοστοιχείων ενός με τόνο γκρι τιμής  $y$  σε μια περιοχή  $R$ .

Επίσης η διασπορά ορίζεται ως:

$$S^2 = \frac{1}{N} \sum_{[r,c] \in R} (I[r,c] - X)^2$$

Υποθέτοντας ότι όλα τα εικονοστοιχεία που ανήκουν στην περιοχή  $R$  και το δοκιμαστικό pixel  $y$  είναι ανεξάρτητα και κατανομημένα με κανονικό τρόπο, λαμβάνουμε τη στατιστική:

$$T = \frac{N-1}{N} \frac{N+1}{2} \frac{y - X}{S^2}$$

Με κατανομή  $T_{n-1}$ . Αν το  $T$  είναι αρκετά μικρό το  $y$  προστίθεται στην περιοχή  $R$  και ο μέσος και η σταθερά αναπροσαρμόζονται με βάση την τιμή  $y$ . Ειδικότερα δίνονται από την αναδρομική σχέση:

$$X_n \leftarrow \frac{N}{N+1} X_0 + \frac{y}{N+1}$$

Και:

$$S_n^2 \leftarrow S_0^2 + \frac{(y - X_n)^2}{N+1} + \frac{N}{N+1} (X_n - X_0)^2$$

Σε περίπτωση που η  $T$  έχει υψηλή τιμή τότε η  $y$  δεν προήλθε από τα στοιχεία της περιοχής  $R$ . Τότε το  $y$  αφού συμπεραίνουμε πως διαφέρει από τα υπόλοιπα pixels δημιουργεί μια δική του περιοχή. Αυστηροποιώντας το κριτήριο που ομαδοποιεί τα στοιχεία προσθέτουμε την απαίτηση όχι μόνο το  $y$  να έχει μια τιμή κοντά

στον μέσο της γειτονικής περιοχής αλλά και ένα εικονοστοιχείο από τη γειτονική περιοχή να παίρνει μια τιμή κοντά σε αυτήν του  $y$ .

Οι παραπάνω μέθοδοι συγκαταλέγονται στις μεθόδους κατάτμησης εικόνας με βάση διαχωριστικές τεχνικές.

### 2.1.3 Ιεραρχικές τεχνικές

Η μεθοδολογία των μεθόδων που έχουν ως βάση ιεραρχικές τεχνικές μπορεί να συνοψιστεί στα εξής σημεία:

- Συγκροτούμε τις οντότητες-εικονοστοιχεία (pixels)
- Ορίζουμε τις αρχικές παραμέτρους των οντοτήτων. Μια καλή επιλογή είναι κάθε οντότητα να αποτελείται από ένα εικονοστοιχείο ώστε αρχικά να καλύπτουμε όλη την εικόνα
- Δίνουμε τα κριτήρια ομοιογένειας και διαφοροποίησης ανάμεσα στις οντότητες
- Δημιουργούμε τον πίνακα συγγένειας  $s(i,j)$  βρίσκοντας το βαθμό συγγένειας κάθε οντότητας με τις άλλες. Κάθε στοιχείο του πίνακα μας δίνει ποσοτικά τη συγγένεια μεταξύ των οντοτήτων ενώ οι δείκτες αντιστοιχούν στο πλήθος των οντοτήτων.
- Ορίζουμε μια τιμή κατωφλίου συσχέτισης στη μήτρα συγγένειας αναλόγως την επιδίωξη μας. Αν θέλουμε να κατατμήσουμε με βάση το χρώμα τότε για να ομαδοποιήσουμε δυο οντότητες θα πρέπει η διαφορά των τιμών χρωματικού περιεχομένου τους να εκφράζεται στο βαθμό ομοιότητας. Τα στοιχεία που έχουν τιμές κάτω από την τιμή κατωφλίου μηδενίζονται.
- Παίρνοντας στοιχεία από τη θεωρία γραφημάτων ορίζουμε τις αρχικές οντότητες ως κόμβους. Συνδέουμε με κλάδους μόνο αν η τιμή στην μήτρα συγγένειας είναι μη μηδενική. Ο κάθε κλάδος σταθμίζεται με βάση την τιμή του βαθμού ομοιότητας των συσχετιζομένων οντοτήτων ανατρέχοντας μέσω των δεικτών στο στοιχείο της μήτρας συγγένειας.
- Όσοι κόμβοι περιλαμβάνονται σε κλάδους που εμφανίζουν μεγάλες τιμές βαθμού συγγένειας ομαδοποιούνται. Εκ νέου θέτουμε ένα κατώφλι διακοπής της διαδικασίας όταν έχουν ομαδοποιηθεί σε οντότητες όλοι οι κόμβοι. Όταν συγχωνεύονται δύο κόμβοι δημιουργείται ένας καινούριος που αντικαθιστά τους προηγούμενους. Η διαδικασία δημιουργίας κλάδων μεταξύ των νέων κόμβων όπως και η στάθμιση τους γίνεται δυναμικά.

Στην περίπτωση αυτόματης κατάτμησης εικόνας παίρνουμε τη χρωματική απόσταση δύο pixels ως μέτρο ομοιότητας, κατασκευάζουμε τον πίνακα συγγένειας και υπολογίζοντας τη νόρμα στον χώρο R,G,B έχουμε τις σχέσεις:

$$d_1 = |R_i - R_j| + |G_i - G_j| + |B_i - B_j|$$

$$d_2 = \sqrt{(R_i - R_j)^2 + (G_i - G_j)^2 + (B_i - B_j)^2}$$

$$d_M = x^T M_x x = [(R_i - R_j), (G_i - G_j), (B_i - B_j)]^T$$



οι τεχνικές για τον υπολογισμό της συγγένειας και η ομοιότητα μεταξύ συνενωμένων οντοτήτων και προηγούμενων σε κάθε βήμα του αλγορίθμου.

Μερικές από τις σημαντικότερες τεχνικές ιεράρχησης είναι:

1. single link method : Κάθε cluster αντιπροσωπεύεται από όλα τα στοιχεία και η ομοιότητα μεταξύ δύο clusters καθορίζεται από την ομοιότητα του πιο κοντινού ζεύγους στοιχείων που ανήκουν σε διαφορετικά clusters.
2. CURE method : Η μέθοδος αυτή είναι ουσιαστικά συνδυασμός των δύο προηγούμενων μεθόδων. Εδώ, ορίζονται αντιπρόσωποι, αλλά περισσότεροι από ένας για κάθε cluster. Η ομοιότητα μετράται μεταξύ των ζευγών των αντιπροσώπων των cluster και το κριτήριο συνένωσης δύο περιοχών εξαρτάται από τον βαθμό ομοιότητας των ζευγών αυτών συνολικά που ανήκουν σε διαφορετικά clusters. Σε κάθε συγχώνευση clusters, δυναμικά ορίζονται νέοι αντιπρόσωποι του νέου cluster που προέκυψε. Ακριβώς επειδή ο αλγόριθμος αυτός αποδίδει πολλούς αντιπροσώπους σε κάθε περιοχή, αλλά και γιατί τους συρρικνώνει προς το κέντρο βάρους, αντιμετωπίζει με μεγαλύτερη επιτυχία καταστάσεις διαφορετικών μεγεθών και σχημάτων clusters. Ο αριθμός των αντιπροσώπων εξαρτάται από το μέγεθος, το σχήμα αλλά και την παρουσία θορύβου σε κάθε cluster.
3. group-average method : Η μέθοδος αυτή υποθέτει διασύνδεση όλων των δεδομένων των clusters και υπολογίζει όλες τις αποστάσεις μεταξύ τους. Κανονικοποιεί τις τιμές σχηματίζοντας ένα  $n \times m$ , πίνακα όπου  $n$  είναι ο αριθμός των δεδομένων στην μία περιοχή και  $m$  ο αντίστοιχος στην δεύτερη, ώστε να αντιμετωπίσει το πρόβλημα που προκύπτει όταν τα clusters έχουν διαφορετικό μέγεθος.
4. ROCK method : Ο αλγόριθμος αυτός μοιάζει με αυτόν της group-average μεθόδου, κανονικοποιεί όμως με βάση κατώφλι ορισμένο από τον χρήστη, αντί να χρησιμοποιεί ένα πίνακα  $n \times m$ .

## 2.2 Αναπαράσταση περιοχών

Κάθε αλγόριθμος κατάτμησης που παράγει κάποιον αριθμό περιοχών εικόνας θα πρέπει να προβλέπει έναν τρόπο να καταχωρεί τις περιοχές για μελλοντική χρήση. Υπάρχουν διάφοροι τρόποι για να επιτευχθεί αυτό όπως η υπερκάλυψη της εικόνας, η προσθήκη ετικετών, encodings στα όρια των περιοχών κ.ά. Ο πιο διαδεδομένος τρόπος είναι η προσθήκη ετικετών.

## 2.3 Αναγνώριση περιγράμματος

Μερικές εφαρμογές ανάλυσης εικόνας δουλεύουν άμεσα με περιοχές ενώ άλλες απαιτούν τα όρια αυτών των περιοχών ή κάποιες άλλες δομές όπως ευθύγραμμα τμήματα ή κυκλικά τόξα. Στην παρούσα ενότητα εξετάζουμε την εξαγωγή τέτοιων δομών από τις εικόνες.

### 2.3.1 Εύρεση ολοκληρωμένων ορίων περιοχών

Αφού ολοκληρωθεί η διαδικασία κατάτμησης μιας εικόνας μπορούμε να εξάγουμε από αυτήν κάποιο περίγραμμα. Σε μικρού μεγέθους εικόνες η υλοποίηση μιας διαδικασίας αφαίρεσης του περιγράμματος είναι απλή. Σαρώνουμε την εικόνα και καταχωρούμε

σε λίστα το πρώτο rixel του περιγράμματος κάθε αντικειμένου. Μετά για κάθε περιοχή ξεκινώντας από το πρώτο rixel του περιγράμματος ακολουθούμε το περίγραμμα με ωρολογιακή φορά μέχρι να ξαναβρεθούμε στο πρώτο rixel, στο σημείο απ' όπου εκκίνησε η διαδικασία. Για μεγάλου μεγέθους εικόνες η διαδικασία αποτυγχάνει διότι υπερχειλίζει την τοπική μονάδα αποθήκευσης. Ένας απλός αλγόριθμος εύρεσης περιγράμματος δίνεται παρακάτω. Η είσοδος του αλγορίθμου μας είναι μια εικόνα που έχει υποστεί κατάτμηση και τα εικονοστοιχεία του φόντου φέρουν μια ετικέτα. Ο αλγόριθμος αυτός δεν μεταπηδά από περιοχή σε περιοχή αλλά εκτελεί τη σάρωση της εικόνας σε όλες τις διευθύνσεις πάνω-κάτω, δεξιά -αριστερά. Καθ' όλο το χρόνο εκτέλεσης οι περιοχές κατηγοριοποιούνται σε *προσωρινές περιοχές* που δεν έχει διαγραφεί το περίγραμμά τους, σε *ολοκληρωμένες περιοχές* όπου έχουμε πάρει το περίγραμμά και *προς σάρωση περιοχές* εκείνες που δεν έχει φθάσει ακόμα η σάρωση.

Ο αλγόριθμος ελέγχει τρεις σειρές ανά εκτέλεση : την τρέχουσα σειρά, την αμέσως από πάνω και την αμέσως από κάτω. Δύο σειρές λειτουργούν ως σειρές από το φόντο. Στη διαδικασία που περιγράφεται  $S$  είναι το όνομα της εικόνας επομένως  $S[R,C]$  είναι το όνομα του rixel που σαρώνει και παίρνει την τιμή (LABEL). Αν η ετικέτα αυτή είναι νέα τότε καταχωρείται στο σύνολο CURRENT. NEIGHB είναι το σύνολο των rixel γειτονικών του  $[R,C]$  που έχουν ετικέτα (LABEL). Η συνάρτηση rixeltype ελέγχει τις τιμές του  $[R,C]$  και των γειτονικών rixel για να αποφασίσει αν το  $[R,C]$  ανήκει στο φόντο. Αν ανήκει η διαδικασία αρχίζει αναζήτηση για μια αλυσίδα με την τιμή LABEL που να έχει ένα γειτονικό rixel του  $[R,C]$  ως τελευταίο και αν το βρεί μέσω της διαδικασίας add σπρώχνει το  $[R,C]$  στο τέλος της αλυσίδας. Στο ενδεχόμενο κανένα από τα γειτονικά του  $[R,C]$  να βρίσκεται στο τέλος κάποιας αλυσίδας δημιουργείται μια καινούρια αλυσίδα με το  $[R,C]$  ως μοναδικό στοιχείο μέσω της make\_new\_chain. Αφου σαρωθεί ολόκληρη η εικόνα οι αλυσίδες από τις περιοχές που έχει βρεθεί το περίγραμμα συγχωνεύονται σε ένα περίγραμμα με αποτέλεσμα να έχουμε εξάγαγει τα όρια της εικόνας.

### 3. Η μέθοδος active contours

#### 3.1 Εισαγωγή

Η τεχνική active contours (snakes) είναι ένας τρόπος να διαχωρίσουμε το περίγραμμα ενός αντικειμένου από μια πιθανώς θορυβώδη εικόνα 2D. Μπορούμε να φανταστούμε τα παραγόμενα πολυώνυμα ως φίδια, τα οποία παραμορφώνονται με τον χρόνο προσπαθώντας να πάρουν το σχήμα του περιγράμματος του αντικειμένου. Ο αλγόριθμος προσεγγίζει επαναληπτικά το περίγραμμα ανάγοντας το πρόβλημα κατάτμησης εικόνας στην ελαχιστοποίηση κάποιου κατάλληλα σχεδιασμένου συναρτησιακού ενέργειας  $J$  που ορίζεται στις υποπεριοχές  $R=R_1, \dots, R_N$  της εικόνας. Η συνάρτηση  $J$  γράφεται ως:

Στον παραπάνω τύπο η συνάρτηση  $f_{x,R}$  ποσοτικοποιεί την ομογένεια ενός στοιχείου  $x$  της εικόνας με τα στοιχεία της εκάστοτε υποπεριοχής  $R$  και η συνάρτηση  $g(S)$  είναι μια φθίνουσα συνάρτηση. Έχουμε ως είσοδο το πλήθος των σημείων, έναν όρο εσωτερικής ελαστικής ενέργειας και έναν όρο που αντιστοιχεί σε εξωτερική ενέργεια.

Για την ελαχιστοποίηση αυτή χρησιμοποιείται η μέθοδος της ταχύτερης καθόδου και η καμπύλη εξελίσσεται σύμφωνα με μία μερική διαφορική εξίσωση μεχρις ότου η καμπύλη να παραμένει χρονικά αμετάβλητη. Ιδανικά, η εξίσωση αυτή ισορροπεί όταν η καμπύλη έχει περιγράψει το υπό κατάτμηση αντικείμενο, δηλαδή το γράμμα. Η κατασκευή της μερικής διαφορικής εξίσωσης γίνεται με

βάση τις εξισώσεις Euler-Lagrange για την ελαχιστοποίηση συναρτησιακού ενέργειας.

Η εξωτερική ενέργεια ελαχιστοποιείται όταν η προσέγγιση μας φθάνει στα όρια του περιγράμματος, συνήθως ο ακριβέστερος τρόπος να το ελέγξουμε είναι να πάρουμε την πρώτη παράγωγο όπου αυτή παίρνει μέγιστες τιμές.

Η εσωτερική ενέργεια ελαχιστοποιείται όταν η καμπύλη έχει πάρει σχήμα παραπλήσιο του σχήματος του αντικειμένου. Ο πιο ακριβής τρόπος δίνει υψηλές τιμές για σχετικά επιμήκη περιγράμματα και για υψηλής καμπυλότητας περιγράμματα θεωρώντας πως το σχήμα είναι όσο πιο ομαλό γίνεται.

Η παραπάνω προσέγγιση χρησιμοποιήθηκε αρχικά στην μέθοδο Snakes (boundary-based active contours) στην οποία η αναπαράσταση του εξελισσόμενου περιγράμματος δίνεται σε παραμετρική μορφή. Ένα μειονέκτημα της μεθόδου αυτής είναι η μεγάλη εξάρτηση του αποτελέσματος από την επιλογή του σημείου εκκίνησης της κατάτμησης εικόνας καθώς επίσης και το γεγονός ότι το μοντέλο δεν μπορεί απευθείας να διαιρέσει ή να συγχωνεύσει τις εξελισσόμενες καμπύλες.

### 3.2 Περιγραφή της μεθόδου

Συνήθως είναι αρκετά δύσκολο να προσδιορίσουμε το πως μεταβάλλονται τοπικά ορισμένες δομές όπως τα χρώματα που περιέχονται σε μια εικόνα. Πράγματι είναι συχνό τα σημεία στα οποία παρατηρούμε απότομες μεταβολές στις παραμέτρους να ανήκουν και στο περίγραμμα του αντικειμένου που προσπαθούμε να διαχωρίσουμε από την υπόλοιπη εικόνα.

### 3.3 Περιγραφή της μεθόδου Geodesic Active Contours

Στόχος της Geodesic Active Contours ήταν να ξεπεράσει τα μειονεκτήματα που εμφάνιζε η μέθοδος active contours κυρίως στην εισαγωγή παραμέτρων γι' αυτό και στη μέθοδο αυτή δεν είναι απαραίτητη κάποια αρχικοποίηση των παραμέτρων. Η μέθοδος αυτή βασίζεται στον υπολογισμό των γεωδαιτικών καμπυλών ή των καμπυλών ελάχιστης απόστασης, με μία μετρική εξαρτώμενη από χαρακτηριστικά της εικόνας.

Πιο συγκεκριμένα, στο μοντέλο Geodesic Active Contours ελαχιστοποιείται το γεωδαιτικό μήκος της καμπύλης. Η ελαχιστοποίηση οδηγεί σε μία μερική διαφορική εξίσωση για την εξέλιξη της καμπύλης η οποία υλοποιείται χρησιμοποιώντας τη μέθοδο των επιπεδοσυνόλων (Level sets). Στην υλοποίηση των επιπεδοσυνόλων, το ενεργό περίγραμμα καθορίζεται έμμεσα ως το μηδενικό επιπεδοσύνολο μιας βαθμωτής συνάρτησης. Έτσι οι εξελισσόμενες καμπύλες διαχωρίζονται και συγχωνεύονται και καθίσταται δυνατό να εντοπίζονται ταυτόχρονα δομές που δεν συνδέονται μεταξύ τους. Βέβαια ακόμα και σε αυτό το μοντέλο απαιτείται μια κατάλληλη αρχικοποίηση, καθώς η αρχική καμπύλη θα πρέπει να βρίσκεται εξολοκλήρου εσωτερικά η εξωτερικά των ορίων του αντικειμένου (ή του γράμματος).

Στη μέθοδο active contours που αναλύθηκε παραπάνω η συνάρτηση ενέργειας μοντελοποιείται εισάγοντας το μήκος της καμπύλης  $C$

και το εμβαδόν της επιφάνειας που περικλείει η καμπύλη  $C$ . Έτσι, έχουμε την παρακάτω συνάντηση ενέργειας:

Έστω ότι έχουμε ένα κλειστό περίγραμμα  $C$ , και  $\omega$  είναι η επιφάνεια που περικλείει η καμπύλη  $C$  ενώ  $\Omega$  είναι το συμπλήρωμα του  $\omega$ . Άρα ισχύει ότι  $\int_C \phi \, ds = \int_{\Omega} \text{grad} \phi \cdot \mathbf{n} \, d\omega$  όπου η καμπύλη  $C$  ορίζεται ως το μηδενικό επιπεδοσύνολο μιας συνάρτησης Lipschitz :

Όπου

Με χρήση της συνάρτησης Heaviside,  $H(z)$  και της συνάρτησης Dirac, οι όροι του συναρτησιακού ενέργειας  $F$  εκφράζονται ως εξής:

Και

Τελικά η  $F(c_1, c_2, \phi)$  γράφεται:

Για τη λύση μπορεί να χρησιμοποιηθεί το παρακάτω διαμορφωμένο επιπεδοσύνολο:

Διατηρώντας σταθερή την συνάρτηση  $\phi$  και μειώνοντας την ενέργεια  $F$  έχουμε:

Αν  $\int_{\Omega} H\phi \, dx \, dy > 0$ , δηλαδή το εσωτερικό της καμπύλης δεν είναι το κενό

Αν  $\int_{\Omega} (1-H\phi) \, dx \, dy > 0$ , το εξωτερικό της καμπύλης δεν είναι το κενό.

Να σημειώσουμε πως σε κάποιες εκφυλισμένες περιπτώσεις, όπου δεν υπάρχουν περιορισμοί για την τιμή των σταθερών και, οι σταθερές αυτές υπολογίζονται από τους παρακάτω τύπους:

## 5 Η μέθοδος Watershed

### 5.1 Βασική σύλληψη της μεθόδου Watershed

Μια από τις κυριότερες μεθόδους κατάτμησης είναι η γραμμή watershed. Στον αλγόριθμο αυτό μια μονοχρωματική εικόνα επιπέδων του γκρι θεωρείται ως ένα τοπογραφικό ανάγλυφο, όπου η φωτεινότητα του κάθε εικονοστοιχείου αντιστοιχίζεται με το ύψος του ανάγλυφου σε εκείνο το σημείο. Η τεχνική αυτή δανείζεται στοιχεία από την τοπολογία έναν κλάδο των μαθηματικών που καταπιάνεται με τη μαθηματικά αναπαράσταση μορφολογιών. Ο αλγόριθμος watershed χρησιμοποιείται κυρίως σε εικόνες κλίσης (gradient images) και εντοπίζει τις λεκάνες απορροής (catchment basins). Η εικόνα προσομοιάζεται με κάποιο τοπογραφικό ανάγλυφο κατά τον Beucher ο οποίος περιέγραψε τη μέθοδο ως εξής: Στα τοπικά ελάχιστα τοποθετούνται πηγές με αποτέλεσμα να

πλημμυρίζεται ομοιόμορφα η εικόνα - τοπογραφικό ανάγλυφο. Όταν εμφανιστεί διαρροή που σηματοδοτεί την συνένωση δύο κοιλάδων υψώνεται ένα φράγμα ανάμεσα στις δυό περιοχές. Άμα ενώσουμε όλα τα φράγματα παίρνουμε τη γραμμή watershed. Ένα μειονέκτημα που εμφάνισε η μέθοδος ήταν η υπερκατάτμηση (oversegmentation) που οφείλονταν στο μεγάλο πλήθος τοπικών ελαχίστων που έβρισκε στην εικόνα τα περισσότερα εκ των οποίων ήταν άσχετα με το επιθυμητό αποτέλεσμα. Τα αποτελέσματα θα μπορούσαν να βελτιωθούν ως ένα βαθμό, αν φιλτράραμε την εικόνα των κλίσεων πριν τον μετασχηματισμό. Το πρόβλημα αντιμετωπίστηκε με τη προσέγγιση δεικτών (markers), σε αυτή την προσέγγιση σε κάθε αντικείμενο ανατίθεται ένας δείκτης συμπεριλαμβανομένου και του φόντου της εικόνας. Τότε η κατάτμηση γίνεται σε δύο βήματα. Πρώτον ο αλγόριθμος ανιχνεύει δείκτες για τα αντικείμενα και το φόντο. Δεύτερον, παράγεται η γραμμή watershed αυτή τη φορά πλημμυρίζοντας τις κοιλάδες με πηγές που ταυτίζονται με τις περιοχές που έχει τοποθετηθεί δείκτης. Αφού γεμίσει η εικόνα έχουν μείνει τα τμήματα της εικόνας που έχουν κάποιο δείκτη.

## 5.2 Ορισμός στον συνεχή χώρο

Για να ορίσουμε τη γραμμή watershed στον συνεχή χώρο θα πρέπει να προσέξουμε έναν αστάθμητο παράγοντα, τον θόρυβο που περιέχουν οι εικόνες που επεξεργαζόμαστε. Το ζήτημα αυτό παρακάμπτεται άμα ορίσουμε τη συνάρτηση της γραμμής watershed ακόμα και σε σημεία που δεν είναι ομαλή.

Ορίζω συνάρτηση  $f : R^n \rightarrow R$  έχω το  $\sup(f)$ , το διάστημα  $T$  στο  $R$  και τη συνεχή συνάρτηση  $\gamma$  από το  $T$  στο  $\sup(f)$ , έχουμε τη διαδρομή  $(T, \gamma)$  που περιέχεται στο  $\text{supp}(f)$  και  $\zeta = (t_1 < t_2 < \dots < t_n)$  ένα πεπερασμένο μέρος του  $T$ .

Επομένως μπορώ να ορίσω:

- Τοπογραφική μεταβολή της συνάρτησης  $f$  κατά μήκος της πολυγωνικής γραμμής  $\zeta$  ως:

$$TV\zeta = \sum_i |f(\gamma_i) - f(\gamma_{i+1})|$$

- Τοπογραφική απόσταση δυο σημείων  $p, q$  θεωρώντας το σύνολο  $\Gamma(p, q)$  υποσύνολο του  $\text{supp}(f)$ :

$$TD_{p,q} = \inf_{\gamma \in \Gamma(p,q)} TV\zeta$$

- Λεκάνη απορροής (catchment basin)  $CB(m_i)$  του τοπικού ελαχίστου  $m_i$  ως το σύνολο των σημείων  $x \in \text{supp}(f)$  τα οποία βρίσκονται πιο κοντά στο τοπικό ελάχιστο  $m_i$  από οποιαδήποτε άλλη τοπογραφική απόσταση. Δηλαδή:

$$\forall j \in T, j \neq i : TD_{x, m_i} < TD_{x, m_j}$$

- Γραμμή Watershed μιας συνάρτησης  $f$  το σύνολο των σημείων του  $\text{supp}(f)$  που δεν ανήκουν σε καμία λεκάνη απορροής.

$$Wshf = \text{supp}(f) \setminus \bigcup_i (CB_{m_i})^c$$

Από τον τελευταίο ορισμό γίνεται φανερό πως ο μετασχηματισμός watershed της  $f$  θέτει έναν ξεχωριστό δείκτη σε κάθε λεκάνη απορροής και ομαδοποιεί τα σημεία της γραμμής watershed της  $f$  βάζοντας μια ετικέτα  $w$  σε κάθε ένα από αυτά.

5.3 Ορίζοντας τη γραμμή watershed με τη βοήθεια της συνάρτησης τοπογραφικής απόστασης

Η συνάρτηση τοπογραφικής απόστασης υπολογίζει σε εικόνες με σταθερή κλίση ποσότητες αντίστοιχες με την γεωδαισιακή απόσταση της αντίστοιχης προβολής πολλαπλασιασμένης με κάποιο συντελεστή .

Θεωρούμε μια συνάρτηση τόνων του γκρι  $f$  από το  $Z^n$  στο  $Z$  και το support της  $\text{supp}(f)$ . Έστω  $G$  το υποκείμενο πλέγμα, το οποίο μπορεί να είναι οποιουδήποτε τύπου. Συσχετίζεται στο  $G$  ένας γειτονικός γράφος  $U$ , όπου είναι ένα υποσύνολο  $Z^n \times Z^n$  το οποίο ορίζεται από  $(a,b) \in U$  αν και μόνο αν τα  $a$  και  $b$  είναι γειτονικά

Ονομάζεται  $N_U(p)$  τα γειτονικά εικονοστοιχεία ενός εικονοστοιχείου  $p$ , ως προς το  $U$  και το  $\text{supp}(f)$ :  $N_U(p) = \{p' \in Z^n, (p,p') \in U \cap \text{supp}(f)\}$ . Το υποσύνολο  $B(p)$  όλων των εικονοστοιχείων του  $N_U(p)$  που βρίσκονται σε απόσταση 1 από το  $p$  καλείται μοναδιαίος δίσκος μεγέθους 1. Το erosion της συνάρτησης  $f$  από αυτόν τον δίσκο είναι το στοιχειώδες erosion  $ef$ .

5.4 Η γραμμή watershed στον διακριτό χώρο

Στον διακριτό χώρο, πάντα υπάρχουν τα ίδια προβλήματα: απειροελάχιστα δομικά στοιχεία δεν υπάρχουν στο ψηφιακό πλέγμα και οι γειτονικές σχέσεις είναι αδύναμες. Παρ' όλα αυτά είναι δυνατόν να γίνουν καλές προσεγγίσεις της γραμμής watershed.

6. **BSDS500** :Μια μέθοδος ανίχνευσης περιγράμματος και ιεραρχικής κατάτμησης εικόνας

6.1 Προηγούμενη έρευνα

Τα προβλήματα που αντιμετωπίζουν οι μέθοδοι της ανίχνευσης περιγράμματος και της κατάτμησης εικόνας είναι ανάλογα χωρίς να ταυτίζονται. Σε γενικές γραμμές οι ανιχνευτές περιγράμματος δεν παράγουν εγγυημένα κάποιο κλειστό περίγραμμα, συνεπώς δεν παρέχουν απαραίτητα μια διαμέριση της εικόνας σε υποπεριοχές. Μπορεί κανείς όμως να ανακτήσει πάντοτε κλειστά περιγράμματα από περιοχές με την μορφή των ορίων αυτών των περιοχών. Η μέθοδος που εισάγεται στο... δείχνει πώς να επιτύχει κανείς το αντίστροφο: πώς δηλ. να ανακτήσει κανείς περιοχές από έναν ανιχνευτή περιγράμματος. Στη συνέχεια παρουσιάζουμε μια σύντομη ιστορική ανασκόπηση των προσεγγίσεων στα δύο αυτά προβλήματα, που γενικά ήταν αρκετά διαφορετικές.

6.1.1) Περιγράμματα.

Οι αρχικές προσεγγίσεις στην ανίχνευση περιγραμμάτων είχε σαν στόχο την ποσοτικοποίηση της παρουσίας ενός ορίου/συνόρου σε μια συγκεκριμένη θέση μιας εικόνας μέσω τοπικών μετρήσεων (πχ. με χρήση zero crossings των τελεστών Laplace ή Gauss, με την χρήση της προσέγγισης Oriented Energy ή του ανιχνευτή Canny που βασίζεται στην απόδοση/μοντελοποίηση των άκρων σε ένα περίγραμμα ως έντονων ασυνεχειών στο κανάλι της φωτεινότητας κ.α.).

Πιο πρόσφατες τοπικές προσεγγίσεις παίρνουν υπόψιν τους πληροφορίες σχετικά με την «υφή» και το χρώμα κάνοντας χρήση τεχνικών εκμάθησης για cue combination (πχ. με τον καθορισμό τελεστών βαθμίδας (gradient operators) για τα κανάλια φωτεινότητας, χρώματος και υφής που χρησιμοποιούνται στη συνέχεια για δεδομένα εισόδου σε έναν regression classifier για την πρόβλεψη της έντασης των ορίων). Άλλοι προτείνουν έναν αλγόριθμο BEL (Boosted Edge Learning) που προσπαθεί να «εκπαιδεύσει» έναν edge classifier μέσω ενός στοχαστικού boosting tree (Dollar et al.) από χιλιάδες απλών χαρακτηριστικών υπολογισμένων από διάφορα image patches.

Ένα πρόβλημα που παραμένει γι' αυτές τις πιο σύγχρονες προσεγγίσεις είναι το μεγάλο εύρος κλίμακας στις οποίες μπορούν να εμφανίζονται τα διάφορα αντικείμενα στις εικόνες. Κάποιοι (πχ. Ren) προσπαθούν να εκμεταλλευτούν τον συνδυασμό πληροφορίας από πολλαπλές κλίμακες των τοπικών τελεστών, προμηθεύοντας στον boundary classifier επιπρόσθετα cues τοπικοποίησης και σχετικής αντίθεσης (contrast), προσδιορισμένων με όρους του output από τους ανιχνευτές πολλαπλής κλίμακας. Για κάθε κλίμακα, η localisation cue «συλλαμβάνει» την απόσταση ενός εικονοστοιχείου από την πλησιέστερη μέγιστη απόκριση. Η σχετική contrast cue κανονικοποιεί κάθε εικονοστοιχείο σε όρους της τοπικής γειτονιάς.

Μια άλλη ομάδα αλγορίθμων συσχετίζει συνοριακά κομμάτια έντονης διαβάθμισης (gradient) ώστε να ταυτοποιήσουν εκτεταμένα, ομαλά περιγράμματα. Αυτοί αντιπροσωπεύουν μια γραμμή προσέγγισης που εστιάζει σε ένα άλλο επίπεδο επεξεργασίας, την καθολικοποίηση (globalisation), αξιοποιώντας τοπικά δεδομένα των ανιχνευτών. Οι πιο πρόσφατες μέθοδοι υψηλής-απόδοσης, μεταξύ των οποίων και η μέθοδος BSD (Berkeley Segmentation Dataset), χαρακτηρίζονται από ακόμα πιο εξεζητημένα επίπεδα καθολικοποίησης, με κοινό γνώρισμα την χρήση τελεστών ανίχνευσης τοπικών άκρων από το [2].

Το Conditional Random Fields framework χρησιμοποιείται από άλλους (Ren et al.) για την επιβολή μιας καμπύλης συνέχειας στα περιγράμματα. Υπολογίζοντας μια περιορισμένη τριγωνοποίηση Delaunay (CDT) πάνω από τοπικά ανιχνευμένα περιγράμματα, παίρνουν ένα γράφημα αποτελούμενο από τα περιγράμματα αυτά μαζί με την καινούρια «συμπληρωμένα» άκρα/όρια που προκύπτουν



από την τριγωνοποίηση. Η μέθοδος CDT είναι αναλλοίωτη ως προς την κλίμακα και τείνει να καλύψει μικρά κενά στα ανιχνευθέντα περιγράμματα. Συσχετίζοντας μια τυχαία μεταβλητή με κάθε περίγραμμα και κάθε συμπληρωμένο άκρο, ορίζουμε ένα CRF με δυναμικά άκρων σε όρους της απόκρισης των ανιχνευτών και δυναμικά κορυφών (των τριγώνων) με όρους τύπου «συνάρθρωσης» (junction) και ομαλότητας συνέχειας. Αναμενόμενες τιμές προκύπτουν τότε με τη χρήση loopy belief propagation.

Μια διαφορετική στρατηγική εξαγωγής προεξάρχουσες ομαλές καμπύλες από τα αποτελέσματα ενός τοπικού ανιχνευτή περιγραμμάτων συνίσταται στο να θεωρήσει κανείς το σύνολο των μικρού μήκους προσανατολισμένων ευθύγραμμων τμημάτων που συνδέουν εικονοστοιχεία μιας εικόνας με γειτονικά τους εικονοστοιχεία. Κάθε τέτοιο τμήμα είναι είτε κομμάτι μιας καμπύλης είτε ένα τμήμα υποβάθρου. Υποθέτοντας ότι οι καμπύλες σχεδιάζονται μέσω μιας διαδικασίας Markov η κύρια κατανομή των καμπυλών ευνοεί μερικές μόνο για κάθε σκηνή, και οι αποκρίσεις του ανιχνευτή είναι ανεξάρτητες από τις συνθήκες δεδομένης της τιτλοποίησης των τμημάτων. Η εύρεση της βέλτιστης «τιτλοποίησης» μεταφράζεται τότε σε ένα γενικό πρόβλημα πρόβλημα εύρεσης weighted min-cover στο οποίο τα τμήματα που καλύπτονται είναι τα ίδια τα τμήματα γραμμών και τα στοιχεία που τα καλύπτουν σχεδιάζονται από το σύνολο όλων των πιθανών καμπυλών και όλων των πιθανών τμημάτων στο υπόβαθρο. Η πολυπλοκότητα του προβλήματος είναι NP οπότε μια προσεγγιστική λύση βρίσκεται με τη χρήση μιας greedy «cost-per-pixel» ευριστικής.

### 6.1.2 Περιοχές

Μια ευρεία οικογένεια προσεγγίσεων στην κατάτμηση εικόνας περιλαμβάνει την ενσωμάτωση γνωρισμάτων όπως η φωτεινότητα, το χρώμα ή η υφή σε τοπικά patches εικόνας και στην ομαδοποίηση (clustering) των γνωρισμάτων αυτών στη συνέχεια με βάση διάφορα μοντέλα όπως τα fitting mixture, mode-finding, graph partitioning κ.α. Οι πιο διαδεδομένοι αλγόριθμοι που χρησιμοποιούνται στην κατηγορία αυτή είναι τρεις.

Ο πρώτος είναι ο αλγόριθμος graph-based region merging (Felz-Hutt), που προσπαθεί να διαμερίσει τα εικονοστοιχεία μιας εικόνας σε συνιστώσες που να μην είναι ούτε ιδιαίτερα «χονδροειδείς» ούτε ιδιαίτερα «λεπτές». Με δεδομένο έναν γράφο στον οποίον οι κόμβοι είναι εικονοστοιχεία και τα «βάρη» των ακμών μετρούν την ανομοιότητα μεταξύ των κόμβων (πχ. διαφορές χρώματος), κάθε κόμβος τίθεται αρχικά στην δική του συνιστώσα.

Ο αλγόριθμος Mean Shift ([34]), παρέχει ένα εναλλακτικό πλαίσιο για clustering. Εδώ τα εικονοστοιχεία αναπαρίστανται στον joint spatial-range χώρο/πεδίο συμπαραθέτοντας/concatenating τις

χωρικές συντεταγμένες και τις χρωματικές τιμές τους σε ένα μοναδικό διάνυσμα. Εφαρμόζοντας ένα mean-shift φιλτράρισμα στο πεδίο αυτό παίρνουμε ένα σημείο σύγκλισης για κάθε εικονοστοιχείο. Περιοχές σχηματίζονται ομαδοποιώντας όλα εκείνα τα εικονοστοιχεία που τα σημεία σύγκλισής τους είναι εγγύτερα από μια απόσταση  $h_s$  στον χώρο αποστάσεων και  $h_r$  στο range domain.

Η φασματική θεωρία γράφων, και συγκεκριμένα το κριτήριο των Normalized Cuts, παρέχει έναν τρόπο ενσωμάτωσης καθολική πληροφορία για μια εικόνα στη διαδικασία ομαδοποίησης. Στο πλαίσιο αυτό, με δεδομένο έναν πίνακα συνάφειας (affinity matrix), του οποίου τα στοιχεία κωδικοποιούν την ομοιότητα ανάμεσα στα εικονοστοιχεία ορίζουμε τον διαγώνιο πίνακα  $D_{ii} = \sum W_{ij}$  και βρίσκουμε τα γενικευμένα ιδιοδιανύσματα του γραμμικού συστήματος:

$$(D-W)v = \lambda D_v$$

Μετά από το βήμα αυτό, παραδοσιακά εφαρμόζεται η μέθοδος K-means clustering για να πάρουμε μια κατάτμηση σε περιοχές.

Μια πιο πρόσφατη εκδοχή της μεθόδου Ncuts είναι η Multiscale Ncuts που αντιμετωπίζει το πρόβλημα του απαγορευτικά «ακριβού» υπολογισμού των ιδιοτιμών λόγω της «αραιότητας» του πίνακα  $W$ , με τον υπολογισμό κατάλληλων πινάκων σε πολλαπλές κλίμακες που οδηγεί σε ένα καινούριο πρόβλημα ιδιοτιμών.

Μια εναλλακτική προσέγγιση με σκοπό τη βελτίωση της υπολογιστικής αποτελεσματικότητας της μεθόδου Ncuts, εμπνεόμενη από το αλγεβρικό multigrid, καθιστά πιο «χονδροειδές» το αρχικό γράφημα/γράφο επιλέγοντας ένα υποσύνολο των αρχικών κόμβων με τέτοιο τρόπο ώστε κάθε μεταβλητή στο «λεπτότερο» επίπεδο να είναι ισχυρά συζευγμένη με μια μεταβλητή στο «χονδροειδέστερο».

Αρκετές προσεγγίσεις στην κατάτμηση εικόνας εντάσσονται σε μια άλλη κατηγορία από αυτές που είδαμε μέχρι τώρα, βασιζόμενες στην διαμόρφωση του προβλήματος σε ένα variational πλαίσιο (πχ. το μοντέλο που προτείνεται από τους Mumford και Shaz και στο οποίο η κατάτμηση της υπό παρατήρηση εικόνας δίνεται από την ελαχιστοποίηση ενός κατάλληλου συναρτησιοειδούς).

### 6.1.3 Benchmarks (Δείκτες Αναφοράς)

Αν και το BSDS έχει βρει ευρεία αποδοχή σαν benchmark στην ανίχνευση περιγράμματος – παρόλο που η σχετική βιβλιογραφία προηγείται χρονικά της εμφάνισής του – ένα αντίστοιχο πρότυπο για την εκτίμηση/αξιολόγηση των αλγορίθμων κατάτμησης είναι λιγότερο ξεκάθαρο.

Μια επιλογή είναι να θεωρήσει κανείς τα ίδια τα όρια των κατατμήσεων σαν περιγράμματα και να τα αξιολογήσει ως τέτοια. Είναι όμως επιθυμητό να διαθέτει κανείς και μια μεθοδολογία που να μετρά άμεσα την ποιότητα των κατατμήσεων. Για παράδειγμα, μερικά σφάλματα όπως η απουσία ενός εικονοστοιχείου στο σύνορο μεταξύ δύο περιοχών μπορεί να μην αντανακλάται στην στην αξιολόγηση ενός περιγράμματος αλλά να έχει πολύ σημαντικές συνέπειες στην ποιότητα μιας κατάτμησης, πχ. να οδηγεί σε μια λανθασμένη συγχώνευση μεγάλων περιοχών. Μπορούμε να πούμε ότι η αξιολόγηση συνόρων/ορίων ευνοεί μάλλον την χρήση ανιχνευτών περιγράμματος παρά μεθόδους κατάτμησης, αφού οι πρώτοι δεν πάσχουν από τον περιορισμό της παραγωγής κλειστών καμπυλών, οπότε πρέπει κανείς να λάβει υπόψιν και μετρικές που βασίζονται σε περιοχές.

Οι βασικές μετρικές είναι:

#### 6.1.3.1 Μεταβολή της Πληροφορίας (Variation of Information)

Η μετρική της μεταβολής της πληροφορίας εισήχθη με σκοπό την σύγκριση των cluster και μετρά την απόσταση μεταξύ δύο κατατμήσεων σε όρους της μέσης conditional εντροπίας τους που δίνεται από την εξίσωση:

$$VI(S,S') = H(S) + H(S') - 2I(S,S')$$

όπου τα  $H$ ,  $I$  αντιπροσωπεύουν αντίστοιχα τις εντροπίες και την αμοιβαία πληροφορία μεταξύ δύο clustering δεδομένων  $S$ ,  $S'$ . Στην περίπτωση του BSDS αυτά τα clustering είναι test και ground-truth κατατμήσεις.

Αν και η συνάρτηση  $VI$  διαθέτει κάποιες ενδιαφέρουσες θεωρητικές ιδιότητες το αντιληπτικό της περιεχόμενο και η εφαρμοσιμότητά της στην περίπτωση αρκετών ground-truth κατατμήσεων παραμένει ασαφής.

#### 6.1.3.2 Δείκτης Κατάταξης (Rank Index)

Ο Δείκτης Κατάταξης εισήχθη αρχικά για γενική αξιολόγηση clustering. Λειτουργεί με την σύγκριση της συμβατότητας των συσχετίσεων ανάμεσα σε ζεύγη στοιχείων των clustering. Ο δείκτης κατάταξης ανάμεσα σε δύο κατατμήσεις test και ground-truth  $S$  και  $G$ , δίνεται από το άθροισμα του αριθμού των ζευγών εικονοστοιχείων που έχουν την ίδια ετικέτα και στις δύο κατατμήσεις και του αριθμού εκείνων που έχουν διαφορετικές ετικέτες και στις δύο κατατμήσεις διαιρεμένο με τον συνολικό αριθμό των ζευγών εικονοστοιχείων. Έχουν προταθεί επίσης παραλλαγές του Δείκτη Κατάταξης για την αντιμετώπιση περιπτώσεων πολλαπλών ground-truth κατατμήσεων.

### 6.1.3.3 Κάλυμμα Κατάτμησης (Segmentation Covering)

Η αλληλεπικάλυψη μεταξύ δύο περιοχών  $R$  και  $R'$  οριζόμενη ως:

έχει χρησιμοποιηθεί για την εκτίμηση της ενέργειας/task ταξινόμησης σε εύρος εικονοστοιχείου στην αναγνώριση εικόνας. Με βάση αυτήν ορίζεται το κάλυμμα μιας κατάτμησης  $S$  από μια κατάτμηση  $S'$ .

Αντίστοιχα το κάλυμμα μιας μηχανικής κατάτμησης  $S$  από μια οικογένεια ground-truth κατατμήσεων  $\{G_i\}$  ορίζεται καλύπτοντας ξεχωριστά πρώτα το  $S$  από κάθε ανθρώπινη κατάτμηση  $G_i$  και στη συνέχεια παίρνοντας μέσες τιμές πάνω στις διαφορετικές ανθρώπινες κατατμήσεις. Για να πετύχει τέλεια κάλυψη η μηχανική κατάτμηση θα πρέπει να εξηγήσει όλα τα ανθρώπινα δεδομένα. Στη συνέχεια μπορεί κανείς να ορίσει δύο ποιοτικούς περιγραφείς (descriptors) για περιοχές, το κάλυμμα της  $S$  από τις  $\{G_i\}$  και το κάλυμμα των  $\{G_i\}$  από την  $S$ .

## 6.2. Ανίχνευση Περιγράμματος

Ξεκινώντας από μια συνάρτηση  $P_b(x,y,\theta)$  (Martin et al.) που προβλέπει την posterior πιθανότητα ενός ορίου με προσανατολισμό  $\theta$  σε κάθε εικονοστοιχείο  $(x,y)$  της εικόνας μετρώντας τη διαφορά στα κανάλια της τοπικής φωτεινότητας, χρώματος και υφής της εικόνας, οι Arbelaez και Fowlkes εισάγουν μια εκδοχή του ανιχνευτή  $P_b$  για πολλαπλές κλίμακες, καθώς και μια καινούρια μέθοδο καθολικοποίησης.

### 6.2.1 Βαθμίδες Φωτεινότητας, Χρώματος και Υφής

Το βασικό στοιχείο στη διαμόρφωση του ανιχνευτή περιγράμματος  $P_b$  είναι ο υπολογισμός ενός προσανατολισμένου σήματος βαθμίδας  $G(x,y,\theta)$  από μια εντασιακή εικόνα  $I$ . Ο υπολογισμός γίνεται με την τοποθέτηση ενός κυκλικού δίσκου στην θέση  $(x,y)$  χωρισμένου σε δύο ημικύκλια από μια διάμετρο στην κλίση της γωνίας  $\theta$ . Για κάθε ημικυκλικό δίσκο παράγεται ένα ιστόγραμμα των τιμών έντασης για τα εικονοστοιχεία της εικόνας  $I$  που καλύπτονται από τον δίσκο αυτό. Το μέγεθος της βαθμίδας  $G$  ορίζεται ως εξής:

Το κίνητρο για έναν τέτοιο υπολογισμό είναι η διαίσθηση ότι τα περιγράμματα αντιστοιχούν σε ασυνέχειες στην εικόνα και ότι τα

ιστογράμματα παρέχουν έναν εύρωστο μηχανισμό για την μοντελοποίηση του περιεχομένου μιας περιοχής της εικόνας. Μια έντονη προσανατολισμένη απόκριση βαθμίδας/διαβάθμισης σημαίνει ότι το εικονοστοιχείο είναι πολύ πιθανόν να βρίσκεται στο σύνορο/όριο μεταξύ δύο διαφορετικών περιοχών.

Ο ανιχνευτής Pb συνδυάζει τα σήματα προσανατολισμένης βαθμίδας που παράγονται από τον μετασχηματισμό μιας εισαγόμενης εικόνας σε τέσσερα ξεχωριστά κανάλια χαρακτηριστικών και επεξεργάζεται το κάθε κανάλι ξεχωριστά. Τα τρία πρώτα αντιστοιχούν στα κανάλια του χρωματικού χώρου CIE Lab (κανάλια φωτεινότητας και χρώματος a και b αντίστοιχα). Το τέταρτο κανάλι είναι κανάλι υψής, που αποδίδει σε κάθε εικονοστοιχείο ένα `textron id`. Αυτά τα στοιχεία υπολογίζονται υποκείμενα σε ένα άλλο στάδιο φιλτραρίσματος που μετατρέπει την εισαγόμενη εικόνα σε βαθμίδες του γκρι και την συνελίσσει (convolutes) με το σύνολο των 17 Gaussian παραγώγων και φίλτρων που δείχνονται στο σχήμα 5. Κάθε εικονοστοιχείο συσχετίζεται με ένα διάνυσμα 17 συνιστωσών αποκρίσεων και τα διανύσματα ομαδοποιούνται με τη χρήση των K-means.

Στη συνέχεια σχηματίζεται μια εικόνα όπου κάθε εικονοστοιχείο έχει μια ακέραια τιμή  $[1, K]$  όπως καθορίζεται από το `textron id`. Στην εικόνα αυτή υπολογίζονται τελικά διαφορές στα ιστογράμματα στους προσανατολισμένους ημικυκλικούς δίσκους με τον ίδιο τρόπο που γίνεται για τα κανάλια φωτεινότητας και χρώματος. Το βασικό κομμάτι της μεθόδου είναι λοιπόν η εξαγωγή μιας συνάρτησης  $G(x,y,\theta)$  για την τυχαία είσοδο  $I$ .

### 6.2.2 Συνδυασμός των cues σε πολλαπλές κλίμακες.

Για να υπολογίσουν λεπτής ή πιο χονδροειδούς υψής δομές οι Arbelaez-Fowlkes θεωρούν βαθμίδες σε τρεις κλίμακες:  $[\sigma/2, \sigma, 2\sigma]$  για καθένα από τα τέσσερα κανάλια και στη συνέχεια συνδυάζουν γραμμικά τις τοπικές αυτές cues σε ένα μοναδικό προσανατολισμένο σήμα:

Παίρνοντας την μέγιστη απόκριση για διάφορους προσανατολισμούς δίνει ένα μέτρο της ισχύος του συνόρου σε κάθε εικονοστοιχείο:

### 6.2.3 Καθολικοποίηση.

Στην καρδιά του μηχανισμού καθολικοποίησης των A-F βρίσκεται η φασματική ομαδοποίηση (spectral clustering) με το κύριο σημείο διαφοροποίησης από άλλες μεθόδους να είναι ο «μαλακός» τρόπος με τον οποίον χρησιμοποιούνται τα ιδιοδιανύσματα που λαμβάνονται από την φασματική διαμέριση.

Η είσοδος στην φασματική ομαδοποίηση είναι ένας sparse συμμετρικός πίνακας συνάφειας  $W$  με τη χρήση του intervening contour cue.

Για να εισαχθεί καθολική πληροφορία ορίζεται ένας διαγώνιος πίνακας  $D_{ii} = \sum_j W_{ij}$  και επιλύεται για τα γενικευμένα ιδιοδιανύσματα του συστήματος

$$(D-W)v = \lambda Dv$$

Στο σημείο αυτό, η συνήθης μέθοδος Normalized Cuts συσχετίζει με κάθε εικονοστοιχείο έναν περιγραφέα μήκους  $n$  σχηματιζόμενου από τα  $n$  ιδιοδιανύσματα και με την χρήση ενός αλγορίθμου ομαδοποίησης όπως ο K-means δημιουργεί μια αδρή διαμέριση της εικόνας.

Για να αποφευχθεί το πρόβλημα της δημιουργίας μιας λανθασμένης κατάτμησης εξαιτίας του «σπασίματος» μεγάλων περιοχών στα οποία τα ιδιοδιανύσματα μεταβάλλονται σχετικά ομαλά, λαμβάνουμε υπόψιν ότι τα ιδιοδιανύσματα μεταφέρουν πληροφορία σχετικά με το περίγραμμα. Αντιμετωπίζοντας τα ίδια τα ιδιοδιανύσματα σαν εικόνες συλίσσουμε με Gaussian φίλτρα των κατευθυντικών παραγώγων για πολλαπλούς προσανατολισμούς παίρνοντας έτσι προσανατολισμένα σήματα  $\{ \}$ . Η χρήση των παραγώγων έχει σαν αποτέλεσμα να αγνοούνται τελικά οι ομαλές μεταβολές που πιθανά οδηγούν σε σφάλματα. Η πληροφορία από διαφορετικά ιδιοδιανύσματα συνδυάζεται τέλος για να πάρουμε την «φασματική» συνιστώσα του ανιχνευτή περιγράμματος:

$$sPb(x,y,\theta) = \dots$$

Τα σήματα  $mPb$  και  $sPb$  μεταφέρουν διαφορετική πληροφορία το καθένα (το πρώτο κυρίως από τις άκρες ενώ το δεύτερο από τις περισσότερο κυρίαρχες καμπύλες της εικόνας) και αρκεί ο γραμμικός τους συνδυασμός για να αξιοποιήσει κανείς και τις δύο αυτές συμπεριφορές. Η τελική καθολικοποιημένη πιθανότητα περιγράμματος γράφεται τελικά σαν το σταθμισμένο άθροισμα (weighted sum) τοπικών και φασματικών σημάτων:

$$gPb(x,y,\theta) =$$

η οποία ανα-κλιμακοποιείται με χρήση μιας σιγμοειδούς ώστε τα αποτελέσματα να ερμηνευθούν όντως σαν πιθανότητες.

#### 6.2.4 Αποτελέσματα.

Από ποιοτική άποψη, ο συνδυασμός των cues πολλαπλών κλιμάκων με τον μηχανισμό καθολικοποίησης έχει σαν αποτέλεσμα τον περιορισμό των clutter edges και την συμπλήρωση των περιγραμμάτων (πχ. Σχήμα 9).

Αναλύοντας τη συνεισφορά των διαφόρων κλιμάκων και των σημάτων στην απόδοση του ανιχνευτή gPb φαίνεται ότι ο περιορισμός των false positives συγκεντρώνεται κυρίως στα υψηλά thresholds ενώ ο ανιχνευτής gPb συνδυάζει το καλλίτερο των δύο «κόσμων», στηριζόμενος στον sPb στο πεδίο της υψηλής ακρίβειας και στον mPb στο πεδίο υψηλού recall.

Μια ματιά στην ανίχνευση περιγράμματος στα τεστ benchmarks του BSDS300 δείχνει μια μέση βελτίωση της ακρίβειας του gPb της τάξης του 10% σε σχέση με την απλή κλίμακα Pb.

#### 6.3. Κατάτμηση.

Τα περιγράμματα που παράγονται με την μέθοδο που περιγράψαμε στην προηγούμενη ενότητα (ανιχνευτής gPb) συχνά δεν είναι κλειστά οπότε και δεν μπορούν να διαμερίσουν μια εικόνα σε κατάλληλες περιοχές (αν και παραμένουν χρήσιμα, πχ. σαν σήματα για τον υπολογισμό περιγραφών της εικόνας). Οι κλειστές περιοχές παρέχουν όμως επιπλέον πλεονεκτήματα, έχουν πχ. Δικές τους εκτιμήσεις κλίμακας και παρέχουν ένα φυσικό πεδίο για τον υπολογισμό χαρακτηριστικών χρήσιμων στην αναγνώριση εικόνας. Ακόμα πολλές εργασίες μπορούν να επωφεληθούν από την μείωση της πολυπλοκότητας που επιτυγχάνεται με τον μετασχηματισμό μιας εικόνας εκατομμυρίων εικονοστοιχείων σε μερικές εκατοντάδες «υπερεικονοστοιχεία».

Οι A-F εισάγουν στην εργασία τους έναν καινούριο αλγόριθμο και μέθοδο για την ανάκτηση κλειστών περιοχών διατηρώντας ταυτόχρονα τα πλεονεκτήματα στην ποιότητα των περιγραμμάτων που επιτυγχάνεται με την διαδικασία της προηγούμενης ενότητας. Ο αλγόριθμος αυτός παράγει μια ιεραρχική κατάτμηση εισάγοντας μια καινούρια εκδοχή του μετασχηματισμού watershed, τον Oriented Watershed Transform (OWT), που παράγει ένα σύνολο αρχικών περιοχών από το προϊόν του ανιχνευτή περιγράμματος και την κατασκευή, στη συνέχεια ενός Ultrametric Contour Map (UCM) από τα σύνορα αυτών των αρχικών περιοχών.

Αυτή η ακολουθία διεργασιών παρέχει έναν γενικό μηχανισμό για να περάσει κανείς από περιγράμματα σε ένα ιεραρχικό δέντρο περιοχής. Τα περιγράμματα που κωδικοποιούνται στην ιεραρχική κατάτμηση που προκύπτει, διατηρούν πραγματικά βάρη (real-valued weights) που υποδεικνύουν την πιθανότητα να είναι πραγματικά σύνορα. Για ένα συγκεκριμένο κατώφλι, το προϊόν είναι ένα σύνολο από κλειστά περιγράμματα που μπορεί να αντιμετωπιστεί είτε σαν μια κατάτμηση της εικόνας είτε σαν ένας ανιχνευτής περιγράμματος για τους σκοπούς ενός benchmarking.

### 6.3.1 Oriented Watershed Transform.

Η μέθοδος OWT έχει να κάνει ακριβώς με την εφαρμογή του συνηθισμένου μετασχηματισμού watershed από την μαθηματική μορφολογία σε μια τοπογραφική επιφάνεια που δίνει ο υπολογισμός της συνάρτησης  $E(x,y)$ , δηλ. της μέγιστης απόκρισης της  $E(x,y,\theta)$  για διάφορους προσανατολισμούς (εξ ου και το oriented), όπου  $E(x,y,\theta)$  είναι φυσικά ο ανιχνευτής περιγράμματος. Είναι τα τοπικά ελάχιστα της  $E(x,y)$  που λαμβάνονται ως seed locations ομογενών κομματιών πάνω στα οποία εφαρμόζεται ο watershed μετασχηματισμός. Τα βυθίσματα (catchment basins) των ελαχίστων, τα οποία συμβολίζονται με  $P_0$ , δίνουν τις περιοχές της λεπτότερης διαμέρισης και τα αντίστοιχα τόξα,  $K_0$ , του watershed μετασχηματισμού δίνουν τις πιθανές τοποθεσίες των συνόρων.

Ένα πρόβλημα που εμφανίζεται είναι σταθμίζοντας (weighting) κάθε τόξο απλά με την μέση τιμή της  $E(x,y)$  στα εικονοστοιχεία του τόξου μπορεί να οδηγήσει σε πλασματικά αποτελέσματα. Αυτό οφείλεται στο ότι ο ανιχνευτής περιγράμματος παράγει μια χωρικά εκτεταμένη απόκριση γύρω από ισχυρά σύνορα. Έτσι μια αρχική υπερκατάτμηση μπορεί να περιέχει τόξα που κανονικά θα έπρεπε να είναι ασθενή αλλά παρόλα αυτά τέμνουν ισχυρά σύνορα.

Η διαδικασία OWT έχει σαν σκοπό να διορθώσει ακριβώς αυτό το πρόβλημα με την επιβολή μιας συνέπειας μεταξύ της ισχύος των περιγραμμάτων του  $K_0$  και του υποκείμενου σήματος  $E(x,y,\theta)$ . Το πρώτο βήμα σε αυτή την επαναστάθμιση είναι η εκτίμηση ενός προσανατολισμού σε κάθε εικονοστοιχείο του τόξου από την τοπική γεωμετρία του ίδιου του τόξου με την προσέγγιση των watershed τόξων με γραμμικά τμήματα όπως φαίνεται στο σχήμα 13. Υποδιαιρώντας αναδρομικά κάθε τόξο που δεν «ταιριάζει καλά» με το τμήμα που ενώνει τα άκρα του, παίρνουμε τους προσανατολισμούς του κάθε τόξου. Εκφράζοντας το κριτήριο προσέγγισης σε όρους της μέγιστης απόστασης ενός εικονοστοιχείου του τόξου από το ευθύγραμμο τμήμα σαν ένα κλάσμα του μήκους του τμήματος αυτού, παίρνουμε μια υποδιαίρεση που είναι ανεξάρτητη-κλίμακας. Σε κάθε εικονοστοιχείο  $(x,y)$  ενός υποδιαιρεμένου τόξου, αντιστοιχούμε τον προσανατολισμό (την γωνία)  $\theta(x,y)$  στο διάστημα  $[0,\pi)$  του αντίστοιχου ευθύγραμμου τμήματος.



### 6.3.2 Ultrametric Contour Map(Υπερμετρική Απεικόνιση Περιγράμματος).

Τα περιγράμματα έχουν το πλεονέκτημα ότι μπορούν να αναπαραστήσουν σχετικά εύκολα την αβεβαιότητα στην παρουσία ενός πραγματικά υποκείμενου περιγράμματος συσχετίζοντας μια δυαδική τυχαία μεταβλητή με αυτό. Μπορούμε να ερμηνεύσουμε στην πραγματικότητα την ισχύ του συνόρου που αποδίδεται σε ένα τόξο με βάση τον αλγόριθμο OWT σαν μια εκτίμηση της πιθανότητας ότι το τόξο αυτό ανήκει σε ένα πραγματικό περίγραμμα.

Η απεικόνιση UCM (Υπερμετρική Απεικόνιση Περιγράμματος) είναι μια δυνατότητα για να αποδοθεί αυτή η αβεβαιότητα σχετικά με μια κατάτμηση. Η απεικόνιση αυτή ορίζει μια δυσκολία ανάμεσα σε κλειστά, σταθμισμένα περιγράμματα που δεν τέμνουν τον εαυτό τους και μια ιεραρχία περιοχών. Τα κατώτερα επίπεδα της ιεραρχίας αυτής σέβονται ακόμα και τα «αδύναμα» περιγράμματα δίνοντας έτσι μιαν υπερκατάτμηση της εικόνας ενώ τα ανώτερα επίπεδα σέβονται μόνο τα ισχυρά/αδρά περιγράμματα δίνοντας έτσι μια υπο-κατάτμηση. Μεταβαίνοντας ανάμεσα στα επίπεδα αυτά παίρνει κανείς μια συνεχή συναλλαγή (trade-off) μεταξύ αυτών των άκρων και η αλλαγή αυτή αναπαράστασης «ελευθερώνει» μεταγενέστερα επεξεργαστικά στάδια ώστε να κάνουν χρήση πληροφορίας από πολλαπλά επίπεδα ή να επιλέξουν ένα επίπεδο με βάση επιπρόσθετη γνώση.

Η ιεραρχία της μεθόδου κατασκευάζεται με τη χρήση ενός greedy βασισμένου σε γράφημα αλγορίθμου «συγχώνευσης περιοχών». Ορίζεται ένα αρχικό γράφημα  $G=(P_0, K_0, W(K_0))$ , με κόμβους τις περιοχές  $P_0$ , συνδέσμους τα τόξα  $K_0$  που διαχωρίζουν διπλανές περιοχές ενώ τα βάρη  $W(K_0)$  είναι ένα μέτρο της ανομοιότητας μεταξύ περιοχών. Ο αλγόριθμος δουλεύει κατατάσσοντας τους συνδέσμους με βάση την ομοιότητα και συγχωνεύοντας αναδρομικά τις πιο όμοιες περιοχές.

Έτσι παράγεται ένα δένδρο από περιοχές στο οποίο τα «φύλλα» είναι τα αρχικά στοιχεία του  $P_0$ , η «ρίζα» είναι η συνολική εικόνα και οι περιοχές διατάσσονται με βάση την σχέση εγκλεισμού (ανήκειν).

Η ανομοιότητα μεταξύ δύο γειτονικών περιοχών ορίζεται σαν η μέση ισχύς του κοινού τους συνόρου στο  $K_0$  με τα βάρη  $W(K_0)$  που αρχικοποιούνται από τον OWT. Το βάρος ενός περιγράμματος που αφαιρείται δεν μπορεί να μειωθεί κατά τη διαδικασία της συγχώνευσης. Συνεπώς η περιοχή που κατασκευάζεται έχει τη δομή μιας δεικτονομημένης (indexed) ιεραρχίας οπότε μπορεί να αναπαρασταθεί από ένα δενδρόγραμμα όπου το ύψος κάθε περιοχής

είναι η τιμή της ανομοιότητας όταν πρωτοεμφανίζεται. Με άλλα λόγια

$$H(R) = W(C)$$

όπου  $C$  είναι το περίγραμμα του οποίου η αφαίρεση σχημάτισε την περιοχή  $R$ . Επιπλέον, η ιεραρχία αυτή επιτρέπει να οριστεί μια μετρική στο καρτεσιανό σύνολο  $P_0 \times P_0$  των περιοχών, δηλ. Μια απόσταση μεταξύ δύο περιοχών που ικανοποιεί την υπερμετρική:

Σαν αποτέλεσμα, ολόκληρη η ιεραρχία μπορεί να παρασταθεί σαν μια Υπερμετρική Απεικόνιση Περιγράμματος (UCM) με την πραγματικής-τιμής εικόνα να προκύπτει σταθμίζοντας κάθε σύνορο με την κλίμακα «εξαφάνισής» του. Το σχήμα 14 δίνει ένα παράδειγμα εφαρμογής της μεθόδου. Το αποτέλεσμα της UCM είναι μία εικόνα σταθμισμένων περιγραμμάτων με την αξιοσημείωτη ιδιότητα ότι παράγει ένα σύνολο κλειστών καμπυλών για κάθε κατώφλι. Αντίστροφα, μπορεί να ειπωθεί σαν μια βολική αναπαράσταση του δένδρου των περιοχών καθώς η κατάτμηση στην κλίμακα  $k$  μπορεί να ανακτηθεί εύκολα ορίζοντας ένα κατώφλι για την UCM στην κλίμακα αυτή. Αν η κλίμακα οριστεί ως η μέση ισχύς του περιγράμματος, τότε οι τιμές της UCM αντανakλούν την αντίθεση ανάμεσα σε γειτονικές περιοχές.

### 6.3.3 Αποτελέσματα

Τα καλλίτερα αποτελέσματα για τον αλγόριθμο OWT-UCM προκύπτουν εφαρμόζοντας τον ανιχνευτή  $gPb$  της προηγούμενης ενότητας, αν και ο αλγόριθμος μπορεί να χρησιμοποιήσει οποιαδήποτε πηγή περιγραμμάτων για το σήμα εισόδου  $E(x,y,\theta)$ . Τα σχήματα 15 και 16 απεικονίζουν αποτελέσματα της μεθόδου  $gPb$ - $owt$ - $ucm$  από το dataset BSDS500. Επειδή ο αλγόριθμος OWT-UCM παράγει ιεραρχικά δένδρα περιοχών για να πάρουμε μια μοναδική κατάτμηση θα πρέπει να επιλέξουμε μια συγκεκριμένη κλίμακα. Μια δυνατότητα είναι να επιλέξει κανείς ένα συγκεκριμένο κατώφλι για όλες τις εικόνες στο dataset, βαθμονομημένη (calibrated) ώστε να παρέχει βέλτιστη απόδοση στο training set. Αυτό αναφέρεται σαν optimal dataset scale (ODS). Η απόδοση αξιολογείται επίσης όταν το βέλτιστο κατώφλι επιλέγεται από ένα «μαντείο» (oracle) με μια βάση-ανά-εικόνα. Με μια τέτοια επιλογή Optimal Image Scale (ois) παίρνουμε μια ακόμα καλλίτερη κατάτμηση.

### 6.3.4 Αξιολόγηση