ΕΡΓΑΣΤΗΡΙΟ ΗΛΕΚΤΡΟΝΙΚΗΣ ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ & ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ ΔΗΜΟΚΡΙΤΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΡΑΚΗΣ

Ανάπτυξη και Υλοποίηση Ηλεκτρονικού Συστήματος Όρασης Βασισμένου στις Αρχές Λειτουργίας του Ανθρώπινου Οπτικού Συστήματος

> Διδακτορική Διατριβή BONIKAKH ΒΑΣΙΛΕΙΟΥ

Τριμελής Συμβουλευτική Επιτροπή Καθ. Ιωάννης Ανδρεάδης (Τμήμα Η.Μ. & Μ.Υ., Δ.Π.Θ.), Πρόεδρος Καθ. Νικόλαος Παπαμάρκος (Τμήμα Η.Μ. & Μ.Υ., Δ.Π.Θ.), Μέλος Αν. Καθ. Ιωάννης Θεοχάρης (Τμήμα Η.Μ. & Μ.Υ., Α.Π.Θ.), Μέλος

Ξάνθη, Μάρτιος 2008

ΕΡΓΑΣΤΗΡΙΟ ΗΛΕΚΤΡΟΝΙΚΗΣ ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ & ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ ΔΗΜΟΚΡΙΤΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΡΑΚΗΣ

Ανάπτυξη και Υλοποίηση Ηλεκτρονικού Συστήματος Όρασης Βασισμένου στις Αρχές Λειτουργίας του Ανθρώπινου Οπτικού Συστήματος

Διδακτορική Διατριβή Βονικάκη Βασιλείου

Τριμελής Συμβουλευτική Επιτροπή Καθ. Ιωάννης Ανδρεάδης (Τμήμα Η.Μ. & Μ.Υ., Δ.Π.Θ.), Πρόεδρος Καθ. Νικόλαος Παπαμάρκος (Τμήμα Η.Μ. & Μ.Υ., Δ.Π.Θ.), Μέλος Αν. Καθ. Ιωάννης Θεοχάρης (Τμήμα Η.Μ. & Μ.Υ., Δ.Π.Θ.), Μέλος

Ξάνθη, Μάρτιος 2008

Περίληψη & Συμβολή στην Επιστήμη

Τα προβλήματα, που πρέπει να αντιμετωπίσουν τα τεχνητά και τα βιολογικά συστήματα όρασης, είναι στη βάση τους ίδια. Και τα δύο πρέπει να αντιμετωπίσουν το πεπερασμένο δυναμικό εύρος των αισθητήρων τους σε σχέση με το ευρύ φάσμα φωτεινών εντάσεων που υπάρχουν στη φύση. Και τα δύο πρέπει να ανιχνεύσουν και να κωδικοποιήσουν ακμές, χρώματα, κίνηση, υφή και βάθος. Πρέπει να επεξεργαστούν έναν τεράστιο όγκο πληροφοριών σε πραγματικό χρόνο και να εκτελέσουν πολλαπλές ταξινομήσεις χαρακτηριστικών. Όλα τα παραπάνω αποτελούν ενεργά πεδία έρευνας στον τομέα της όρασης μηχανών. Εντούτοις, αυτά έχουν ήδη αντιμετωπιστεί από τη φύση μέσω της διαδικασίας της εξέλιξης. Έχοντας ήδη ένα σύστημα το οποίο λειτουργεί σχεδόν ιδανικά, η μελέτη των ιδιοτήτων και των χαρακτηριστικών του μπορεί να οδηγήσει σε πιθανές λύσεις για πολλά από τα προβλήματα με τα οποία πραγματεύεται η σημερινή επιστήμη. Βασικός στόχος της παρούσας διατριβής είναι η διερεύνηση νέων τεχνικών στο πεδίο της όρασης μηχανών, οι οποίες έχουν εμπνευστεί από το οπτικό σύστημα των πρωτευόντων θηλαστικών και ειδικότερα από το Ανθρώπινο Οπτικό Σύστημα (ΑΟΣ).

Το πρώτο πρόβλημα, το οποίο διερευνάται στην παρούσα διατριβή, είναι η εξαγωγή των εξεχόντων περιγραμμάτων. Οι ακμές μιας εικόνας αποτελούν πολύ χρήσιμο στοιχείο στην ανάλυση και την επεξεργασία της. Ωστόσο, ο μεγάλος αριθμός τους μπορεί να αποτελέσει σημαντικό πρόβλημα σε ένα αυτόματο σύστημα. Αντίθετα, τα βιολογικά οπτικά συστήματα μπορούν να ξεχωρίσουν τις εξέχουσες ακμές σε μια σκηνή και να εξάγουν αβίαστα το αρχικό σχεδιάγραμμά της. Στο πλαίσιο αυτό, παρουσιάζεται μια νέα μέθοδος για την εξαγωγή των εξεχόντων περιγραμμάτων, βασισμένη στις διασυνδέσεις των νευρώνων κλίσης του πρωτοταγούς οπτικού φλοιού. Στα πλαίσια αυτής της μεθόδου σχεδιάστηκε ένα νέο σύνολο μασκών για τη μοντελοποίηση των νευρώνων κλίσης, που βελτιώνει κατά 20 φορές τη μέχρι τώρα υπάρχουσα προσέγγιση, η οποία βασίζεται στη συνέλιξη με φίλτρα Gabor. Επίσης, ο προτεινόμενος τρόπος υπολογισμού της κλίσης των ακμών ευνοεί την παραλληλία και ταιριάζει περισσότερο στο διακριτό πλέγμα υπερστηλών του πρωτοταγούς οπτικού φλοιού. Στα πλαίνος υπολογισμού της κλίσης των ακμών ευνοεί την παραλληλία και ταιριάζει περισσότερο στο διακριτό πλέγμα υπεροτηλών του πρωτοταγούς οπο το τροτεινόμενος τρόπος υπολογισμού της κλίσης των ακμών εριγραφιά της παρουσιάζεται ένας καινούριος περιγραφέας απλών σχημάτων βασισμένος στο προτεινόμενο σύνολο μασκών, ο οποίος παρουσιάζεται και χαιρουσιάζεται ένας καινούριος περιγραφέας απλών σχημάτων βασισμένος στο προτεινόμενος σύνολο μασκών, ο οποίος παρουσιάζει ανοχή στην περιστροφή και στην κλιμάκωση, ενώ επιδεικνύει ανοχή σε στρεβλώσεις και

παραμορφώσεις σχημάτων. Εισάγεται, επίσης, μια νέα μορφή νευρωνικού δικτύου, που δέχεται ως εισόδους τα αποτελέσματα των προτεινόμενων μασκών κλίσης για κάθε σημείο της εικόνας, ενώ με την πάροδο των επαναλήψεων αυξάνει τις τιμές εκείνων των μασκών που κωδικοποιούν εξέχοντα περιγράμματα. Οι κανόνες αλληλεπίδρασης των μασκών εμπεριέχουν τους κανόνες Gestalt για την ομαλή συνέχεια και την εγγύτητα και στοχεύουν στο σχηματισμό καμπυλών εκθετικής μορφής. Η προσέγγιση αυτή διαπιστώθηκε ότι δίνει καλύτερα αποτελέσματα από τη μέχρι τώρα προσέγγιση, η οποία βασίζεται στο σχηματισμό καμπυλών κυκλικής μορφής. Παρουσιάζεται στη συνέχεια μια καινούρια συνάρτηση διέγερσης για τους νευρώνες του δικτύου, η οποία στηρίζεται στην ομοιότητα της διέγερσης των δύο λοβών τους. Αυτό σημαίνει ότι οι νευρώνες που δέχονται το ίδιο ποσοστό διέγερσης στο δεξιό και στο αριστερό τμήμα της ακμής που κωδικοποιούν, παρουσιάζουν εντονότερη διέγερση. Χαρακτηριστικό της νέας αυτής συνάρτησης είναι ότι ευνοεί τη σωστή συνέχεια των εξεχόντων περιγραμμάτων. Το προτεινόμενο δίκτυο έχει τη δυνατότητα, επίσης, να τονίζει συγκεκριμένους βαθμούς καμπυλότητας. Με τον τρόπο αυτό, μπορεί να χρησιμοποιηθεί για το φιλτράρισμα των εξεχόντων περιγραμμάτων, ώστε να αναδεικνύεται μια μόνο κατηγορία, όπως ευθείες ή κύκλοι. Τέλος, η προτεινόμενη μέθοδος παρουσιάζει ικανοποιητικά αποτελέσματα για μια ευρεία γκάμα πραγματικών και συνθετικών εικόνων, σε συνδυασμό πάντα με τους συντομότερους χρόνους εκτέλεσης που έχουν αναφερθεί στην υπάρχουσα βιβλιογραφία.

Το επόμενο πεδίο, που πραγματεύεται η παρούσα διατριβή, είναι το πρόβλημα του περιορισμένου δυναμικού εύρους των σημερινών αισθητήρων εικόνων, το οποίο οδηγεί πολλές φορές στην υποβάθμιση της ποιότητας των καταγεγραμμένων εικόνων. Στις περιπτώσεις αυτές το δυναμικό εύρος του αισθητήρα της κάμερας είναι περιορισμένο σε σχέση με το δυναμικό εύρος της σκηνής που καλείται να καταγράψει. Ως αποτέλεσμα, η καταγεγραμμένη εικόνα παρουσιάζει υποεκτεθειμένα σημεία, τα οποία είναι πολύ σκοτεινά και στα οποία δε διακρίνονται λεπτομέρειες, είτε υπερεκτεθειμένα σημεία, τα οποία είναι πολύ φωτεινά και στα οποία, επίσης, οι λεπτομέρειες δεν είναι διακριτές. Στο πλαίσιο αυτό προτείνεται μια νέα μέθοδος για τη βελτίωση εκείνων των περιοχών των εικόνων που έχουν υποστεί υποέκθεση ή υπερέκθεση. Κύριος στόχος της μεθόδου είναι να βελτιώσει τα τμήματα που παρουσιάζουν το συγκεκριμένο πρόβλημα χωρίς να επηρεάζει τις σωστά εκτεθειμένες περιοχές. Σαν τμήμα της νέας μεθόδου παρουσιάζεται, επίσης, μια νέα συνάρτηση απεικόνισης αποχρώσεων, βασισμένη στη διακλαδισμένη αναστολή των δικτύων ανταγωνισμού κέντρου-περιφέρειας του ΑΟΣ. Η συνάρτηση αυτή χρησιμοποιεί ρυθμιζόμενες μη γραμμικότητες, οι οποίες μεταβάλλουν την τιμή έντασης κάθε εικονοστοιχείου βάσει της ευρύτερης περιφέρειάς του και των γενικότερων χαρακτηριστικών της εικόνας. Παρουσιάζεται ακόμη ένας νέος τρόπος για τον υπολογισμό της περιφέρειας κάθε εικονοστοιχείου, βασιζόμενος σε μια περιοχή η οποία προσαρμόζει το σχήμα της στις τοπικές ακμές της εικόνας. Η προσαρμοστικότητα του σχήματος αυτής της περιοχής βασίζεται στα κύτταρα κλίσης του πρωτοταγούς οπτικού φλοιού. Η προτεινόμενη μέθοδος παρουσιάζει βελτιωμένα αποτελέσματα σε σχέση με τις υπάρχουσες μεθόδους της βιβλιογραφίας. Πιο συγκεκριμένα, δεν εξάγονται φαινόμενα άλω (γενική της λέξης «άλως») στις ισχυρές μεταβάσεις εντάσεως. Επιπλέον, οι σωστά εκτεθειμένες περιοχές δεν επηρεάζονται, ενώ βελτιώνονται όχι μόνο οι υποεκτεθειμένες περιοχές, αλλά ταυτόχρονα και οι υπερεκτεθειμένες. Ο χρόνος εκτέλεσης της

προτεινόμενης μεθόδου είναι τουλάχιστο μια τάξη μεγέθους μικρότερος από τον αντίστοιχο των άλλων μεθόδων. Τέλος, παρουσιάζονται τα αποτελέσματα σχετικής ψυχοφυσιολογικής μελέτης σε ανθρώπινους παρατηρητές, τα οποία ανέδειξαν τα αποτελέσματα της προτεινόμενης μεθόδου ως τα πιο «φυσικά» και εκείνα που «αναδεικνύουν τις περισσότερες λεπτομέρειες στις φωτεινές και στις σκοτεινές περιοχές» σε σχέση με τα υπόλοιπα. Το πεδίο εφαρμογής της προτεινόμενης μεθόδου εκτείνεται από την όραση μηχανών έως και τη βελτίωση φωτογραφιών για αισθητικούς λόγους.

Η δυαδικοποίηση εγγράφων είναι ένα άλλο πεδίο το οποίο πραγματεύεται η παρούσα διατριβή. Το πρόβλημα του αυτόματου διαχωρισμού των χαρακτήρων ενός ψηφιακού κειμένου από το φόντο παίζει σημαντικό ρόλο στα αποτελέσματα της οπτικής αναγνώρισης χαρακτήρων. Παρουσιάζεται, λοιπόν, μια νέα μέθοδος δυαδικοποίησης εγγράφων, η οποία υιοθετεί τους βιολογικούς εκείνους μηχανισμούς, που είναι υπεύθυνοι για την αντίληψη του σκοτεινού και του φωτεινού. Οι βιολογικοί αυτοί μηχανισμοί είναι τα γαγγλιακά κύτταρα σκοτεινού κέντρου του ΑΟΣ. Η προτεινόμενη μέθοδος χρησιμοποιεί δομές ανταγωνισμού κέντρου-περιφέρειας ανάλογες με τα γαγγλιακά κύτταρα αυτού του τύπου. Εισάγεται έτσι μια νέα συνάρτηση απόκρισης των προτεινόμενων κυττάρων, η οποία αυξάνει την ευαισθησία τους στις σκοτεινές περιοχές, μειώνοντας τις επιπτώσεις των σκιάσεων και των λεκέδων. Η εφαρμογή αυτών των κυττάρων μετατρέπει την αρχική εικόνα σε μια ενδιάμεση μορφή, ικανή να δυαδικοποιηθεί από τις κλασικές καθολικές μεθόδους δυαδικοποίησης. Παράλληλα, υιοθετείται ο συνδυασμός διαφορετικών χωρικών κλιμάκων επεξεργασίας, με τη χρήση κυττάρων δύο διαφορετικών μεγεθών, τα οποία επεξεργάζονται πληροφορίες διαφορετικής συχνότητας το καθένα. Η προτεινόμενη μέθοδος υπερτερεί των ήδη υπαρχόντων μεθόδων της βιβλιογραφίας, ειδικότερα για έγγραφα με μεγάλο βαθμό υποβάθμισης. Αυτό συμβαίνει γιατί παρουσιάζει μεγαλύτερη ανοχή στο θόρυβο, στις σκιάσεις και στη χαμηλή αντίθεση. Για το λόγο αυτό μπορεί να χρησιμοποιηθεί ιδιαίτερα σε ιστορικά έγγραφα και έγγραφα τα οποία ψηφιοποιήθηκαν με φωτογραφική μηχανή.

Τέλος, παρουσιάζεται η υλοποίηση σε FPGA της μεθόδου βελτίωσης αποχρώσεων. Η προτεινόμενη υλοποίηση έχει τη δυνατότητα για επεξεργασία σήματος βίντεο μεγέθους μέχρι 2,5 εκατομμυρίων εικονοστοιχείων σε πραγματικό χρόνο (25 πλαίσια ανά δευτερόλεπτο), χαρακτηριστικά που είναι αδύνατο να επιτευχθούν από το λογισμικό. Η συσκευή που χρησιμοποιήθηκε ανήκει στην κατηγορία Stratix II GX της ALTERA. Οι επιδόσεις του κυκλώματος υπερτερούν σε σχέση με τις υπάρχουσες υλοποιήσεις ανάλογων αλγορίθμων, αφού οι αλγόριθμοι αυτοί επεξεργάζονται ασπρόμαυρες εικόνες σε μια μόνο χωρική κλίμακα και με ανάλυση 200×200 εικονοστοιχεία. Αντίθετα, η προτεινόμενη μέθοδος, εκτός από τη μεγαλύτερη ανάλυση, επεξεργάζεται έγχρωμες εικόνες σε τρεις χωρικές κλίμακες.

Εν κατακλείδι, η παρούσα διδακτορική διατριβή παρουσιάζει λύσεις σε υπάρχοντα προβλήματα όρασης μηχανών, όπως η τμηματοποίηση περιγραμμάτων σε μη ομοιόμορφο φόντο, η καταγραφή και επεξεργασία σκηνών με μεγάλο δυναμικό εύρος σε πραγματικό χρόνο και η δυαδικοποίηση εγγράφων. Οι λύσεις αυτές προέρχονται από το πεδίο των βιολογικών οπτικών συστημάτων και μπορούν να αποτελέσουν νέες κατευθύνσεις στον

τρόπο αντιμετώπισης και άλλων κλασικών προβλημάτων που υπάρχουν στο πεδίο όρασης μηχανών.

Summary & Contribution to the state of the art

Both biological and artificial visual systems have to face the same problems on their basis. They both have to compensate for the limited dynamic range of their sensors, compared to the immense range of intensities that are present in the natural scenes. They both have to extract and encode edges, colors, motion, texture and depth. They have to process a huge amount of visual information in real-time and carry out multiple feature classifications. All the above problems are active research fields of machine vision. Yet, they have already been solved by nature, through the process of evolution. Having already a visual system that functions almost ideally, analyzing its characteristics can shed light to many of the contemporary scientific problems. The primary objective of this dissertation is the development of new techniques in the fields of machine vision and image processing, which have been inspired by the characteristics of the Human Visual System (HVS), and generally, the visual system of the primates.

The first problem, which is investigated in this dissertation, is the extraction of salient contours. The edges of an image can be a valuable tool in its analysis and processing. Nevertheless, their vast number can pose a problem to an automatic system. Moreover, biological visual systems can easily distinguish between the most important contours in a scene and create its primal sketch. In this context, a new method is presented for the extraction of salient contours, based on the non-classical receptive field of the simple cells of the primary visual cortex. As part of this method, a new set of binary kernels was designed for the implementation of the simple orientation cells. This set improves by a factor of 20 the usual approach, which is based in the convolution of the image with Gabor filters. Additionally, the proposed implementation of the oriented cells favors parallelism and is more compatible with the discrete lattice of positions and orientations of the hypercolumns, found in the primary visual cortex. A new shape descriptor, based on the proposed kernel set is also introduced. This shape descriptor exhibits scale and rotation invariance and can handle distortions in the contours of the shapes. Moreover, a new type of neural network is introduced having as inputs the results of the proposed oriented kernels, for every image region. With every iteration the network increases the values of the kernels which encode salient contours. The network incorporates the Gestalt laws of good continuity and proximity, by affiliating the orientation kernels according to their coexponentiality. This approach was found to exhibit better results than the usual co-circular approach found in the literature. A new affinity function, based in the similarity between the two lobes of the oriented neurons, is also introduced. This means that neurons with the same excitation degree in both their left and right part increase more their output compared to the other neurons. As a result, better continuity in the extracted contours is achieved. The proposed neural network has also the ability to enhance a certain degree of curvature. This can be used for filtering the extracted salient contours, in order to output only one type of contour (e.g. circles, straight lines etch.). Finally, the proposed method exhibits satisfactory results in a great variety of real and synthetic images, as well as, the fastest execution times reported until now in the literature.

The present dissertation also addresses the problem of image quality degradation, caused by the limited dynamic range of the image sensor. In these cases, the dynamic range of the sensor is lower than the dynamic range of the scene, causing either under-exposed or overexposed regions to the image and, thus, reducing the visible information. In this context, a new method is proposed for the enhancement of images with under or over-exposed regions. A new contrast modification function is introduced, derived by the shunting inhibition characteristics of the center-surround cells of the HVS. This function employs adjustable non-linearities, which vary according to the pixels' surrounding neighborhood and the global image statistics. Additionally, a novel way for the computation of the surrounding neighborhood is introduced, which is based in a region that adapts its shape to the sharp intensity transitions of the image. The adaptability of this region is achieved through orientation-like kernels, similar to the orientation neurons of the V1 region. The proposed approach exhibits better results from other reported methods. Particularly, no halo effects are extracted in the sharp intensity transitions, the correctly exposed regions are not affected, while both under and over-exposed regions are enhanced. The execution time of the proposed method is at least an order of magnitude lesser than the one of the existing methods. Finally, a psychophysical experiment with human observers is carried out, which shows that the results of the proposed method are ranked as being more "natural" compared to the ones of the existing methods and that they reveal more details in the light and dark image regions. The proposed method can be applied to machine vision applications and photography enhancement.

Document binarization is another research field which the proposed method addresses. The automatic segmentation of the document characters and its background plays a crucial role in optical character recognition systems. In those contexts, a novel document binarization method is introduced, which employs the basic biological mechanisms that subserve the perception of brightness and darkness. These biological mechanisms are the OFF-center ganglion cells of the retina. The proposed method uses similar center-surround architectures. A new response function is also introduced, that allows the artificial cells to increase their sensitivity in the dark image regions, decreasing the results of shadows or smears. The processing by the proposed OFF-center cells transforms the original image to an intermediate state, which can be successfully binarized by the classical global binarization methods (e.g. Otsu's method). Additionally, two different scales are employed in this processing; each one specialized for different frequencies. The proposed method

outperforms all the existing document binarization methods, since it exhibits superior response to noise, shadows and low contrast. Consequently, it can be successfully used for the binarization of degraded historical documents, as well as for documents that were captured by digital still cameras.

Finally, an FPGA implementation of the image enhancement algorithm is introduced. The proposed implementation can process images of sizes up to 2.5 MPixels in real time (25 fps), which otherwise, is impossible by the software implementation. The FPGA device used belongs to the Stratix II GX family from ALTERA. The performance of the proposed system outperforms other reported systems that implement similar algorithms, since these implementations can process grayscale images of 200×200 pixels in one spatial scale, while the proposed circuit processes color images in three spatial scales.

Concluding, the present dissertation presents solutions in contemporary problems of machine vision, which are all inspired by the field of biological visual systems and may introduce new directions to other similar problems in this field.

Πρόλογος

Ο κλάδος της επιστήμης που ασχολείται με την όραση μηχανών, την επεξεργασία εικόνας και γενικότερα με την επεξεργασία δισδιάστατου σήματος, έχει ήδη συμπληρώσει μισό περίπου αιώνα ύπαρξης και συνεχίζει να αναπτύσσεται ραγδαία. Καινούριες τεχνικές παρουσιάζονται καθημερινά, ο αριθμός των ερευνητών που βρίσκονται ενεργά στο πεδίο συνεχώς μεγαλώνει, ενώ η υπολογιστική ισχύς, το βασικότερο εργαλείο του ερευνητικού αυτού πεδίου, εξακολουθεί να διπλασιάζεται κάθε δύο χρόνια σύμφωνα με το νόμο του Moore. Εντούτοις, ακόμη και υπό αυτές τις ομολογουμένως ευνοϊκές συνθήκες, το Ανθρώπινο Οπτικό Σύστημα (ΑΟΣ) εξακολουθεί να υπερτερεί μακράν όλων σχεδόν των ηλεκτρονικών οπτικών συστημάτων και των αλγορίθμων που έχουν αναπτυχθεί μέχρι σήμερα. Για το λόγο αυτό, τα τελευταία χρόνια έχει παρατηρηθεί μια έντονη στροφή προς τεχνικές, οι οποίες είναι εμπνευσμένες στη βάση τους από τα χαρακτηριστικά του ΑΟΣ. Η κατανόηση της λειτουργίας και της δομής του ΑΟΣ είναι πεδίο έρευνας της οπτικής νευροεπιστήμης (visual neuroscience). Η πρόοδος σε αυτόν τον τομέα της επιστήμης είναι εξαιρετικά ραγδαία. Καθημερινά ανακαλύπτονται όλο και περισσότερα κομμάτια του «παζλ», συμπληρώνοντας συνεχώς την εικόνα που έχουμε για τη λειτουργία του ανθρώπινου εγκεφάλου και του ΑΟΣ. Οι δύο αυτοί κλάδοι της επιστήμης, μπορούν να επωφεληθούν από τη μεταξύ τους αλληλεπίδραση. Ειδικότερα, για το πεδίο της όρασης μηχανών και της επεξεργασίας εικόνας, υπάρχει πλήθος βιολογικών ευρημάτων τα οποία κρύβουν απαντήσεις σε πολλά από τα βασικότερα προβλήματα, που προσπαθεί να λύσει ο συγκεκριμένος τομέας της επιστήμης. Εντούτοις, τις περισσότερες φορές αυτά μένουν αχρησιμοποίητα, λόγω της μη χαρτογράφησης του αντίστοιχου βιολογικού ερευνητικού πεδίου από τους επιστήμονες στο χώρο της όρασης μηχανών. Στόχος της παρούσας διδακτορικής διατριβής είναι να χρησιμοποιήσει ορισμένα από τα ευρήματά του πεδίου της οπτικής νευροεπιστήμης, παρουσιάζοντας νέες μεθόδους για την αντιμετώπιση μερικών κλασικών προβλημάτων της όρασης μηχανών και της επεξεργασίας εικόνας.

Το **πρώτο κεφάλαιο** είναι μια εισαγωγή στις ομοιότητες και τις διαφορές που υπάρχουν μεταξύ των ηλεκτρονικών οπτικών συστημάτων και του ΑΟΣ. Παρουσιάζεται αρχικά μια απαραίτητη διασαφήνιση των βασικών εννοιών, οι οποίες χρησιμοποιούνται στα πεδία της όρασης μηχανών και της οπτικής νευροεπιστήμης και οι οποίες πολλές φορές παρερμηνεύονται ή χρησιμοποιούνται με λανθασμένο τρόπο. Στη συνέχεια, ακολουθεί μια σύντομη παρουσίαση των βασικών τμημάτων του ΑΟΣ, η οποία όμως εστιάζει στις βασικές

διαφορές του με τους αλγορίθμους επεξεργασίας εικόνας και τα ηλεκτρονικά οπτικά συστήματα, με κύριο στόχο να γίνει κατανοητό στο σύνολο των μηχανικών.

Ένα από τα βασικότερα προβλήματα στην κωδικοποίηση σχήματος, την τμηματοποίηση και την ανάλυση εικόνων είναι το μη ομοιόμορφο φόντο που αναπόφευκτα υπάρχει στις πραγματικές συνθήκες λειτουργίας των οπτικών συστημάτων. Επιφάνειες με έντονη υφή οδηγούν στην εξαγωγή ακμών που δεν ανήκουν στο κεντρικό αντικείμενο του ενδιαφέροντος. Στο **δεύτερο κεφάλαιο** παρουσιάζεται μια νέα μέθοδος για την εξαγωγή των περιγραμμάτων που τραβούν την προσοχή του παρατηρητή, ανεξάρτητα από το φόντο στο οποίο βρίσκεται το αντικείμενο από το οποίο προέρχονται. Η μέθοδος αυτή είναι εμπνευσμένη από τις αλληλεπιδράσεις των νευρώνων κλίσης του πρωτοταγούς οπτικού φλοιού.

Σημαντικό είναι, επίσης, και το πρόβλημα που εισάγει το περιορισμένο δυναμικό εύρος του αισθητήρα της κάμερας, σε σχέση με το δυναμικό εύρος της σκηνής που θέλουμε να καταγραφεί. Ως αποτέλεσμα αυτού, πολλές από τις εικόνες που καταγράφονται από την κάμερα έχουν σημαντικές διαφορές στην ποιότητα των λεπτομερειών, σε σχέση με αυτό που αντιλαμβάνεται ο άνθρωπος. Αυτό μπορεί να υποβαθμίσει την ποιότητα των αποτελεσμάτων μετέπειτα αλγορίθμων που ακολουθούν στη ροή επεξεργασίας. Στο **τρίτο κεφάλαιο** παρουσιάζεται μια νέα μέθοδος για βελτίωση των αποχρώσεων εικόνων, οι οποίες παρουσιάζουν προβλήματα λόγω του περιορισμένου δυναμικού εύρους του αισθητήρα της κάμερας. Η μέθοδος αυτή είναι εμπνευσμένη από την επεξεργασία που πραγματοποιείται στον αμφιβληστροειδή χιτώνα του ανθρώπινου ματιού και πιο συγκεκριμένα, από τις ιδιότητες των γαγγλιακών κυττάρων.

Η αντίληψη του φωτεινού και του σκοτεινού από τα νευρικά κυκλώματα του ΑΟΣ, χρησιμοποιείται στο **τέταρτο κεφάλαιο** για τη δυαδικοποίηση εγγράφων με υποβαθμισμένη ποιότητα. Η δυαδικοποίηση εγγράφων είναι μια σημαντική διεργασία η οποία επηρεάζει τα αποτελέσματα της οπτικής αναγνώρισης χαρακτήρων. Η νέα μέθοδος που παρουσιάζεται, εκμεταλλεύεται τα χαρακτηριστικά των γαγγλιακών κυττάρων σκοτεινού κέντρου, τα οποία σχετίζονται με την αντίληψη του σκοτεινού και του φωτεινού.

Τέλος, στο πέμπτο κεφάλαιο παρουσιάζεται η υλοποίηση σε FPGA του αλγορίθμου βελτίωσης αποχρώσεων του κεφαλαίου τρία. Η υλοποίηση αυτή έχει τη δυνατότητα να επεξεργάζεται σε πραγματικό χρόνο εικόνες βίντεο, οι οποίες είναι αδύνατο να βελτιωθούν στους απαιτούμενους χρονικούς περιορισμούς από το λογισμικό.

Σε αυτό το σημείο, θα ήθελα να ευχαριστήσω θερμά όλους όσοι βοήθησαν κατά τη διάρκεια εκπόνησης της παρούσας διδακτορικής διατριβής. Περισσότερο από όλους θα ήθελα να ευχαριστήσω τον πρόεδρο της τριμελούς συμβουλευτικής επιτροπής **Καθ.** Ιωάννη Ανδρεάδη, ο οποίος με το μεγάλο του ενδιαφέρον και την καταλυτική του καθοδήγηση, συνέβαλε αποφασιστικά στην περάτωση αυτής της διατριβής. Η εποικοδομητική του κριτική βοήθησε σημαντικά στην κατανόηση από μέρους μου της επιστημονικής μεθόδου και συντέλεσε στη βελτίωση της ποιότητας της διατριβής. Επίσης, θα ήθελα να τον ευχαριστήσω για τη συνεχή προσπάθειά του να μου εξασφαλίσει οικονομικούς πόρους καθ' όλη τη διάρκεια των μεταπτυχιακών μου σπουδών, γεγονός που μου επέτρεψε να αφοσιωθώ πλήρως σε αυτές. Τέλος, θα ήθελα να τον ευχαριστήσω για την άψογη συνεργασία κατά τη διάρκεια όλων των χρόνων διεξαγωγής της διατριβής αυτής.

Ακόμη, θα ήθελα να ευχαριστήσω τον **Καθ. Νικόλαο Παπαμάρκο**, ο οποίος ως μέλος της τριμελούς συμβουλευτικής επιτροπής, έδειξε ιδιαίτερο ενδιαφέρον σε όλη τη διάρκεια εκπόνησης αυτής της διατριβής, ενώ η βοήθειά του σε συγκεκριμένες ερευνητικές περιοχές ήταν καταλυτική.

Επιπλέον, θα ήθελα να ευχαριστήσω τον **Αν. Καθ. Ιωάννη Θεοχάρη** για τη συμμετοχή του στην τριμελή συμβουλευτική επιτροπή και την άμεση βοήθειά του στην επίλυση διαφόρων γραφειοκρατικών θεμάτων που ανέκυψαν στη διάρκεια εκπόνησης της διατριβής.

Επίσης, θα ήθελα να ευχαριστήσω τον **Επ. Καθ. Αντώνιο Γαστεράτο** για την επιμέλεια της ερευνητική πρόταση ΠΕΝΕΔ με τίτλο «Ανάπτυξη και Υλοποίηση Νέων Αλγορίθμων Αναγνώρισης Προτύπων Βασισμένων σε Βιολογικά Εμπνευσμένα Μοντέλα και σε Ευφυή Συστήματα», η οποία μου εξασφάλισε επιπλέον οικονομική άνεση για την απρόσκοπτη εκπόνηση της παρούσας διατριβής.

Σημαντική ήταν επίσης η βοήθεια του υποψήφιου διδάκτορα **Σάββα Χατζηχριστοφή**, στη μετατροπή του κώδικα της διατριβής σε visual περιβάλλον, τον οποίο και θα ήθελα να ευχαριστήσω.

Ακόμη, θα ήθελα να ευχαριστήσω τη **Δρ. Μηχανικό Περιβάλλοντος κ. Ελένη Βαϊοπούλου** για τη σημαντική βοήθειά της στη διόρθωση των κειμένων της παρούσας διατριβής. Περισσότερο όμως θα ήθελα να την ευχαριστήσω για το μεγάλο της ενδιαφέρον, τη συμπαράστασή της, αλλά και την αγάπη που με περιέβαλε από την έναρξη ως την περάτωση αυτής της διατριβής.

Τέλος, θα ήθελα να ευχαριστήσω **τους γονείς μου, Γεώργιο και Ελένη,** που με στήριξαν με κάθε τρόπο σε όλη τη διάρκεια των μεταπτυχιακών μου σπουδών.

Περιεχόμενα

1. Βιολογικά & Ηλεκτρονικά Οπτικά Συστήματα	3
1.1 Διασαφήνιση Βασικών Εννοιών	4
1.1.1 Μη χρωματικά μεγέθη	5
1.1.2 Χρώμα	10
1.1.2.1 Το παράδοξο της έγχρωμης όρασης	10
1.1.2.2 Το πείραμα του Edwin Land	12
1.2 Τεχνητά συστήματα όραση	15
1.3 Το Ανθρώπινο Οπτικό Σύστημα	16
1.3.1 Ροή οπτικής πληροφορίας	16
1.3.2 Ο οφθαλμός	17
1.3.3 Ο αμφιβληστροειδής χιτώνας	18
1.3.3.1 Φωτοϋποδοχείς	19
1.3.3.2 Γαγγλιακά κύτταρα	21
1.3.4 Ο πρωτοταγής οπτικός φλοιός (περιοχή V1)	26
1.3.4.1 Υπερστήλες (Hypercolumns)	26
1.3.4.2 Απλά κύτταρα κλίσης (simple cells)	27
1.3.4.3 Σύνθετα κύτταρα (complex cells)	28
1.3.5 Ανώτερες οπτικές περιοχές	29
1.4 Συμπεράσματα	

2. Εξέχοντα Περιγράμματα	
2.1 Διατύπωση του προβλήματος	34
2.2 Νευροφυσιολογικά στοιχεία του ΑΟΣ	34
2.3 Ανασκόπηση υπάρχουσας βιβλιογραφίας	37
2.4 Μοντελοποίηση κυττάρων κλίσης	40
2.4.1 Υπάρχουσα μοντελοποίηση κυττάρων κλίσης	40
2.4.2 Προτεινόμενη μοντελοποίηση κυττάρων κλίσης	42
2.5 Περιγραφή και αναγνώριση απλών σχημάτων με μάσκες κλίσης	50
2.5.1 Περιγραφή της μεθόδου	50
2.5.2 Παράδειγμα ταξινόμησης	52
2.6 Μέθοδος εξαγωγής εξεχόντων περιγραμμάτων	54
2.6.1 Δομή του δικτύου	54
2.6.2 Κανόνες διέγερσης	59
2.6.2.1 Αναστολή	59
2.6.2.2 Διέγερση των δύο λοβών της μάσκας	60
2.6.2.3 Συνολική διέγερση μάσκας	60
2.6.2.4 Έξοδος μάσκας	61
2.6.2.5 Εσωτερικός ανταγωνισμός υπερστήλης	62
2.6.2.6 Προσαρμοσμένη διαρροή δυναμικού	62
2.6.2.7 Τελική έξοδος του δικτύου	63
2.7 Πειραματικά αποτελέσματα – συγκρίσεις	65
2.7.1 Αξιολόγηση απόδοσης	66
2.7.2 Αποτελέσματα για πραγματικές εικόνες	67
2.7.3 Αποτελέσματα για συνθετικές εικόνες	68
2.7.4 Τονισμός συγκεκριμένων καμπυλοτήτων	69
2.7.5 Εκθετικό και κυκλικό μοτίβο διασύνδεσης	69
2.7.6 Σύγκριση με υπάρχουσες μεθόδους	70
2.8 Συμπεράσματα	76

3. Βελτιστοποίηση αποχρώσεων7			
3.1 Διατύπωση του προβλήματος	80		
3.1.1 Χαρακτηριστικά CCD αισθητήρων	80		

3.1.2 Υποέκθεση & Υπερέκθεση	81
3.1.3 Σκηνές Μεγάλου Δυναμικού Εύρους	82
3.2 Ανασκόπηση Υπάρχουσας Βιβλιογραφίας	83
3.2.1 Εξοπλισμός Μεγάλου Δυναμικού Εύρους	
3.2.2 Πολλαπλή Έκθεση Σκηνών	84
3.2.3 Βελτίωση Εικόνων από κάμερες ΣΔΕ	85
3.2.3.1 Αλγόριθμοι Retinex	85
3.2.3.2 Retinex Πολλαπλής Κλίμακας	86
3.2.3.3 Ο αλγόριθμος ACE	87
3.2.3.4 Διόρθωση γ	
3.2.3.5 Λογαριθμική κωδικοποίηση	90
3.2.3.6 Εξισορρόπηση ιστογράμματος (Histogram Equalization)	91
3.2.4 Αδυναμίες υπαρχόντων λύσεων	92
3.3 Η λύση του ΑΟΣ	95
3.4 Προτεινόμενη μέθοδος	
3.4.1 Υιοθέτηση στοιχείων του ΑΟΣ	
3.4.2 Αναλυτική περιγραφή της μεθόδου	
3.4.2.1 Μετατροπή χρωματικών χώρων	
3.4.2.2 Κλιμάκωση (0-255)	
3.4.2.3 Απεικόνιση αποχρώσεων	
3.4.2.4 Υπολογισμός της περιφέρειας ${m {\cal S}}_{ij}$	
3.4.2.5 Υπολογισμός Παραμέτρων	
3.5 Συγκριτικά αποτελέσματα	114
3.5.1 Αποτελέσματα σε εικόνες με υποέκθεση και υπερέκθεση	
3.5.2 Αποτελέσματα για άλλες εικόνες	124
3.5.3 Ψυχοφυσιολογική σύγκριση	
3.5.4 Πολυπλοκότητα και χρόνος εκτέλεσης	
3.6 Συμπεράσματα	

4. Δυαδικοποίηση Εγγράφων	133
4.1 Διατύπωση του προβλήματος	
4.2 Ανασκόπηση υπάρχουσας βιβλιογραφίας	135
4.2.1 Καθολικές μέθοδοι δυαδικοποίησης	135
4.2.2 Τοπικές μέθοδοι δυαδικοποίησης	136

4.2.2.1 Μέθοδος Niblack	137
4.2.2.2 Μέθοδος Sauvola	137
4.2.2.3 Μέθοδος προσαρμοσμένης λογικής (ALM)	
4.2.2.4 Μέθοδος Gatos	137
4.2.3 Αδυναμίες των υπαρχόντων μεθόδων	
4.3 Ανίχνευση κειμένου από το ΑΟΣ	139
4.4 Περιγραφή της προτεινόμενης μεθόδου	143
4.4.1 Προεπεξεργασία	143
4.4.2 Τεχνητά γαγγλιακά κύτταρα σκοτεινού κέντρου	
4.4.2.1 Υποδεκτικά πεδία	145
4.4.2.2 Συνάρτηση απόκρισης	146
4.4.2.3 Επιλογή παραμέτρων	149
4.4.3 Δυαδικοποίηση Otsu	151
4.4.4 Διορθωτική επεξεργασία	151
4.5 Πειραματικά αποτελέσματα	152
4.5.1 Παραγωγή των υποβαθμισμένων εικόνων	
4.5.2 Σύγκριση του PSNR	155
4.5.3 Σύγκριση αποτελεσμάτων OCR	159
4.5.4 Ποιοτική σύγκριση αποτελεσμάτων	160
4.5.5 Συνολική συμπεριφορά των μεθόδων	
4.6 Συμπεράσματα	164

5. Υλοποίηση σε FPGA				
5.1 Αναγκαιότητα της υλοποίησης σε υλικό	168			
5.2 Υπάρχουσες υλοποιήσεις				
5.3 Προτεινόμενη υλοποίηση	170			
5.3.1 Ανάλυση δομής του αλγορίθμου βελτίωσης αποχρώσεων	170			
5.3.2 Τροποποίηση του αλγορίθμου	171			
5.3.3 Τροποποιήσεις για επίτευξη διαδοχικής διοχέτευση	173			
5.3.4 Μετατροπές RGB→YCbCr και YCbCr→RGB	176			
5.3.5 Εύρεση μεγίστου-ελαχίστου	179			
5.3.6 Εύρεση ιστογράμματος				
5.3.7 Υπολογισμός των πινάκων Stretching				
5.3.8 Υπολογισμός ιστογράμματος Η				

5.3.9 Υπολογισμός παραμέτρων	
5.3.10 Σερπαντίνα και υπολογισμός της περιφέρειας	
5.3.11 Πίνακας απεικόνισης και τελικός μετασχηματισμός	193
5.4 Σύγκριση με τα αποτελέσματα του λογισμικού	196
5.5 Εναλλακτική υλοποίηση του κυκλώματος	198
5.6 Χαρακτηριστικά του κυκλώματος και της συσκευής FPGA	202
5.7 Συμπεράσματα	203
6. Συμπεράσματα	205
6.1 Συμπεράσματα διατριβής	206
6.2 Προτάσεις για μελλοντική έρευνα	208
Βιβλιογραφία	211
Κατάλογος Δημοσιεύσεων	219
Κατάλογος Αγγλικών Όρων	221
Κατάλογος Ακρωνυμίων	225

- 4 Αποσαφήνιση εννοιών
- 5 Μη χρωματικά μεγέθη
- 10 Χρώμα
- 15 Τεχνητά συστήματα όραση
- 16 Το Ανθρώπινο Οπτικό Σύστημα
- 17 Ο οφθαλμός
- 26 Πρωτοταγής οπτικός φλοιός
- 30 Συμπεράσματα



1

Βιολογικά & Ηλεκτρονικά Οπτικά Συστήματα

Το κεφάλαιο αυτό αναφέρεται στη βασική δομή και τα χαρακτηριστικά των βιολογικών οπτικών συστημάτων, με έμφαση στο Ανθρώπινο Οπτικό Σύστημα (ΑΟΣ) και το οπτικό σύστημα των πρωτευόντων θηλαστικών. Γίνεται αντιπαραβολή σε κάθε βασικό τους στάδιο με τα τεχνητά οπτικά συστήματα και παρατίθενται οι ομοιότητες και οι διαφορές τους. Επίσης, διασαφηνίζονται έννοιες της φωτομετρίας και της οπτικής αντίληψης, οι οποίες είναι διφορούμενες και χρησιμοποιούνται πολλές φορές λανθασμένα. Τέλος, παρουσιάζονται σε κάθε στάδιο, οπτικά παραδείγματα που αποκαλύπτουν την επεξεργασία που πραγματοποιεί το ΑΟΣ και που πολλές φορές διαφέρει σημαντικά από το αναμενόμενο. Κύριος σκοπός είναι η γεφύρωση των δύο πλευρών της επιστήμης που ασχολούνται με το πρόβλημα της όρασης: της οπτικής νευροεπιστήμης

1.1 Διασαφήνιση βασικών εννοιών

Με τη μελέτη της όρασης ασχολούνται δύο γενικότεροι κλάδοι της επιστήμης, οι οποίοι, το αντιμετωπίζουν από διαφορετικές οπτικές γωνίες. Ο πρώτος είναι ο κλάδος των βιολογικών επιστημών, στον οποίο περιλαμβάνονται η νευροφυσιολογία, που αναλύει το ΑΟΣ σε φυσικό επίπεδο (τρόποι διασύνδεσης νευρώνων) και η ψυχοφυσιολογία, που αναλύει τις αντιλήψεις των παρατηρητών σε διάφορα σήματα εισόδου. Στόχος τους είναι να αποκρυπτογραφήσουν το ΑΟΣ και γενικότερα τα οπτικά συστήματα των πρωτευόντων θηλαστικών, να αναδείξουν και να ερμηνεύσουν τις λειτουργίες των επιμέρους τμημάτων τους και να δώσουν απαντήσεις στο ερώτημα «πώς σχηματίζεται η οπτική αντίληψη». Απώτερος στόχος είναι η δημιουργία ενός μαθηματικού μοντέλου, το οποίο θα μπορεί να προβλέψει με ακρίβεια την οπτική αντίληψη για κάθε εικόνα εισόδου στο ΑΟΣ. Όλα αυτά πραγματοποιούνται παράλληλα σε δύο κατευθύνσεις. Η μια κατεύθυνση περιλαμβάνει μακροσκοπικές παρατηρήσεις μέσω πειραμάτων οπτικής αντίληψης σε ανθρώπους ή ζώα (ψυχοφυσιολογία), και οδηγεί στη δημιουργία μαθηματικών μοντέλων, που στοχεύουν στην όσο το δυνατό πιο πιστή αναπαραγωγή των αποτελεσμάτων των ερευνών αυτών. Η άλλη κατεύθυνση περιλαμβάνει μικροσκοπικές παρατηρήσεις, μέσω χαρτογράφησης των οπτικών νευρικών κυκλωμάτων με τεχνικές απεικόνισης ή ηλεκτρικής διέγερσης (νευροφυσιολογία) και οδηγεί στην άμεση κατανόηση των χαρακτηριστικών του ΑΟΣ και στην επαλήθευση ή αναθεώρηση των μαθηματικών μοντέλων που αναπτύχθηκαν από τις μακροσκοπικές παρατηρήσεις. Έτσι, δημιουργείται μια συνεχή αλληλεπίδραση μεταξύ των δύο αυτών κατευθύνσεων. Η όλη διαδικασία είναι ίδια με τη σχέση που έχουν η θεωρητική και η πειραματική φυσική, στην οποία η πρώτη αναπτύσσει θεωρίες και η δεύτερη τις επαληθεύει ή τις αναιρεί. Δεν είναι λίγες οι φορές που κάποια θεωρία οπτικής αντίληψης προέβλεψε την ύπαρξη κάποιου νευρικού κυκλώματος και μετέπειτα νευροφυσιολογικές έρευνες επαλήθευσαν την πρόβλεψη αυτή. Πρέπει επίσης να σημειωθεί ότι ο κλάδος των βιολογικών επιστημών ήταν ο πρώτος που ασχολήθηκε με τη μελέτη της όρασης. Ήδη από τον 19° αιώνα εκπονούνταν συστηματικές ψυχοφυσιολογικές μελέτες και αναπτύσσονταν μοντέλα για την οπτική αντίληψη (Helmholtz, 1866), (Hering, 1874, 1964).

Ο δεύτερος κλάδος που ασχολείται με τη μελέτη της όρασης είναι ο κλάδος των θετικών επιστημών, ο οποίος περιλαμβάνει τη φυσική, τη φωτομετρία και την όραση μηχανών. Κύριος στόχος του είναι η ανάπτυξη εφαρμογών σε αυτόματα συστήματα που περιλαμβάνουν κάμερα και μονάδα επεξεργασίας, καθώς και η επίλυση προβλημάτων που προκύπτουν από τις εφαρμογές αυτές.

Η ταυτόχρονη ανάπτυξη των δύο αυτών κλάδων, έχει οδηγήσει, όπως είναι φυσικό, στο διαφορετικό ορισμό βασικών εννοιών του πεδίου της όρασης, οι οποίες πολλές φορές είναι αντικρουόμενες. Για το λόγο αυτό ακολουθεί μια αναλυτική διασαφήνιση αυτών των όρων, οι οποίοι και είναι απαραίτητοι για την κατανόηση της διατριβής.

1.1.1 Μη χρωματικά μεγέθη

Η Εικόνα 1.1 δείχνει παραστατικά τα βασικά μη χρωματικά μεγέθη που σχετίζονται με την όραση, ενώ ο Πίνακας 1.1 τις μονάδες αυτών των μεγεθών. Ακολουθούν οι αναλυτικοί ορισμοί του κάθε μεγέθους και αναφέρονται οι πιθανές λανθασμένες χρήσεις τους, καθώς και παραδείγματα που δικαιολογούν τους ορισμούς αυτούς.



	Μέγεθος	Σύμβολο	Μονάδα SI	Σχέση	Παρατηρήσεις
Πίνακας 1.1 Μονάδες και σχέσεις φωτομετρικών μεγεθών .	Φωτισμός (Illuminance)	E	lux	$lux = lm/m^2$	
	Φωτοβολία (Luminance)	L	cd/m ²		
	Φωτεινή ροή (Luminous flux)	F	lm	$lumen = cd \cdot sr$	
	Φωτεινή ένταση (Luminous intensity)	I	cd	кпріо candela = lm/sr	Βασική μονάδα του SI
	Στερεά γωνία (Solid angle)		sr	στερακτίνιο steradian	Είναι αδιάστατο, αφού 1 $sr=m^2\cdot m^{-2}=1$

Illuminance E(x, y): Φωτισμός (φωτομετρικό μέγεθος)

Φωτισμός (Illuminance) είναι «η συνολική ροή του φωτός που προσπίπτει πάνω σε μια επιφάνεια, ανά μονάδα επιφάνειας», μετρούμενη σε lux (Πίνακας 1.1).

Luminance L(x, y): Φωτοβολία (φωτομετρικό μέγεθος)

Φωτοβολία είναι «η ροή φωτός εκπεμπόμενη ανά μονάδα στερεάς γωνίας, από τη μονάδα επιφάνειας ενός αντικειμένου, προβαλλόμενη κάθετα στην κατεύθυνση παρατήρησης», μετρούμενη σε cd/m². Η φωτοβολία μιας ομοιόμορφα φωτισμένης διαχυτικής (diffusive) επιφάνειας, είναι το γινόμενο του προσπίπτοντος φωτός E(x, y) και της ανακλαστικότητάς της R(x, y) (Fiorentini, et al., 1990).

Σύμφωνα με τη CIE (Commission Internationale de l'Eclairage – Διεθνής Επιτροπή Φωτισμού), η φωτοβολία μπορεί να υπολογιστεί από το σταθμισμένο άθροισμα των γραμμικών αποκρίσεων των τριών βασικών χρωματικών συνιστωσών μιας κάμερας ως εξής:

$$Y(x, y) = 0.2125 RED(x, y) + 0.7154 GREEN(x, y) + 0.0721 BLUE(x, y)$$
(1.1)

όπου Y(x, y) η υπολογισμένη φωτοβολία και *RED*, *GREEN*, *BLUE* οι γραμμικές αποκρίσεις μιας κάμερας, πριν την εφαρμογή της διόρθωσης γ (Poynton, 1996). Οι σταθεροί συντελεστές στην εξίσωση (1.1) προκύπτουν από την ευαισθησία του τυπικού ανθρώπινου παρατηρητή στις τρεις βασικές χρωματικές συνιστώσες. Έτσι προσεγγίζεται ο τρόπος με τον οποίο καταγράφει το ανθρώπινο μάτι τη φωτοβολία μιας επιφάνειας. Εδώ πρέπει να σημειωθεί ότι η λέξη «καταγράφει» δεν εμπεριέχει την έννοια της αντίληψης, η οποία, όπως θα δούμε παρακάτω, συσχετίζεται με την έννοια της φωτεινότητας (lightness).

Reflectance R(x, y): Ανακλαστικότητα (φωτομετρικό μέγεθος)

Ανακλαστικότητα μιας ματ διαχυτικής (diffusive) επιφάνειας, είναι «ο λόγος της ροής του ανακλώμενου φωτός L(x, y) από την επιφάνεια, προς τη ροή του προσπίπτοντος φωτός E(x, y) στην επιφάνεια, και τα δύο σε σχέση με τη μονάδα επιφάνειας». Η ανακλαστικότητα σχετίζεται άμεσα με τις ιδιότητες της επιφάνειας του αντικειμένου (π.χ. απόχρωση) και είναι ανεξάρτητη του φωτισμού (Fiorentini, et al., 1990). Για παράδειγμα, μια επιφάνεια ανακλά το 25% του φωτός, ενός συγκεκριμένου μήκους κύματος, που προσπίπτει πάνω της (ιδιότητα της επιφάνειας). Αν ο προσπίπτον φωτισμός είναι 10 cd/m², τότε θα ανακλάσει 2.5 cd/m². Αν ο προσπίπτον φωτισμός γίνει 20 cd/m², τότε θα ανακλάσει 5 cd/m².

Intensity I(x, y): Ένταση (φωτομετρικό μέγεθος)

Ένταση (Intensity) είναι «το μέτρο της ροής ισχύος, σε ένα διάστημα του ηλεκτρομαγνητικού φάσματος, το οποίο εκπέμπεται από μια επιφάνεια ή προσπίπτει πάνω της» μετρούμενο σε watt/m². Η ένταση είναι μια γραμμική μέτρηση του φωτός (linear-light measure). Για το λόγο αυτό, η γραμμική απόκριση του αισθητήρα μιας

κάμερας, πριν την εφαρμογή της διόρθωσης γ, μπορεί να αναπαριστά την ένταση *I* του φωτός (Poynton, 1996).

Όλα τα παραπάνω μεγέθη είναι αντικειμενικά και μπορούν να μετρηθούν με ακρίβεια χρησιμοποιώντας φωτόμετρο ή κάμερα. Αντίθετα, η φωτεινότητα και η λαμπρότητα είναι αντιληπτικά (perceptual) μεγέθη και είναι το αποτέλεσμα της αντίληψης της φωτοβολίας ενός αντικειμένου από έναν παρατηρητή (Purves, et al., 2004). Όπως όλες οι αντιλήψεις, η φωτεινότητα και η λαμπρότητα είναι αδύνατο να μετρηθούν άμεσα και μπορούν μόνο να αποτιμηθούν με το να περιγράψει ο παρατηρητής την αντίληψη ενός ερεθίσματος σε σχέση με κάποιο άλλο που χρησιμοποιείται σαν αναφορά. Οι έννοιες της φωτεινότητας και της λαμπρότητας έχουν χρησιμοποιηθεί λανθασμένα πολλές φορές, συνήθως στην προσπάθεια να περιγραφούν οι έννοιες της έντασης ή της φωτοβολίας.

Lightness LS(x, y): Φωτεινότητα (αντιληπτικό μέγεθος)

Φωτεινότητα είναι «η ιδιότητα της οπτικής αντίληψης, σύμφωνα με την οποία ένα αντικείμενο φαίνεται να ανακλά περισσότερο ή λιγότερο φως, σε σχέση με ένα αντικείμενο που φαίνεται άσπρο» (Fiorentini, et al., 1990).

Η CIE (Commission Internationale de l'Eclairage – Διεθνής Επιτροπή Φωτισμού) περιγράφει τη φωτεινότητα L^* με αντικειμενικό τρόπο, ως συνάρτηση της φωτοβολίας Y (CIE, 1986).

$$L^* = 116 \left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} - 16, \quad \forall \frac{Y}{Y_n} > 0.008856$$
 (1.2)

όπου Y_n είναι η φωτοβολία του λευκού αναφοράς (reference white). Ο παραπάνω ορισμός βασίζεται στο γεγονός ότι η ανθρώπινη αντίληψη έχει μη γραμμική σχέση με τη φωτοβολία Y, η οποία περιγράφεται από ύψωση σε μια δύναμη.

Εντούτοις, αυτό ισχύει πρακτικά μόνο σε πειραματικές διατάξεις, στις οποίες υπάρχουν δύο μόνο επιφάνειες και από τις οποίες, η μια χρησιμοποιείται ως αναφορά για την απόχρωση της άλλης. Στην πραγματικότητα η σχέση της φωτεινότητας με τη φωτοβολία είναι πολύ πιο πολύπλοκη και εξαρτάται από τη χωρική διάταξη της σκηνής. Όλες οι ψυχοφυσιολογικές μελέτες, από τις αρχές του αιώνα, υποδεικνύουν ότι η φωτεινότητα είναι μια πολύπλοκη υποκειμενική αντίληψη. Κατά συνέπεια, δεν είναι δυνατό να περιγραφεί με αντικειμενικό τρόπο από μια απλή εξίσωση όπως η (1.2). Αυτό γίνεται εμφανές στις δύο επόμενες εικόνες, οι οποίες καταρρίπτουν την εξίσωση (1.2).

Στην Εικόνα 1.2, σύμφωνα με τη Φωτεινότητα *LS* που αντιλαμβάνεται ο ανθρώπινος παρατηρητής, το τετράγωνο Α είναι σκούρο και το Β ανοιχτό. Όμως, η μετρούμενη φωτοβολία τους (οπότε και η CIE *L**, σύμφωνα με την εξίσωση (1.2)) είναι ίδια (78). Αυτό συμβαίνει γιατί η CIE *L** είναι στην ουσία μια 1-1 απεικόνιση της φωτοβολίας, η οποία εμπεριέχει την επίδραση του προσπίπτοντος φωτισμού *E*, που στην Εικόνα 1.2 επηρεάζεται από τη σκιά του κυλίνδρου. Αντίθετα, η φωτεινότητα *LS* υπολογίζει την ανακλαστικότητα της επιφάνειας, βάσει των χωρικών σχέσεων της σκηνής, που στην περίπτωση της Εικόνας 1.2 είναι ότι το Α ανακλά λιγότερο από το Β. Έτσι, το ΑΟΣ εξάγει τα

αμετάβλητα χαρακτηριστικά της επιφάνειας (ανακλαστικότητα R) ανεξάρτητα από τις συνθήκες φωτισμού (E).



Εικόνα 1.2

LS ≠ CIE L* Ο ανθρώπινος παρατηρητής αντιλαμβάνεται το Α ως σκούρο και το Β ως ανοιχτό, ενώ η μετρούμενη φωτοβολία τους (οπότε και η CIE L*) είναι ίδια (78).



Εικόνα 1.3

Οι κύβοι του Logvinenko. Ακόμη μια απόδειξη ότι **LS ≠ CIE L***.

Ο ανθρώπινος παρατηρητής αντιλαμβάνεται το Α ως ανοιχτό και το Β ως σκοτεινό, ενώ η μετρούμενη φωταύγειά τους (οπότε και η CIE L*) είναι ίδια (139). Επίσης, η λωρίδα Γ έχει την ίδια σταθερή απόχρωση με τα Α και Β.

Η Εικόνα 1.3 δείχνει επίσης μια πολύ γνωστή οφθαλμαπάτη, η οποία αποδεικνύει ξεκάθαρα ότι το ΑΟΣ βασίζεται στις τοπικές χωρικές σχέσεις για τον υπολογισμό των αντιλήψεων του φωτεινού και του σκοτεινού. Ενώ οι τιμές φωτοβολίας, οπότε και η CIE *L*^{*}, των τμημάτων Α και Β είναι ίδιες (139), εντούτοις αντιλαμβανόμαστε το Α σαν φωτεινό και

το B σαν σκοτεινό. Η επίδραση των τοπικών σχέσεων είναι τόσο έντονη, που, όταν η λωρίδα Γ, η οποία έχει ίδια απόχρωση με τα Α και Β, τα ενώσει, φαίνεται σταδιακά να αλλάζει την απόχρωσή της, καθώς προοδευτικά αλλάζουν και οι γειτονικές της περιοχές.

Κατά συνέπεια, φωτεινότητα είναι η Αντιληπτή Ανακλαστικότητα (Perceived Reflectance). Επειδή όμως η ανακλαστικότητα έχει άμεση σχέση με την απόχρωση, μπορεί να διατυπωθεί ότι φωτεινότητα είναι η Αντιληπτή Απόχρωση (Perceived Pigmentation) (Fiorentini, et al., 1990).

Σε αντίθεση με την απλή εξίσωση (1.2), οι υπάρχουσες μέθοδοι για τον υπολογισμό της φωτεινότητας *LS* είναι πολύ πιο πολύπλοκες και εισάγουν πολλαπλές χωρικές συγκρίσεις (μέθοδοι Retinex και ACE – κεφάλαιο 3).

Brightness B(x, y): Λαμπρότητα (αντιληπτικό μέγεθος)

Λαμπρότητα είναι «η ιδιότητα της οπτικής αντίληψης σύμφωνα με την οποία μια περιοχή του οπτικού πεδίου φαίνεται να *εκπέμπει* περισσότερο ή λιγότερο φως σε σχέση με τις γειτονικές της περιοχές» (Fiorentini, et al., 1990).

Η λαμπρότητα διέπεται από τους ίδιους ψυχοφυσιολογικούς κανόνες που διέπουν και τη φωτεινότητα. Η μόνη τους διαφορά είναι ότι η μεν λαμπρότητα αναφέρεται σε αυτόφωτα αντικείμενα ή αντικείμενα με πολύ έντονη ανάκλαση σε σχέση με τη γειτονιά τους (π.χ. το φεγγάρι), ενώ η φωτεινότητα σε αντικείμενα με ματ διαχυτική επιφάνεια. Κατά συνέπεια, θα μπορούσε να ισχυριστεί κανείς ότι η λαμπρότητα αφορά τα αυτόφωτα αντικείμενα, και ότι η φωτεινότητα τις διαχυτικές επιφάνειες. Εν κατακλείδι, λαμπρότητα είναι ο *Αντιληπτός Φωτισμός (Perceived Illumination)* (Fiorentini, et al., 1990), (Purves, et al., 2004). Η Εικόνα 1.4 δείχνει τις ομοιότητες και τις διαφορές της λαμπρότητας και της φωτεινότητας.

Εικόνα 1.4

Τα βασικά χαρακτηριστικά των αντιληπτικών μεγεθών της Λαμπρότητας και της Φωτεινότητας.

Λαμπρότητα (Brightness)

(αμυδρό↔λαμπρό) (dim⇔bright)

Αντιληπτός Φωτισμός (Perceived Illumination)

Φωτεινότητα (Lightness)

(σκοτεινό↔φωτεινό)(dark↔light)

Αντιληπτή Απόχρωση (Perceived Pigmentation)

1.1.2 Χρώμα

1.1.2.1 Το παράδοξο της έγχρωμης όρασης

Το χρώμα είναι ίσως η πιο πολύπλοκη έννοια της όρασης. Και αυτό, γιατί διέπεται από ένα παράδοξο. Το παράδοξο προκύπτει από το γεγονός ότι το οπτικό φάσμα είναι έγχρωμο. Φως συγκεκριμένου μήκους κύματος έχει συγκεκριμένο χρώμα π.χ. το μεγάλο μήκος κύματος φαίνεται κόκκινο, το μεσαίο πράσινο και το μικρό μήκος κύματος ιώδες. Είναι λοιπόν πολύ φυσικό να υποθέσουμε ότι τα αντικείμενα και οι επιφάνειές τους, αποκτούν το χρώμα τους από το επικρατών μήκος κύματος (dominant wavelength) του φωτός που ανακλούν. Σύμφωνα με αυτή την προσέγγιση λοιπόν, ένα κόκκινο αντικείμενο φαίνεται κόκκινο ηιοθέτου μη που, μέχρι και σήμερα, έχει γίνει μέρος των πεποιθήσεων του πολιτισμού μας.

Και εδώ ακριβώς είναι το παράδοξο. Όταν βλέπουμε αντικείμενα κάτω από διαφορετικούς φωτισμούς, η σύνθεση του μήκους κύματος, που ανακλάται από αυτά, αλλάζει δραματικά. Εάν για παράδειγμα κάποιος παρατηρούσε ένα πορτοκάλι, ή μια μπανάνα σε ένα δωμάτιο που φωτίζονταν από μια λάμπα πυρακτώσεως, στη συνέχεια από μια λάμπα φθορισμού, μετά στο φως της ημέρας και στη συνέχεια στο σούρουπο, θα διαπίστωνε ότι το πορτοκάλι θα εξακολουθούσε να έχει πορτοκαλί χρώμα και η μπανάνα κίτρινο. Θα υπήρχαν κάποιες αλλαγές στις αποχρώσεις του πορτοκαλί και του κίτρινου, όμως το βασικό χρώμα θα παρέμενε το ίδιο. Εντούτοις, εάν μετρούσαμε τη σύνθεση του μήκους κύματος του ανακλώμενου φωτός σε όλες αυτές τις συνθήκες φωτισμού θα βρίσκαμε μεγάλες διαφορές. Σε φυσικές συνθήκες παρατήρησης, λοιπόν, δεν υπάρχει κάποια προκαθορισμένη σύνθεση μήκους κύματος, ή κάποιος κώδικας που να αντιστοιχεί σε ένα και μόνο χρώμα. Πράγματι, εάν τα χρώματα των αντικειμένων άλλαζαν με την παραμικρή αλλαγή του φωτισμού, τότε το χρώμα θα έχανε τη σημασία του σαν παράγοντας ένδειξης των χαρακτηριστικών ιδιοτήτων των αντικειμένων. Αυτό αναιρεί την ύπαρξη της έγχρωμης όρασης από εξελικτική άποψη, αφού θα ήταν δυνατό να διακριθεί ο κυνηγός ή το θήραμα από τα χρώματά τους, καθώς αυτά θα άλλαζαν ανάλογα με το φωτισμό. Η ιδιότητα της έγχρωμης όρασης να διατηρεί σταθερή την αντίληψη του χρώματος των αντικειμένων κάτω από διαφορετικούς φωτισμούς, ανεξάρτητα από το επικρατών μήκος κύματος που ανακλάται, είναι γνωστή ως χρωματική σταθερότητα (color constancy) και είναι μια από τις πιο σημαντικές ιδιότητες των βιολογικών οπτικών συστημάτων.

Ο λόγος για τον οποίο έχει επικρατήσει ως σήμερα η λανθασμένη άποψη, ότι το χρώμα εξαρτάται από το επικρατών μήκος κύματος, είναι ότι υπάρχει μια σπάνια περίπτωση στην οποία αυτό μπορεί να συμβεί. Η περίπτωση αυτή είναι γνωστή ως συνθήκη διαφράγματος (aperture condition) και εμφανίζεται όταν ένα χρώμα παρατηρείται σε πλήρη απομόνωση από όλα τα άλλα. Στην πράξη κάτι τέτοιο είναι εξαιρετικά σπάνιο να συμβεί στην καθημερινή ζωή. Σε συνθήκες εργαστηρίου όμως, αυτό μπορεί να συμβεί όταν όλο το οπτικό πεδίου του παρατηρητή καλύπτεται από μαύρο χρώμα, εκτός από ένα πολύ μικρό τμήμα στο οποίο υπάρχει η προς παρατήρηση επιφάνεια. Στην περίπτωση αυτή, διαπιστώνεται ότι το χρώμα της επιφάνειας αλλάζει ανάλογα με το κυρίαρχο μήκος κύματος του προσπίπτοντος φωτισμού. Έτσι, στη συνθήκη διαφράγματος, οποιαδήποτε

Χρωματική σταθερότητα

Μια εξαίρεση

επιφάνεια ανεξαρτήτως χρώματος, μπορεί να αποκτήσει οποιοδήποτε χρώμα, ανάλογα με το φωτισμό.



1.1.2.2 Το πείραμα του Land

Η μεγαλύτερη εξέλιξη στο χώρο του χρώματος, πραγματοποιήθηκε στη δεκαετία του 1950 από τον Edwin Land, και άλλαξε σημαντικά την αντίληψή μας για το «τι είναι χρώμα», επιβεβαιώνοντας και πειραματικά την ανεξαρτησία του χρώματος από το κυρίαρχο μήκος κύματος του ανακλώμενου φωτός. Η Εικόνα 1.5 περιγράφει παραστατικά το αξιοσημείωτο αυτό πείραμα. Ο Land χρησιμοποίησε εικόνες με έγχρωμα παραλληλόγραμμα σχήματα, τις οποίες ονόμασε Mondrians, από το όνομα του Ολλανδού ζωγράφου Piet Mondrian, ο οποίος σχεδίαζε πίνακες παρόμοιας τεχνοτροπίας. Τα παραλληλόγραμμα που σχημάτιζαν την έγχρωμη εικόνα είχαν διαφορετικά μεγέθη και σχήματα, έτσι ώστε να σχηματίζουν μια αόριστη σκηνή, χωρίς κάποιο χαρακτηριστικό αντικείμενο. Αυτό εξάλειφε τους παράγοντες της μάθησης και της μνήμης. Κανένα παραλληλόγραμμο δεν περιβάλλονταν από ένα και μόνο άλλο. Η προσέγγιση αυτή εξάλειφε τους παράγοντες της χρωματικής αντίθεσης (color contrast) ή του προκαλούμενου χρώματος (induced color). Οι επιφάνειες των παραλληλογράμμων ήταν φτιαγμένες από ματ χαρτί, ώστε να ανακλούν την ίδια ποσότητα ακτινοβολίας σε όλες της κατευθύνσεις και να επιτρέπουν την παρατήρηση από όλες τις γωνίες.

Η αόριστη αυτή έγχρωμη εικόνα φωτίζονταν από τρεις προβολείς, καθένας εκ των οποίων είχε έναν ροοστάτη που ρύθμιζε την ένταση του εκπεμπόμενου φωτός. Κάθε προβολέας είχε ειδικό φίλτρο, το οποίο επέτρεπε τη διέλευση φωτός ενός μόνο μήκους κύματος. Ο πρώτος προβολέας εξέπεμπε φως μεγάλου μήκους κύματος, ο δεύτερος μεσαίου και ο τρίτος μικρού μήκους κύματος. Έτσι, το Mondrian φωτιζόταν από τρία μονοχρωματικά φώτα τα οποία μπορούσαν να αναμειχθούν σε οποιαδήποτε αναλογία. Ένα χρωματόμετρο μπορούσε να μετράει με ακρίβεια το μήκος κύματος το οποίο ανακλούσαν τα διάφορα παραλληλόγραμμα της εικόνας.

Σε ένα σκοτεινό δωμάτιο, ζητήθηκε από μια ομάδα παρατηρητών να εστιάσει σε ένα συγκεκριμένο παραλληλόγραμμο του Mondrian. Ρυθμίζοντας αρχικά την ένταση των τριών προβολέων στην ίδια αναλογία (255, 255, 255 – λευκό φως), μέτρησε με χρωματόμετρο την αναλογία των χρωμάτων για ένα λευκό και ένα πράσινο παραλληλόγραμμο της εικόνας. Όπως ήταν αναμενόμενο, σύμφωνα με το χρωματόμετρο, το λευκό παραλληλόγραμμο κάτω από το λευκό φωτισμό είχε αναλογία (255, 255, 255) ενώ το πράσινο παραλληλόγραμμο (0,255,80). Οι παρατηρητές που έβλεπαν το λευκό παραλληλόγραμμο ανάφεραν ότι έχει άσπρο χρώμα. Στη συνέχεια, άλλαξε την αναλογία έντασης των τριών προβολέων, και αντί για λευκό φως, τους ρύθμισε ώστε να φωτίζουν με την αναλογία που ακτινοβολούσε το πράσινο παραλληλόγραμμο, δηλαδή (0,255,80). Έστρεψε το χρωματόμετρο στο λευκό παραλληλόγραμμο και όπως ήταν αναμενόμενο, η αναλογία που ακτινοβολούνταν κάτω από τον πράσινο φωτισμό ήταν ίδια με την αναλογία της πράσινης περιοχής κάτω από το λευκό φωτισμό (0,255,80). Φωτομετρικά, το λευκό παραλληλόγραμμο στον πράσινο φωτισμό ήταν πανομοιότυπο με το πράσινο παραλληλόγραμμο στο λευκό φωτισμό. Σύμφωνα λοιπόν με την τριχρωματική θεωρία, τα δύο παραλληλόγραμμα είχαν ακριβώς το ίδιο χρώμα. Εντούτοις όμως, οι παρατηρητές που έβλεπαν το λευκό παραλληλόγραμμο κάτω από τον πράσινο φωτισμό, εξακολουθούσαν να αναφέρουν ότι έχει άσπρο χρώμα και όχι πράσινο, όπως υποδείκνυε το χρωματόμετρο. Το πείραμα επαναλήφθηκε για όλα τα παραλληλόγραμμα, κάνοντάς τα να ακτινοβολούν

Mondrians

Το πείραμα

Προβολείς

διαφορετικό κυρίαρχο μήκος κύματος από αυτό που πραγματικά ήταν (το κόκκινο παραλληλόγραμμο να ακτινοβολεί πράσινο κυρίαρχο μήκος κύματος, το μπλε κόκκινο κ.ο.κ.). Η αναφορά των παρατηρητών ήταν πάντοτε ίδια: το αντιληπτό χρώμα του παραλληλογράμμου ήταν το πραγματικό του και όχι αυτό που υποδείκνυε το κυρίαρχο μήκος κύματος. Το συμπέρασμα από το πείραμα αυτό είναι το εξής:

Όταν μια επιφάνεια είναι μέρος μιας πολύχρωμης και πολύπλοκης σκηνής, δεν υπάρχει Συμπέρασμα προφανής και άμεση σχέση του χρώματός της με το κυρίαρχο μήκος κύματος του φωτός που εκπέμπεται από αυτή (Zeki, 1993).

Το πείραμα επαναλήφθηκε για συνθήκη διαφράγματος, κατά την οποία ολόκληρη η σκηνή καλύφθηκε από μια μαύρη μάσκα, εκτός από το προς παρατήρηση παραλληλόγραμμο. Σε όλες τις περιπτώσεις, οι παρατηρητές ανέφεραν ότι το εκάστοτε παραλληλόγραμμο άλλαζε το χρώμα του ανάλογα με τις αλλαγές του προσπίπτοντος φωτισμού. Μόλις απομακρύνονταν η μάσκα και εμφανιζόταν η συνολική εικόνα, το χρώμα του παραλληλογράμμου επανερχόταν στο αρχικό του ανεξάρτητα από το φωτισμό.

*Ανακλαστικό*τητα Ο Land διαπίστωσε ότι η ιδιότητα, βάσει της οποίας ο εγκέφαλος αποδίδει ένα χρώμα σε μια επιφάνεια, είναι η ανακλαστικότητά της, και όχι το κυρίαρχο μήκος κύματος που αυτή ανακλά. Όπως αναφέρθηκε και στην ενότητα 1.1.1, ανακλαστικότητα είναι το ποσοστό του κάθε μήκους κύματος που ανακλά μια επιφάνεια. Το ποσοστό αυτό είναι σταθερό και εξαρτάται μόνο από τις ιδιότητες του υλικού της επιφάνειας. Έτσι, ακόμα και αν ο προσπίπτων φωτισμός αλλάξει τόσο ώστε να μεταβληθεί το κυρίαρχο μήκος κύματος της ανάκλασης, τα ποσοστά του κάθε μήκους κύματος που ανακλώνται παραμένουν σταθερά. Αυτό γίνεται έκδηλο στην Εικόνα 1.6.



Εικόνα 1.6

Η ανακλαστικότητα μιας επιφάνειας παραμένει σταθερή και ανεξάρτητη του κυρίαρχου μήκους κύματος που ανακλά. Στην πρώτη περίπτωση ανακλάται περισσότερο φως μεγάλου μήκους κύματος. Στη δεύτερη περίπτωση, περισσότερο φως μικρού μήκους κύματος. Τα ποσοστά όμως παραμένουν πάντα σταθερά. Ένα ξεκάθαρο παράδειγμα του γεγονότος ότι το χρώμα δεν εξαρτάται από το κυρίαρχο μήκος κύματος παρουσιάζεται στην Εικόνα 1.7, η οποία δείχνει μια σκηνή κάτω από δύο διαφορετικούς φωτισμούς.



Εικόνα 1.7*

Πρώτη γραμμή: Μια σκηνή κάτω από δύο διαφορετικούς φωτισμούς.

Δεύτερη γραμμή:

Τα μπλε τετράγωνα στον κίτρινο φωτισμό και τα κίτρινα τετράγωνα στο μπλε φωτισμός, χωρίς την παρουσία των άλλων τετραγώνων.

Τρίτη γραμμή: Τα κόκκινα τετράγωνα κάθε εικόνας χωρίς την παρουσία των άλλων τετραγώνων.

* Η Εικόνα 1.7 είναι σχεδιασμένη για απεικόνιση στο χρωματικό μοντέλο RGB (οθόνη). Ως εκ τούτου, είναι αδύνατο να απεικονιστεί σωστά σε μια συμβατική εκτύπωση, αφού πολλά από τα χρώματά της βρίσκονται εκτός της γκάμας του μοντέλου CMYK. Για να διαπιστωθεί το πραγματικό μέγεθος αυτής της σημαντικής οφθαλμαπάτης, ο αναγνώστης καλείται να επισκεφτεί την ιστοσελίδα http://www.purveslab.net/resources/download.html#ccc από την οποία προήλθε και η εικόνα. Είναι φανερό ότι ο ανθρώπινος παρατηρητής αντιλαμβάνεται ξεκάθαρα το μπλε χρώμα στον κίτρινο φωτισμό και το κίτρινο χρώμα στο μπλε φωτισμό. Εντούτοις, και τα δύο χρώματα, σύμφωνα με την τριχρωματική θεωρία, θα έπρεπε να είναι γκρι, αφού και τα δύο έχουν ακριβώς τις ίδιες τιμές των βασικών χρωμάτων (136, 136, 136). Αυτό συμβαίνει γιατί το ΑΟΣ εξάγει τις ανακλαστικότητες του κάθε τετραγώνου σε σχέση με τα γειτονικά του. Αν όμως απαλειφθεί η χρωματική πληροφορία των γειτονικών τετραγώνων (συνθήκη διαφράγματος), τότε η απόχρωση που γίνεται αντιληπτή συμβαδίζει με το ποσοστό των τριών βασικών χρωμάτων. Το ίδιο συμβαίνει και με τα κόκκινα τετράγωνα. Όταν ο παρατηρητής τα βλέπει σε σχέση με τα υπόλοιπα, έχουν ξεκάθαρο κόκκινο χρώμα. Όταν όμως παρατηρηθούν μεμονωμένα, διαπιστώνεται ότι αποκτούν πορτοκαλί χρώμα (στον κίτρινο φωτισμό) και μωβ (στο μπλε φωτισμό).

1.2 Τεχνητά συστήματα όρασης

Η βάση των τεχνητών συστημάτων όρασης είναι η κάμερα. Αποτελεί το ηλεκτρονικό ανάλογο του ανθρώπινου ματιού και για το λόγο αυτό παρουσιάζονται ορισμένες ομοιότητες στη βασική δομή τους. Η Εικόνα 1.8 δείχνει ένα τυπικό ηλεκτρονικό σύστημα όρασης.



Αρχικά, το φως εστιάζεται με τη βοήθεια ενός φακού. Στη συνέχεια ακολουθεί μια διάταξη η οποία ρυθμίζει το ποσοστό του φωτός που θα περάσει στο σύστημα και θα διεγείρει τον αισθητήρα της εικόνας. Η πιο διαδεδομένη κατηγορία αισθητήρων είναι οι αισθητήρες CCD (Charge-Coupled Device). Στις περισσότερες κάμερες, δε χρησιμοποιούνται τρία διαφορετικά είδη αισθητήρων, ένα για κάθε βασικό μήκος κύματος. Αντίθετα, χρησιμοποιείται ένα επίπεδο αισθητήρων, μπροστά από το οποίο υπάρχει ένα φίλτρο. Το πιο διαδεδομένο φίλτρο είναι το φίλτρο Bayer (Bayer, 1976). Το φίλτρο αυτό αποτελεί ουσιαστικά τη χωρική υπέρθεση τριών επιμέρους μωσαϊκών: του φίλτρου για τα μεγάλα μήκη κύματος (R), του φίλτρου για τα μεσαία μήκη κύματος (G) και του φίλτρου για τα μικρά μήκη (B). Το μωσαϊκό των μεσαίων μηκών κύματος έχει διπλάσια ανάλυση σε σχέση με τα άλλα δύο, αφού, όπως θα παρουσιαστεί και παρακάτω, το μεγαλύτερο μέρος της φασματικής απορρόφησης του ΑΟΣ πραγματοποιείται στα μεσαία μήκη κύματος. Όταν το φίλτρο Bayer τοποθετηθεί μπροστά από τον αισθητήρα της κάμερας, το επιμέρους φίλτρο, που τυχαίνει να βρίσκεται μπροστά από το κάθε εικονοστοιχείο, καθορίζει και το φάσμα απορρόφησης για τη συγκεκριμένη θέση. Με τον τρόπο αυτό αποφεύγεται η χρήση τριών διαφορετικών αισθητήρων, περιορίζοντας όμως την ανάλυση στη χρωματική συνιστώσα. Ο αισθητήρας εξάγει τάση ανάλογη της ποσότητας φωτός που θα δεχθεί. Περισσότερες πληροφορίες σχετικά με τα χαρακτηριστικά των αισθητήρων CCD και το δυναμικό τους εύρος παρουσιάζονται στο κεφάλαιο 3. Η αναλογική τάση που εξάγει ο αισθητήρας, ψηφιοποιείται στη συνέχεια ώστε να είναι δυνατή η επεξεργασία της. Αν και οι περισσότερες κάμερες του εμπορίου έχουν ακρίβεια 8 bit για κάθε χρώμα, η ψηφιοποίηση γίνεται στα 10 bit, ώστε να υπάρχει αρκετή ακρίβεια κατά το στάδιο της επεξεργασίας (Yamada, 2006). Στη συνέχεια, ακολουθεί η μονάδα ψηφιακής επεξεργασίας σήματος (DSP), η οποία υλοποιεί ορισμένους απαραίτητους αλγορίθμους, ώστε να εξαχθεί ικανοποιητικό αποτέλεσμα στην έξοδο της κάμερας. Ένας από τους πιο σημαντικούς αλγορίθμους που εφαρμόζονται είναι η απομωσαϊκοποίηση (demosaicing) (Alleysson, et al., 2005). Ο αλγόριθμος αυτός πραγματοποιεί το διαχωρισμό των τριών χρωματικών συνιστωσών από την εικόνα που υπέρθεσης που κατέγραψε ο αισθητήρας. Επιπλέον, συνήθως πραγματοποιούνται ειδικές βελτιώσεις, όπως αποθορύβωση, ενώ υλοποιείται και ο αλγόριθμος συμπίεσης της εικόνας, ο οποίος συνήθως είναι ο JPEG. Τέλος, γίνεται ανάλυση της ποιότητας της εικόνας από ειδικούς αλγορίθμους οι οποίοι ελέγχουν το μηχανισμό έκθεσης του αισθητήρα, ώστε να επιτυγχάνεται βέλτιστη ποσότητα του προσπίπτοντος φωτός στον αισθητήρα της εικόνας (Yamada, 2006).

1.3 Το Ανθρώπινο Οπτικό Σύστημα

Στην ενότητα αυτή θα γίνει μια περιγραφή των βασικών τμημάτων του ΑΟΣ και θα τονιστούν οι όποιες διαφορές και ομοιότητες με τα συστήματα όρασης μηχανής. Παράλληλα, θα αναφερθούν συνοπτικά τα σημεία εκείνα του ΑΟΣ τα οποία χρησιμοποιήθηκαν ως αφετηρία για το σχεδιασμό των προτεινόμενων αλγορίθμων της παρούσας διατριβής. Η αναλυτική περιγραφή τους θα πραγματοποιηθεί στο κάθε κεφάλαιο ξεχωριστά.

1.3.1 Ροή οπτικής πληροφορίας

Η Εικόνα 1.9 δείχνει ένα γενικό διάγραμμα της ροής της οπτικής πληροφορίας στο ΑΟΣ.


Κεφάλαιο 1

Η οπτική πληροφορία προσλαμβάνεται αρχικά από τον αμφιβληστροειδή χιτώνα των δύο οφθαλμών. Κάθε οφθαλμός έχει το δικό του οπτικό πεδίο. Τα δύο πεδία έχουν ένα μεγάλο ποσοστό επικάλυψης, το οποίο δίνει τη δυνατότητα σε επόμενα επίπεδα επεξεργασίας, για υπολογισμό της απόστασης των αντικειμένων που παρατηρούνται. Η οπτική πληροφορία, που προσλαμβάνεται από τους δύο οφθαλμούς, μεταβιβάζεται μέσω του οπτικού νεύρου και του οπτικού χιάσματος στον Πλευρικό Γονατώδη Πυρήνα. Στο οπτικό χίασμα, οι οπτικές πληροφορίες διαχωρίζουν την πορεία τους σε δύο κατευθύνσεις. Αυτές που προέρχονται από το αριστερό μισό του συνολικού οπτικού πεδίου κατευθύνονται στο δεξιό ημισφαίριο του εγκεφάλου. Αντίθετα, αυτές που προέρχονται από το δεξιό μισό του συνολικού οπτικού πεδίου κατευθύνονται στο αριστερό ημισφαίριο του εγκεφάλου (Rodieck, 1998). Στον Πλευρικό Γονατώδη Πυρήνα οι οπτικές πληροφορίες αναδιοργανώνονται και κατευθύνονται στον Πρωτοταγή Οπτικό Φλοιό. Ο Πρωτοταγής Οπτικός Φλοιός, είναι γνωστότερος και ως περιοχή V1, από το Visual Area 1, αφού είναι ουσιαστικά η πρώτη περιοχή του ΑΟΣ που ανήκει στον εγκέφαλο και στην οποία γίνεται το πρώτο βασικό στάδιο της επεξεργασίας. Μετά την περιοχή V1 ακολουθούν οι ανώτερες περιοχές του ΑΟΣ, στις οποίες περιλαμβάνονται οι V2 και V3, οι οποίες σχετίζονται με την επεξεργασία περιγράμματος. Ακολουθεί η περιοχή V4 η οποία σχετίζεται με την αντίληψη του χρώματος, η V5/MT η οποία σχετίζεται με την αντίληψη της κίνησης και η IT, που σχετίζεται με την αντίληψη ολόκληρων αντικειμένων.

1.3.2 Ο οφθαλμός

Η Εικόνα 1.10 δείχνει τα βασικά μέρη του οφθαλμού και μια μεγέθυνση του αμφιβληστροειδή χιτώνα.



Τα οπτικά χαρακτηριστικά του οφθαλμού παρουσιάζουν μεγάλη ομοιότητα με αυτά της κάμερας. Το φως εισέρχεται μέσω του κερατοειδούς χιτώνα και της κόρης σε ένα φακό, ο οποίος εστιάζει το είδωλο στο πίσω μέρος του οφθαλμού, όπως ακριβώς και ο φακός εστίασης μιας κάμερας. Το διάφραγμα της κόρης εξασφαλίζει τον έλεγχο της ποσότητας του φωτός που εισέρχεται στον οφθαλμό και βρίσκεται σε πλήρη αναλογία με το διάφραγμα των καμερών. Στο πίσω μέρος του οφθαλμού βρίσκεται ο αμφιβληστροειδής χιτώνας, ο οποίος αποτελείται από τέσσερα είδη φωτοϋποδοχέων, για την απορρόφηση φωτός με μεγάλο, μεσαίο και μικρό μήκος κύματος. Σε ανάλογη θέση βρίσκονται και οι τρεις φωτοευαίσθητοι αισθητήρες της κάμερας, στους οποίους φωτόνια διαφορετικού μήκους κύματος μετατρέπονται σε τάση.

1.3.3 Ο αμφιβληστροειδής χιτώνας

Ο αμφιβληστροειδής είναι ένα πολύπλοκο νευρικό κύκλωμα το οποίο διαφέρει πολύ σε σχέση με τους αισθητήρες της κάμερας τόσο σε δομή όσο και στη λειτουργία. Σε αντίθεση με τις κάμερες, στις οποίες οι αισθητήρες πραγματοποιούν μόνο ανίχνευση, ο αμφιβληστροειδής είναι έτσι σχεδιασμένος ώστε να πραγματοποιεί και επεξεργασία, εκτός από την ανίχνευση. Η επεξεργασία αυτή πραγματοποιείται με τη μορφή τοπικών συγκρίσεων και κανονικοποιήσεων των σημάτων που ανιχνεύονται από τους φωτοϋποδοχείς (Rodieck, 1998). Στην πραγματικότητα, ο αμφιβληστροειδής είναι το πρώτο βασικό στάδιο επεξεργασίας του ΑΟΣ. Όλες οι διαφορετικές οδοί επεξεργασίας έχουν την αφετηρία τους σε αυτόν και από εδώ ξεκινούν οι πρώτες εξειδικεύσεις των νευρικών κυττάρων.

1.3.3.1 Φωτοϋποδοχείς

Η Εικόνα 1.11 δείχνει τη φασματική απόκριση των φωτοϋποδοχέων του ΑΟΣ και μιας γνωστής κάμερας.



Φασματικές αποκρίσεις Το ΑΟΣ διαθέτει τέσσερα είδη φωτοϋποδοχέων, σε σχέση με την κάμερα που διαθέτει τρία. Αυτά είναι τα κωνία (cones), τα οποία αποκρίνονται σε τρία διαφορετικά μήκη κύματος, στα μεγάλα (L – long), τα μεσαία (M – middle) και τα μικρά (S – small), και τα ραβδία (rods), τα οποία αποκρίνονται σε ένα μόνο μήκος κύματος, μεταξύ των Μ και L κωνίων. Είναι φανερό ότι υπάρχουν σημαντικές διαφορές μεταξύ των δύο διαγραμμάτων φασματικής απόκρισης του ΑΟΣ και της κάμερας. Στο διάγραμμα του ΑΟΣ υπάρχει σημαντική επικάλυψη μεταξύ της απόκρισης των κωνίων μεγάλου και μεσαίου μήκους κύματος, αφού η μέγιστη απόκρισή τους απέχει μόνο 30nm. Επίσης, καθώς κινούμαστε προς τα μικρά μήκη κύματος, οι δύο αυτές αποκρίσεις συμπίπτουν κοντά στα 410nm. Αντίθετα, οι φασματικές αποκρίσεις της κάμερας είναι πολύ πιο ομοιόμορφα διατεταγμένες. Έχουν σχεδόν το ίδιο σχήμα και οι μέγιστες αποκρίσεις σχεδόν ισαπέχουν μεταξύ τους.

Διαφορές κωνίων και ραβδίων

Οι διαφορές μεταξύ κωνίων και ραβδίων είναι πολλές. Η βασική τους διαφορά είναι ότι έχουν σχεδιαστεί για να λειτουργούν σε διαφορετικές συνθήκες φωτισμού. Τα κωνία λειτουργούν σε υψηλά επίπεδα φωτισμού (photopic vision), στα οποία τα ραβδία είναι κορεσμένα και δεν αποκρίνονται. Αντίθετα, τα ραβδία λειτουργούν σε χαμηλά επίπεδα φωτισμού (scotopic vision). Τα επίπεδα αυτά είναι πολύ χαμηλά για να διεγείρουν τα κωνία, τα οποία στην περίπτωση αυτή παραμένουν ανενεργά. Υπάρχει φυσικά και η ενδιάμεση κατάσταση, κατά την οποία τόσο τα κωνία όσο και τα ραβδία μπορούν να δώσουν αποκρίσεις (mesopic vision) (Wandell, 1995), (Rodieck, 1998). Μεγάλη είναι επίσης και η διαφορά στους πληθυσμούς των δύο φωτοϋποδοχέων. Υπάρχουν 5 εκατομμύρια περίπου κωνία και 100 εκατομμύρια ραβδία σε κάθε οφθαλμό. Η κατανομή αυτών των κυττάρων στον αμφιβληστροειδή δεν είναι ομοιόμορφη, αλλά διαφέρει ανάλογα με την απόσταση από τον κεντρικό βόθρο. Η Εικόνα 1.12 δείχνει αυτή την κατανομή.



Όπως φαίνεται από το διάγραμμα, τα κωνία και τα ραβδία έχουν συμπληρωματικές κατανομές. Στον κεντρικό βόθρο υπάρχουν μόνο κωνία. Καθώς απομακρυνόμαστε από το σημείο αυτό, ο αριθμός των κωνίων μειώνεται και ο αριθμός των ραβδίων αυξάνεται. Στις 20° περίπου μοίρες από τον κεντρικό βόθρο, τα ραβδία φτάνουν στο μέγιστο της κατανομής τους, ενώ στη συνέχεια καθώς η εκκεντρότητα αυξάνει, η κατανομή τους μειώνεται. Η μορφή της κατανομής, υποδεικνύει ότι, σε χαμηλά επίπεδα φωτισμού, στα οποία διεγείρονται μόνο τα ραβδία, είναι αδύνατο να αντιληφθούμε οποιαδήποτε πληροφορία στο κεντρικό σημείο του οφθαλμού. Πρέπει να σημειωθεί ότι στον αμφιβληστροειδή υπάρχει ένα σημείο που δε διαθέτει φωτοϋποδοχείς. Το σημείο αυτό ονομάζεται «τυφλό σημείο», γιατί είναι αδύνατο εκεί να καταγραφεί οποιαδήποτε οπτική πληροφορία (Wandell, 1995). Αυτό συμβαίνει γιατί ο αμφιβληστροειδής έχει ανάποδη διάταξη. Οι φωτοϋποδοχείς βρίσκονται στο πίσω μέρος, και οι νευρικές ίνες, που αποτελούν την έξοδο του αμφιβληστροειδούς, πρέπει να εξέλθουν από ένα σημείο του οφθαλμού. Στο σημείο εκείνο είναι αδύνατο να υπάρχουν ταυτόχρονα και φωτοϋποδοχείς (Εικόνα 1.9). Υπό κανονικές συνθήκες είναι αδύνατο να αντιληφθούμε την ύπαρξη του τυφλού σημείου, καθώς μηχανισμοί «γεμίσματος» (filling-in) στον οπτικό φλοιό συμπληρώνουν τη χαμένη οπτική πληροφορία. Η κατανομή της Εικόνας 1.12 διαφέρει πολύ από την κατανομή των φωτοϋποδοχέων μιας κάμερας, αφού εκεί η κατανομή είναι ομοιόμορφή σε όλη την επιφάνεια του αισθητήρα.

Η Εικόνα 1.13 δείχνει τη μορφή του μωσαϊκού των φωτοϋποδοχέων του ΑΟΣ σε σχέση με το φίλτρο Bayer που χρησιμοποιείται στις περισσότερες ψηφιακές κάμερες. Αντίθετα με τις κάμερες, το μωσαϊκό των φωτοϋποδοχέων του ΑΟΣ είναι πολύ διαφορετικό. Οι φωτοϋποδοχείς έχουν εξαγωνικό σχήμα, σε αντίθεση με τα εικονοστοιχεία του αισθητήρα τα οποία είναι τετράγωνα. Επίσης, η διάταξή τους δεν παρουσιάζει κάποια περιοδικότητα. Έτσι, πολλές φορές δημιουργούνται συμπλέγματα από φωτοϋποδοχείς του ίδιου είδους, εμποδίζοντας τη δειγματοληψία στο σημείο εκείνο από φωτοϋποδοχείς άλλου μήκους κύματος. Επιπλέον, το μωσαϊκό των φωτοϋποδοχέων μικρού μήκους κύματος (S) παρουσιάζει μεγάλες ανομοιομορφίες. Στο κέντρο του κεντρικού βόθρου δεν υπάρχουν καθόλου S φωτοϋποδοχείς, ενώ στις υπόλοιπες θέσεις είναι πολύ πιο αραιά διατεταγμένοι από του L και M. Αυτό συμβαίνει γιατί η χρωματική εκτροπή (chromatic aberration) του φακού είναι τέτοια που θολώνει το τμήμα της εικόνας που αντιστοιχεί στα μικρά μήκη

Το μωσαϊκό των κωνίων κύματος. Έτσι, δεν υπάρχει λόγος για μεγαλύτερη πυκνότητα των S φωτοϋποδοχέων, αφού το σήμα που πρέπει να δειγματοληπτήσουν έχει μικρότερη μέγιστη συχνότητα (Wandell, 1995). Η αναλογία επίσης των φωτοϋποδοχέων διαφέρει πολύ σε σχέση με την αναλογία της κάμερας. Αν και οι ακριβείς αναλογίες διαφέρουν σε κάθε άτομο, έχουν αναφερθεί αναλογίες L/M=1.2 (Stockman, et al., 1993) και (L+M)/S=14 (Wandell, 1995).



Φωτοϋποδοχείς ΑΟΣ

Φίλτρο Bayer κάμερας



1.3.3.2 Γαγγλιακά κύτταρα

Όπως αναφέρθηκε στην αρχή της ενότητας, σε αντίθεση με τον αισθητήρα της κάμερας, ο αμφιβληστροειδής, εκτός από ανίχνευση, εκτελεί και επεξεργασία του οπτικού σήματος. Η επεξεργασία αυτή γίνεται με τη μορφή χωρικών και χρωματικών συγκρίσεων και λαμβάνει χώρα στα γαγγλιακά κύτταρα. Τα γαγγλιακά κύτταρα αποτελούν τη μόνη έξοδο από τον

Εικόνα 1.13

Το μωσαϊκό των φωτοϋποδοχέων του ΑΟΣ και το φίλτρο Bayer που χρησιμοποιείται στις ψηφιακές κάμερες. αμφιβληστροειδή. Οι νευρικοί άξονές τους σχηματίζουν το οπτικό νεύρο και οδηγούν το οπτικό σήμα στον Πλευρικό Γονατώδη Πυρήνα και στη συνέχεια στον πρωτοταγή οπτικό φλοιό (V1). Όλες οι εξειδικεύσεις του ΑΟΣ ξεκινούν από τον αμφιβληστροειδή και εκφράζονται από το είδος της επεξεργασίας που εκτελούν τα είδη των γαγγλιακών κυττάρων. Μέχρι σήμερα έχουν βρεθεί τρεις βασικές κατηγορίες γαγγλιακών κυττάρων στον αμφιβληστροειδή των πρωτευόντων θηλαστικών (Lee, 1996), (Dacey, 1999), (Calkins, 2001). Η Εικόνα 1.14 δείχνει τις τρεις βασικές αυτές κατηγορίες, καθώς και τις χαρακτηριστικές τους ιδιότητες.

Το σημαντικότερο χαρακτηριστικό των γαγγλιακών κυττάρων, που καθορίζει το είδος της επεξεργασίας που εκτελούν, είναι το υποδεκτικό τους πεδίο. Υποδεκτικό πεδίο ενός Υποδεκτικό πεδίο κυττάρου, είναι το σύνολο των φωτοϋποδοχέων και ο τρόπος με τον οποίο το κύτταρο συνδέεται άμεσα ή έμμεσα με αυτούς. Έτσι, ακόμα και αν κάποιο κύτταρο βρίσκεται μακριά από τον αμφιβληστροειδή, π.χ. στον πρωτοταγή οπτικό φλοιό, ακολουθώντας τη διαδρομή των συνδέσεών του με κατεύθυνση προς τον οφθαλμό, βρίσκουμε το σύνολο των φωτοϋποδοχέων και τον τρόπο με τον οποίο συνδέεται σε αυτούς.

Τα γαγγλιακά κύτταρα τύπου P (από το Parvi = μικρό) είναι η κατηγορία με το μικρότερο μέγεθος υποδεκτικού πεδίου, το οποίο παρουσιάζει ταυτόχρονα χωρικό και χρωματικό ανταγωνισμό κέντρου-περιφέρειας. Ο χωρικός ανταγωνισμός πραγματοποιείται αφαιρώντας τα σήματα δύο ομόκεντρων κυκλικών περιοχών του μωσαϊκού των φωτοϋποδοχέων: του κέντρου και της περιφέρειας. Επιπλέον, το κέντρο λαμβάνει πάντα διαφορετικό χρωματικό σήμα από την περιφέρεια. Έτσι, πραγματοποιείται ταυτόχρονα και χρωματικός ανταγωνισμός. Η Εικόνα 1.15 δείχνει το υποδεκτικό πεδίο δύο γαγγλιακών κυττάρων, εκ των οποίων το ένα είναι τύπου Ρ.



Κέντρο(+L+M+S)|Περιφέρεια(-L-M-S) Γαγγλιακό κύτταρο τύπου Μ (Magni) Γαγγλιακό κύτταρο τύπου Ρ (Parvi) Κέντρο(+L) Περιφέρεια(-Μ) Εικόνα 1.15 ενός γαγγλιακού Μωσαϊκό Φωτοϋποδοχέων

Το υποδεκτικό πεδίο κυττάρου Μ και ενός Ρ.

Τα γαγγλιακά κύτταρα Ρ αποτελούν τη βάση της όρασης υψηλής ευκρίνειας του ΑΟΣ, για το λόγο αυτό και διαθέτουν το μικρότερο υποδεκτικό πεδίο. Στον κεντρικό βόθρο, το κέντρο του υποδεκτικού τους πεδίου αποτελείται από ένα και μόνο κωνίο. Καθώς αυξάνεται η απόσταση από το σημείο αυτό, το υποδεκτικό πεδίο τους μεγαλώνει. Τα γαγγλιακά κύτταρα τύπου Ρ είναι επίσης και φορείς του χρωματικού ανταγωνισμού μεταξύ κόκκινου και πράσινου. Αυτό σημαίνει ότι το κέντρο και η περιφέρεια λαμβάνουν πάντα σήματα διαφορετικού μήκους κύματος: είτε σήματα από τα L κωνία (κόκκινο), είτε σήματα από τα Μ κωνία (πράσινο), όχι όμως ταυτόχρονα και από τα δύο. Η ταυτόχρονη συνύπαρξη του χωρικού και του χρωματικού ανταγωνισμού συνεπάγεται την ύπαρξη τεσσάρων διαφορετικών πληθυσμών κυττάρων Ρ με διαφορετικό υποδεκτικό πεδίο, όπως παρουσιάζεται στην Εικόνα 1.16.



Τα τέσσερα είδη υποδεκτικού πεδίου των γαγγλιακών κυττάρων τύπου Ρ.

> Η κάθε μια από τις τέσσερεις αυτές μορφές κυττάρων Ρ σχηματίζει και ένα διαφορετικό μωσαϊκό, το οποίο επεξεργάζεται ανεξάρτητα τα σήματα του αμφιβληστροειδούς. Έτσι, για παράδειγμα, τα γαγγλιακά κύτταρα Κ:(+L)|Π:(-M) αυξάνουν τη διέγερσή τους όταν δέχονται στο κέντρο τους σήμα από ένα κωνίο μεγάλου μήκους κύματος (L), ενώ μειώνουν την απόκρισή τους όταν δέχονται στην περιφέρειά τους σήμα από κωνία μεσαίου μήκους κύματος (M). Παρομοίως, τα κύτταρα K:(-L)|Π:(+M) μειώνουν τη διέγερσή τους όταν δέχονται στο κέντρο τους σήμα από κωνίο L, ενώ αυξάνουν την απόκρισή τους όταν δέχονται στην περιφέρειά τους σήμα από κωνία Μ. Τα κύτταρα Κ:(+Μ)|Π:(-L) αυξάνουν τη διέγερσή τους όταν δέχονται στο κέντρο τους σήμα από κωνίο Μ, ενώ μειώνουν την απόκρισή τους όταν δέχονται στην περιφέρειά τους σήμα από κωνία L. Τέλος, τα κύτταρα K:(-M)|Π:(+L) μειώνουν τη διέγερσή τους όταν δέχονται στο κέντρο τους σήμα από κωνίο Μ, ενώ αυξάνουν την απόκρισή τους όταν δέχονται στην περιφέρειά τους σήμα από κωνία L.

Γαγγλιακά κύτταρα Μ Τα γαγγλιακά κύτταρα τύπου M (από το Magni = μεγάλο) διαθέτουν μεγαλύτερο υποδεκτικό πεδίο από τα Ρ. Όπως και τα προηγούμενα, αυξάνουν το μέγεθος του υποδεκτικού τους πεδίου καθώς αυξάνεται και η απόστασή τους από τον κεντρικό βόθρο. Σε κάθε σημείο όμως του αμφιβληστροειδούς, τα γαγγλιακά κύτταρα Μ έχουν 3 φορές μεγαλύτερη διάμετρο υποδεκτικού πεδίου από τα κύτταρα Ρ (Εικόνα 1.15). Η μορφή του υποδεκτικού τους πεδίου περιλαμβάνει χωρικό ανταγωνισμό κέντρου περιφέρειας, όχι όμως και χρωματικό ανταγωνισμό, όπως στα κύτταρα Ρ. Ο χωρικός ανταγωνισμός πραγματοποιείται όπως και πριν με την αφαίρεση των σημάτων του κέντρου και της περιφέρειας. Η έλλειψη χρωματικού ανταγωνισμού υποδεικνύει ότι τα κύτταρα τύπου Μ

μπορούν να δέχονται τόσο στο κέντρο τους όσο και στην περιφέρειά τους σήματα από οποιοδήποτε είδος κωνίων. Στην Εικόνα 1.17 παρουσιάζονται τα δύο είδη υποδεκτικού πεδίου γαγγλιακών κυττάρων Μ.





Τα δύο είδη υποδεκτικού πεδίου των γαγγλιακών κυττάρων τύπου Μ.

Τα δύο αυτά είδη υποδεκτικών πεδίων υποδεικνύουν την ύπαρξη δύο διαφορετικών μωσαϊκών από κύτταρα M, τα οποία επεξεργάζονται ανεξάρτητα τα σήματα του μωσαϊκού των φωτοϋποδοχέων. Τα κύτταρα K:(+L+M+S)|Π:(-L-M-S) αυξάνουν τη διέγερσή τους όταν δεχθούν σήματα από οποιοδήποτε είδος κωνίων στο κέντρο τους, ενώ μειώνουν την απόκρισή τους όταν δεχθούν οποιοδήποτε σήμα στην περιφέρειά τους. Αντίθετα, τα κύτταρα K:(-L-M-S)|Π:(+L+M+S) μειώνουν τη διέγερσή τους όταν δεχθούν σήματα από οποιοδήποτε σήμα στην περιφέρειά τους. Αντίθετα, τα κύτταρα K:(-L-M-S)|Π:(+L+M+S) μειώνουν τη διέγερσή τους όταν δεχθούν σήματα από οποιοδήποτε σήμα στην περιφέρειά τους όταν δεχθούν σήματα από οποιοδήποτε είδος κωνίων στο κέντρο τους, ενώ μειώνουν την απόκρισή τους όταν δεχθούν στο κέντρο τους, ενώ αυξάνουν την απόκρισή τους όταν δεχθούν οποιοδήποτε σήμα στην περιφέρειά τους. Η διαφορική αυτή μορφή επεξεργασίας ουσιαστικά καταγράφει τις μεταβολές φωτεινής έντασης στο μωσαϊκό των φωτοϋποδοχέων και ισοδυναμεί με τον τελεστή της διαφοράς Γκαουσιανών (Difference of Gaussians – DoG) ή της λαπλασιανής της Γκαουσιανής κατανομής (Laplacian of Gaussian – LoG). Τα γαγγλιακά κύτταρα τύπου Μ αποτελούν τη βάση για την αντίληψη της φωτεινότητας, του βάθους και της κίνησης στο ΑΟΣ, αφού τα μεγάλα υποδεκτικά τους πεδία και οι χρονικές ιδιότητες απόκρισής τους επιτρέπουν τη σύγκριση φωτεινών εντάσεων σε μεγάλες σχετικά αποστάσεις.

Τόσο τα γαγγλιακά κύτταρα M όσο και τα γαγγλιακά κύτταρα P έχουν αποτελέσει την βάση για δύο από τις προτεινόμενες μεθόδους της παρούσας διατριβής. Στο κεφάλαιο 3 παρουσιάζεται ένας αλγόριθμος για τη βελτίωση των υποεκτεθειμένων και υπερεκτεθειμένων περιοχών των εικόνων, ο οποίος εκμεταλλεύεται τη μορφή της απόκρισης των γαγγλιακών κυττάρων με χωρικό ανταγωνισμό κέντρου-περιφέρειας. Στο κεφάλαιο 4 παρουσιάζεται μια μέθοδος δυαδικοποίησης εγγράφων που έχει ως βάση την αντίληψη της φωτεινότητας από το ΑΟΣ και ως εκ τούτου, χρησιμοποιεί αντίστοιχα γαγγλιακά κύτταρα M και P με αυτά του ΑΟΣ. Κατά συνέπεια, εκτεταμένη ανάλυση των

Η βάση των προτεινόμενων μεθόδων μαθηματικών και βιολογικών ιδιοτήτων αυτών των κυττάρων, περιλαμβάνεται στα κεφάλαια 3 και 4.

Η τρίτη μεγάλη κατηγορία γαγγλιακών κυττάρων είναι τα κύτταρα Κ (Konio). Η Εικόνα 1.18 Γαγγλιακά παρουσιάζει το υποδεκτικό πεδίο ενός γαγγλιακού κυττάρου Κ. Όπως και τα Μ, τα κύτταρα Κ γαγγλιακά κύτταρα Κ διαθέτουν σε κάθε σημείο του αμφιβληστροειδούς 3 φορές μεγαλύτερη διάμετρο υποδεκτικού πεδίου από τα Ρ. Η μορφή του υποδεκτικού τους πεδίου δεν περιλαμβάνει χωρικό ανταγωνισμό κέντρου περιφέρειας, αλλά μόνο χρωματικό ανταγωνισμό μπλε-κίτρινου. Κατά συνέπεια, υπάρχουν δύο μορφές υποδεκτικών πεδίων για τα κύτταρα Κ. Τα κύτταρα (+S-L-M) διεγείρονται όταν δέχονται στο υποδεκτικό τους πεδίο σήματα από κωνία S (μπλε), ενώ μειώνουν την απόκρισή τους όταν δεχθούν σήματα από κωνία Μ και L (πράσινο και κόκκινο). Αντίθετα, τα κύτταρα (+L+M-S) διεγείρονται όταν δέχονται στο υποδεκτικό τους πεδίο σήματα από κωνία Μ και L, ενώ αναστέλλουν την απόκρισή τους όταν δέχονται σήματα από κωνία S. Τα κύτταρα K σχηματίζουν δύο διαφορετικά μωσαϊκά, τα οποία επεξεργάζονται ανεξάρτητα τα σήματα του μωσαϊκού των φωτοϋποδοχέων, ανιχνεύοντας τις διαφορές μπλε-κίτρινου και κίτρινου-μπλε.



Συνοψίζοντας, τα γαγγλιακά κύτταρα είναι η μόνη έξοδος από τον αμφιβληστροειδή προς τον πρωτοταγή οπτικό φλοιό. Οι τρεις βασικές κατηγορίες γαγγλιακών κυττάρων σχηματίζουν 8 διαφορετικά μωσαϊκά υποδεκτικών πεδίων, τα οποία επεξεργάζονται τα σήματα του μωσαϊκού των φωτοϋποδοχέων με μοναδικό τρόπο το καθένα. Έτσι, 8 παράλληλες εικόνες, η κάθε μια με διαφορετικές ιδιότητες, αποστέλλονται από τον αμφιβληστροειδή στον οπτικό φλοιό μέσω του οπτικού νεύρου. Η διαφορική επεξεργασία, που πραγματοποιείται σε χωρικό και χρωματικό επίπεδο από τα γαγγλιακά κύτταρα, εξασφαλίζει τη σταθερότητα των οπτικών σημάτων που μεταδίδονται στον εγκέφαλο.

1.3.4 Ο πρωτοταγής οπτικός φλοιός (περιοχή V1)

Ο πρωτοταγής οπτικός φλοιός είναι η πρώτη βασική περιοχή οπτικής επεξεργασίας στον εγκέφαλο. Βασικός στόχος της επεξεργασίας που λαμβάνει χώρα στη V1 είναι η εξαγωγή απλών οπτικών χαρακτηριστικών, όπως ακμές συγκεκριμένης κλίσης, ανίχνευση κίνησης προς μια συγκεκριμένη κατεύθυνση, ανίχνευση φωτός συγκεκριμένου μήκους κύματος και πρώιμη ανίχνευση βάθους. Για το λόγο αυτό, υπάρχουν εξειδικευμένα είδη νευρώνων για το καθένα από τα παραπάνω χαρακτηριστικά, τα οποία σχηματίζουν παράλληλους χάρτες αυτών των χαρακτηριστικών. Ένα βασικό χαρακτηριστικό της περιοχής V1 είναι ότι έχει την οπτική οργάνωση του αμφιβληστροειδούς (retinotopically organized). Αυτό σημαίνει ότι ένα σημείο του οπτικού πεδίου αντιστοιχίζεται στο ίδιο σημείο του μωσαϊκού του αμφιβληστροειδούς θέση του χάρτη χαρακτηριστικών της περιοχής V1. Θα μπορούσαμε να πούμε ότι πρόκειται για δύο εικόνες με ίδιο σύστημα συντεταγμένων.

1.3.4.1 Υπερστήλες (Hypercolumns)

Το κυριότερο ίσως χαρακτηριστικό της περιοχής V1 είναι η κρυσταλλική της δομή. Αυτό σημαίνει ότι αποτελείται από ένα αυτόνομο σύνολο νευρώνων, το οποίο επαναλαμβάνεται περιοδικά σε όλο το εύρος της περιοχής V1. Κάθε τέτοιο σύνολο νευρώνων ονομάζεται υπερστήλη και περιλαμβάνει κύτταρα ανίχνευσης όλων των χαρακτηριστικών που προαναφέρθηκαν (κλίσης ακμών, μήκους κύματος, κατεύθυνσης κίνησης και βάθους). Έτσι, κάθε υπερστήλη αποτελεί μια ανεξάρτητη μονάδα επεξεργασίας για το συγκεκριμένο σημείο του οπτικού πεδίου το οποίο επεξεργάζεται. Η Εικόνα 1.19 παρουσιάζει τα δύο υπάρχοντα μοντέλα μιας υπερστήλης.





Εικόνα 1.19

i. Αναπαράσταση υπερστήλης σύμφωνα με το μοντέλο «ανεμόμυλος» (pinwheel).

ii. Αναπαράσταση υπερστήλης σύμφωνα με το μοντέλο «παγάκι» (ice cube). Πρέπει να σημειωθεί ότι η γεωμετρική δομή των υπερστηλών είναι στην πραγματικότητα εξαιρετικά πολύπλοκη και δεν είναι δυνατό να αποδοθεί σωστά από απλά γεωμετρικά σχήματα. Έχει δημοσιευτεί δε η άποψη ότι τα κύτταρα μιας υπερστήλης είναι ουσιαστικά σημεία μιας κλειστής επιφάνειας σε μη Ευκλείδειο χώρο, γνωστή στο πεδίο των μαθηματικών ως μπουκάλι του Klein (Klein bottle) (Swindale, 1996). Παρόλα αυτά, το μοντέλο που χρησιμοποιείται περισσότερο για την απεικόνιση των υπερστηλών είναι αυτό του «ανεμόμυλου» (pinwheel). Σύμφωνα με αυτό, στο κέντρο της υπερστήλης βρίσκονται κύτταρα που είναι ευαίσθητα στο μήκος κύματος του φωτός που προσπίπτει στο μέρος του αμφιβληστροειδούς με το οποίο είναι συνδεδεμένα. Εδώ πρέπει να τονιστεί ότι τα κύτταρα αυτά δεν είναι εξειδικευμένα για την αντίληψη του χρώματος, όπως αυτό ορίστηκε στην παράγραφο 1.1.2, αλλά μόνο για το μήκος κύματος. Ο υπολογισμός της αντίληψης του χρώματος πραγματοποιείται σε επόμενο στάδιο επεξεργασίας, στην περιοχή V4 του εγκεφάλου. Ακτινικά, γύρω από τα κύτταρα του μήκους κύματος, βρίσκονται κύτταρα ανίχνευσης κλίσης των ακμών και κίνησης προς συγκεκριμένες κατευθύνσεις. Η γωνία της κλίσης, που ανιχνεύει κάθε νευρώνας, καθορίζει και τη γωνία θέσης του. Καθώς κινούμαστε προς τη βάση της υπερστήλης βρίσκουμε κύτταρα κλίσης ακμών μεγαλύτερης κλίμακας. Έτσι, σε κάθε σημείο του οπτικού πεδίου, είναι δυνατή η ανίχνευση ακμών με διαφορετικά μεγέθη και κλίσεις.

1.3.4.2 Απλά κύτταρα κλίσης (simple cells)

Τα κύτταρα κλίσης χωρίζονται σε δύο βασικές κατηγορίες. Τα απλά κύτταρα και τα σύνθετα κύτταρα (complex cells). Τα απλά κύτταρα σχηματίζονται με την κατάλληλη σύνδεση κυττάρων ανταγωνισμού κέντρου-περιφέρειας, των οποίων η διάταξη των υποδεκτικών τους πεδίων σχηματίζει συγκεκριμένες κλίσεις (Εικόνα 1.20) (Hubel, et al., 1965), (Hubel, et al., 1968).



Εικόνα 1.20

Ο σχηματισμός υποδεκτικών πεδίων των απλών κυττάρων της περιοχής V1, από υποδεκτικά πεδία ανταγωνισμού κέντρου-περιφέρειας. Όπως δείχνει και η Εικόνα 1.20, όταν στο σημείο του αμφιβληστροειδούς που συνδέεται ένα απλό κύτταρο παρουσιαστεί μια ακμή με τη συγκεκριμένη κλίση που ανιχνεύει το κύτταρο, τότε και μόνο τότε το κύτταρο θα διεγερθεί. Σε όλες τις άλλες περιπτώσεις η απόκρισή του είναι ελάχιστη.

1.3.4.3 Σύνθετα κύτταρα (complex cells)

Τα απλά κύτταρα ίδιας κλίσης συνδέονται με τη σειρά τους έτσι ώστε τα υποδεκτικά τους πεδία να σχηματίσουν περιοχές στις οποίες ανιχνεύονται ακμές συγκεκριμένης κλίσης με μεγαλύτερη ανοχή στη θέση (Hubel, et al., 1965), (Hubel, et al., 1968). Η Εικόνα 1.21 παρουσιάζει τη δημιουργία των υποδεκτικών πεδίων των σύνθετων κυττάρων της περιοχής V1.



Όπως φαίνεται από την προηγούμενη εικόνα, τα υποδεκτικά πεδία που προκύπτουν μπορούν να ανιχνεύουν ακμές συγκεκριμένης κλίσης σε ένα «παράθυρο» του μωσαϊκού των φωτοϋποδοχέων. Επιπλέον, ο τρόπος σύνδεσης των κυττάρων είναι τέτοιος που επιτρέπει την ανίχνευση κίνησης προς μια κατεύθυνση. Έτσι, τα σύνθετα κύτταρα της περιοχής V1 αποτελούν τη βάση για την αντίληψη σχήματος και κίνησης, που πραγματοποιείται σε επόμενα επίπεδα επεξεργασίας, υψηλότερων οπτικών περιοχών του εγκεφάλου.

Τα σύνθετα κύτταρα όμως, εκτός από το κλασικό υποδεκτικό τους πεδίο, διαθέτουν και πεδίο συσχέτισης (association field). Αυτό σημαίνει ότι μεταβάλλουν τις αποκρίσεις τους ανάλογα με τις αποκρίσεις των γειτονικών τους κυττάρων. Η αλληλεπίδραση αυτή είναι τέτοια που ευνοεί τη σωστή συνέχεια των περιγραμμάτων (Hess, et al., 2003). Στην ιδιότητά τους αυτή, βασίζεται και η προτεινόμενη μέθοδος της διατριβής για την εξαγωγή των εξεχόντων περιγραμμάτων που παρουσιάζεται στο κεφάλαιο 2. Περισσότερες λεπτομέρειες για τις μαθηματικές και βιολογικές ιδιότητες των κυττάρων αυτών δίνονται στο κεφάλαιο 2.

1.3.5 Ανώτερες οπτικές περιοχές

Μετά τον πρωτοταγή οπτικό φλοιό (V1), στον οποίο πραγματοποιείται βασική επεξεργασία για όλα τα είδη οπτικών χαρακτηριστικών, ακολουθούν οπτικές περιοχές με μεγαλύτερη εξειδίκευση, οι οποίες επεξεργάζονται συγκεκριμένα μόνο οπτικά χαρακτηριστικά.

Οι περιοχές V2 και V3 ασχολούνται κυρίως με την επεξεργασία των περιγραμμάτων. Τα απλά και σύνθετα κύτταρα διαφόρων κλίσεων της περιοχής V1 σχηματίζουν ακόμα πιο σύνθετα κύτταρα με μεγαλύτερα και πιο πολύπλοκα υποδεκτικά πεδία, όπως παρουσιάζει και η επόμενη εικόνα.



Εικόνα 1.22

Ο σχηματισμός πολύπλοκων υποδεκτικών πεδίων στις περιοχές V2 και V3, από το συνδυασμό των υποδεκτικών πεδίων της περιοχής V1.

Στις περιοχές αυτές έχουν βρεθεί υποδεκτικά πεδία, τα οποία έχουν το σχήμα συγκεκριμένων γωνιών και διακλαδώσεων ακμών. Επίσης, στην περιοχή αυτή πραγματοποιείται η επεξεργασία των ψευδαισθητικών περιγραμμάτων (illusionary contours), τα οποία προκαλούν την αίσθηση ακμών, χωρίς όμως να υπάρχει ανάλογο ερέθισμα στον αμφιβληστροειδή (Grossberg, 2004).

Η περιοχή V4 περιέχει νευρώνες ευαίσθητους σε ακμές, αλλά πολύ περισσότερο εξειδικευμένους στο αντιληπτό χρώμα. Εδώ πρέπει να τονιστεί ότι οι νευρώνες αυτοί δεν αποκρίνονται στο μήκος κύματος του φωτός που προσπίπτει στον αμφιβληστροειδή, αλλά στο χρώμα το οποίο αντιλαμβάνεται ο άνθρωπος. Για το λόγο αυτό, η περιοχή V4 έχει ονομαστεί και «χρωματικό κέντρο» του ΑΟΣ (Zeki, 1993).

Η περιοχή V5, γνωστότερη και ως MT, είναι εξειδικευμένη στην αντίληψη της κίνησης. Τα σύνθετα κύτταρα της περιοχής V1 συνεργάζονται ώστε να δημιουργήσουν πιο πολύπλοκες αντιλήψεις κίνησης. Ασθενείς με πρόβλημα στην περιοχή MT βλέπουν τον κόσμο χωρίς κίνηση και αντιλαμβάνονται την κίνηση σαν ένα βίντεο με μικρό αριθμό από καρέ (Zeki, 1993).

Τέλος, η περιοχή ΙΤ (Inferotemporal Cortex – Κάτω Κροταφικός Λοβός) έχει συσχετιστεί με την αντίληψη συγκεκριμένων αντικειμένων. Στην περιοχή αυτή υπάρχουν νευρώνες που

δέχονται εισόδους τόσο από νευρώνες περιγραμμάτων όσο και από νευρώνες χρώματος. Έτσι, προκύπτουν πληθυσμοί νευρώνων από εξειδικευμένους συνδυασμούς χαρακτηριστικών, οι οποίοι αντιστοιχούν σε συγκεκριμένα αντικείμενα.

1.4 Συμπεράσματα

Στο κεφάλαιο αυτό παρουσιάστηκαν τα βασικότερα στοιχεία του οπτικού συστήματος των πρωτευόντων θηλαστικών, σε αντιπαράθεση με αυτά των τεχνητών συστημάτων όρασης και των καμερών. Διαπιστώνεται ότι οι διαφορές είναι πολύ περισσότερες από τις ομοιότητες. Αυτό πρέπει να αποτελέσει βάση για την εξέλιξη των τεχνητών οπτικών συστημάτων προς την κατεύθυνση των βιολογικών, αφού τα δεύτερα υπερέχουν ξεκάθαρα των πρώτων. Οι προτεινόμενες μέθοδοι στη διατριβή αυτή χρησιμοποιούν ως βάση κάποια από τα χαρακτηριστικά ορισμένων κυττάρων του ΑΟΣ ώστε να κινηθούν προς αυτήν την κατεύθυνση. Συγκεκριμένα, εκμεταλλεύονται τις ιδιότητες των γαγγλιακών κυττάρων στην ανίχνευση μεταβάσεων φωτεινότητας (κεφάλαιο 4 – δυαδικοποίηση εγγράφων), αντιγράφουν την απόκριση σε σχέση με την περιφέρεια των γαγγλιακών κυττάρων κυττάρων κυττάρων το πεδίο συσχέτισης των πολύπλοκων κυττάρων της περιοχής V1 (κεφάλαιο 2 – εξαγωγή αδρών περιγραμμάτων).



- 76 Συμπεράσματα
- 65 Συγκρίσεις με υπάρχουσες μεθόδους
- 54 Προτεινόμενη μέθοδος
- 50 Αναγνώριση απλών σχημάτων με μάσκες κλίσης
- 40 Μοντελοποίηση κυττάρων κλίσης
- 37 Ανασκόπηση υπάρχουσας βιβλιογραφίας
- 34 Νευροφυσιολογικά στοιχεία του ΑΟΣ
- 34 Διατύπωση του προβλήματος

Εξέχοντα περιγράμματα

Το κεφάλαιο αυτό αναφέρεται στο πρόβλημα της τμηματοποίησης ακμών που ανήκουν σε περιγράμματα αντικειμένων από το γενικότερο σύνολο των ακμών που εξάγονται σε μια εικόνα. Δίνεται ο αναλυτικός ορισμός του προβλήματος και παρουσιάζεται ο σημαντικός του ρόλος στα προβλήματα όρασης μηχανής. Στη συνέχεια, παρατίθενται στοιχεία από νευροφυσιολογικές μελέτες του εγκεφάλου, τα οποία σκιαγραφούν τον τρόπο με τον οποίο η εξέλιξη έχει αντιμετωπίσει το πρόβλημα αυτό. Επίσης, γίνεται μια σύντομη περιγραφή των μεθόδων που έχουν αναπτυχθεί μέχρι τώρα για την επίλυση αυτού του προβλήματος. Ακολουθεί η περιγραφή της προτεινόμενης μεθόδου, η οποία υιοθετεί στοιχεία από το ΑΟΣ, καθώς και σειρά συγκριτικών πειραμάτων. Τέλος, παρατίθενται συμπεράσματα που αναλύουν τη συνεισφορά της προτεινόμενης μεθόδου.

2.1 Διατύπωση του προβλήματος

Η εξαγωγή των ακμών μιας ψηφιακής εικόνας είναι θεμελιώδης διεργασία στην επεξεργασία εικόνας και στην όραση μηχανής. Εντούτοις, ένα από τα βασικά προβλήματα των σημερινών εξαγωγέων ακμών είναι ότι δεν κάνουν διάκριση μεταξύ των περιγραμμάτων των αντικειμένων, τα οποία είναι αυτά που ουσιαστικά ενδιαφέρουν, και στις ακμές που προέρχονται από υφή ή που δεν αντιστοιχούν σε κάποιο συγκεκριμένο αντικείμενο. Αποτέλεσμα αυτού είναι ότι όταν πραγματοποιείται η εξαγωγή ακμών σε μη ελεγχόμενες συνθήκες, όπου το φόντο πίσω από το αντικείμενο δεν είναι συνήθως ομοιόμορφο, να εξάγονται πολλές ακμές οι οποίες δεν αντιστοιχούν στο περίγραμμα του αντικειμένου. Αυτό μειώνει δραματικά την απόδοση άλλων αλγορίθμων που έπονται της εξαγωγής ακμών, όπως ταξινομητές σχήματος. Είναι λοιπόν σημαντικό να παρουσιαστούν τεχνικές οι οποίες θα βοηθήσουν στο διαχωρισμό των ακμών του φόντου από τις ακμές των αντικειμένων.



Το παραπάνω πρόβλημα είναι γενικότερα γνωστό ως πρόβλημα εξεχόντων περιγραμμάτων (salient contours problem). Η επίλυση αυτού του προβλήματος θα περιελάμβανε τον ορισμό μιας τιμής σε κάθε ακμή της εικόνας, η τιμή οποία θα σχετίζεται με το εάν η εκάστοτε ακμή αποτελεί τμήμα κάποιου αντικειμένου. Ανάλογα με την κατανομή αυτών των τιμών, είναι συνήθως δυνατό να επιλεγεί ένα κατώφλι, το οποίο θα διαχωρίσει το σύνολο των ακμών της εικόνας σε κλάσεις σχήματος και φόντου.

2.2 Νευροφυσιολογικά στοιχεία του ΑΟΣ

Σε αντίθεση με τα τεχνητά συστήματα όρασης, τα βιολογικά συστήματα επιλύουν με ιδιαίτερη ευκολία το πρόβλημα των εξεχόντων περιγραμμάτων. Το ΑΟΣ μπορεί να διακρίνει τα περιγράμματα αντικειμένων άσχετα με το πόσο ανομοιόμορφο είναι το φόντο στο οποίο βρίσκονται. Η διαδικασία εξαγωγής των εξεχόντων περιγραμμάτων φαίνεται να συμβαίνει ακούσια, χωρίς να προϋποθέτει το συστηματικό έλεγχο της εικόνας, ούτε προηγούμενη γνώση του σχήματος των αντικειμένων. Αυτό υποδεικνύει ότι η αντίληψη των εξεχόντων περιγραμμάτων είναι μια άμεση διαδικασία χαμηλού επιπέδου του ΑΟΣ, η οποία δεν περιλαμβάνει νοητικές διεργασίες υψηλού επιπέδου. Αυτό φαίνεται ξεκάθαρα στην Εικόνα 2.2.

Εικόνα 2.2

Ένα σύνολο ευθύγραμμων τμημάτων με τυχαίες κλίσεις, ορισμένα από τα οποία σχηματίζουν μια έλλειψη. Ο τυπικός ανθρώπινος παρατηρητής μπορεί εύκολα να τη διακρίνει χωρίς ιδιαίτερη προσπάθεια, έστω και αν τα τμήματα που την αποτελούν δε διαφέρουν σε αντίθεση από όλα τα άλλα.



Τα τελευταία είκοσι χρόνια οι νευροφυσιολογικές έρευνες του ΑΟΣ άρχισαν να αποκωδικοποιούν τα νευρικά κυκλώματα που σχετίζονται με το μηχανισμό εξαγωγής των εξεχόντων περιγραμμάτων. Έτσι, σκιαγραφείται ο τρόπος με τον οποίο τα βιολογικά οπτικά συστήματα αντιμετωπίζουν το πρόβλημα αυτό. Τα αποτελέσματα όλων των νευροφυσιολογικών ερευνών συγκλίνουν στο ότι οι νευρώνες του πρωτοταγούς οπτικού φλοιού ενσωματώνουν οπτική πληροφορία και έξω από το κλασικό υποδεκτικό τους πεδίο (receptive field), με τέτοιο τρόπο ώστε να ευνοείται η εξαγωγή των εξεχόντων περιγραμμάτων (Field, et al., 2004). Πιο συγκεκριμένα, έχει αποδειχθεί ότι η μεγάλης εμβέλειας οριζόντιες διασυνδέσεις του πρωτοταγούς οπτικού φλοιού ευθύνονται για τη ρύθμιση των κατωφλίων ανίχνευσης αντίθεσης και κατ' επέκταση, για την εξαγωγή εξεχόντων περιγραμμάτων(Newsome, et al., 1989),(Watamaniuk, et al., 1992), (Kovács, et al., 1993), (Field, et al., 1993), (Kovács, et al., 1994), (Pettet, et al., 1998), (Hess, et al., 2003),(Ben-Shahar, et al., 2003). Αυτό σημαίνει ότι οι νευρώνες της περιοχής V1, οι οποίοι είναι οι κωδικοποιητές της κλίσης των ακμών του ΑΟΣ, μεταβάλλουν τις αποκρίσεις και την ευαισθησία τους ανάλογα με τα γειτονικά, στο υποδεκτικό τους πεδίο, οπτικά ερεθίσματα. Εάν ένας V1 νευρώνας διεγερθεί από μια ακμή συγκεκριμένης κλίσης, τότε μια δεύτερη συγγραμμική (collinear) ακμή σε γειτονικό σημείο του μπορεί να αυξήσει την απόκριση αυτού του νευρώνα. Αντίθετα, μια γειτονική ακμή με κάθετη κλίση στον άξονα της ακμής του νευρώνα θα προκαλέσει αναστολή ή το λιγότερο μικτότερη απόκριση (Rockland, et al., 1982),(Rockland, et al., 1983),(Kapadia, et al., 1995),(Singer, et al., 1995),(Gilbert, et al., 1996),(Fitzpatrick, 1996),(Polat, et al., 1998),(Kapadia, et al., 2000). Αυτή η συμπεριφορά των νευρώνων υποδεικνύει την ύπαρξη, παράλληλα με το κλασσικό υποδεκτικό τους πεδίο, ενός δεύτερου πεδίου που ονομάζεται πεδίο συσχέτισης (association field) (Field, et al., 1993). Η Εικόνα 2.3 παρουσάζει το πεδίο συσχέτισης ενός νευρώνα οριζόντιας κλίσης.



Εικόνα 2.3

Μέρος του πεδίου συσχέτισης ενός νευρώνα οριζόντιας κλίσης (Field, et al., 1993). Ο νευρώνας συνδέεται διεγερτικά με γειτονικούς νευρώνες, των οποίων οι κλίσεις σχηματίζουν καμπύλες που διέρχονται από τον άξονά του. Αντίθετα αναστέλλεται από νευρώνες κάθετης κλίσης στον άξονά του.

Η εγκυρότητα των μελετών αυτών επαληθεύεται από τα αποτελέσματα της μελέτης (Geisler, et al., 2001), στην οποία ακολουθήθηκε ακριβώς η αντίστροφη διαδικασία: αναλύθηκε η χωρική στατιστική κατανομή της σύμπτωσης ακμών από ένα σύνολο εικόνων που περιείχαν φυσικά αντικείμενα. Η κατανομή που προέκυψε φαίνεται στην Εικόνα 2.4 και ουσιαστικά δείχνει την προτίμηση σε θέση και κλίση που έχει μια οριζόντια ακμή στο να συνδέεται με άλλες ακμές για το σχηματισμό φυσικών περιγραμμάτων.



Εικόνα 2.4

Η χωρική στατιστική κατανομή που προέκυψε για μια ακμή κλίσης 0°(Geisler, et al., 2001). Στο κέντρο της κατανομής βρίσκεται η οριζόντια ακμή. Η κλίση των ευθύγραμμων τμημάτων σε κάθε θέση, δείχνει την κλίση της ακμής που είναι πιθανότερο να βρεθεί στο σημείο εκείνο. Το χρώμα των ευθύγραμμων τμημάτων καθορίζει την πιθανότητα εμφάνισής τους στο εκάστοτε σημείο.

Είναι φανερό από την κατανομή της Εικόνας 2.4 ότι οι ακμές των φυσικών αντικειμένων συνδέονται με αμβλείες γωνίες έτσι ώστε να ευνοούν τα ομαλά περιγράμματα αποφεύγοντας τις οξείες γωνίες. Σημαντική είναι η ομοιότητα της χωρικής κατανομής των φυσικών περιγραμμάτων (Εικόνα 2.4) με το σχήμα του πεδίου συσχέτισης (Εικόνα 2.3) στην περιοχή V1 του ΑΟΣ. Αυτό δείχνει ότι στη διάρκεια εξέλιξης των βιολογικών οπτικών συστημάτων, οι συνάψεις των νευρώνων προσαρμόστηκαν με τέτοιο τρόπο ώστε να κωδικοποιούν τη χωρική κατανομή των ακμών για τα φυσικά περιγράμματα. Έτσι, όταν ενεργοποιούνται οι νευρώνες κλίσης από τις ακμές ενός αντικειμένου, η τελική απόκρισή τους διαμορφώνεται από τις συνάψεις του πεδίου συσχέτισης. Αν η κλίση ενός

διεγερμένου νευρώνα συμβαδίζει με αυτή του πεδίου συσχέτισης ενός άλλου, τότε οι δύο νευρώνες συγχρονίζουν τα δυναμικά ενέργειάς τους έχοντας παρόμοια συχνότητα εκπόλωσης. Ο συγχρονισμός αυτός μπορεί να σχηματιστεί σε ολόκληρες αλυσίδες νευρώνων και είναι υπεύθυνος για την εξαγωγή των εξεχόντων περιγραμμάτων. Σε διαφορετική περίπτωση, όπου τα πεδία συσχέτισης δεν ταιριάζουν, τα δυναμικά δε συγχρονίζονται και οι νευρώνες αναστέλλουν την απόκρισή τους.

Είναι σημαντικό να επισημανθεί ότι όλοι οι νευρώνες κάθε υπερστήλης διαθέτουν πεδίο συσχετισμού ανάλογο με την κλίση της ακμής που κωδικοποιούν. Έτσι, η ταυτόχρονη αλληλεπίδραση πολλών πεδίων συσχέτισης σε κάθε θέση του οπτικού πεδίου και για κάθε κλίση ακμής οδηγεί στο συγχρονισμό μόνο των νευρώνων που διεγείρονται από εξέχοντα περιγράμματα, ενώ παράλληλα όλοι οι άλλοι νευρώνες παραμένουν αποσυγχρονισμένοι στις αποκρίσεις τους. Αυτός ο μηχανισμός εξαγωγής φαίνεται στην Εικόνα 2.5.

Εικόνα 2.5

Τα υποδεκτικά πεδία Α-Β-Γ-Δ είναι έτσι διατεταγμένα ώστε τα πεδία συσχέτισής τους να συνδέονται. Ως αποτέλεσμα, τα δυναμικά των αντίστοιχων νευρώνων συγχρονίζονται και σχηματίζουν την καμπύλη ΑΒΓΔ. Αντίθετα, το πεδίο συσχέτισης του υποδεκτικού πεδίου Ε δε συνδέεται με κανένα από τα παραπάνω, με αποτέλεσμα να μη συμμετέχει στην αλυσίδα της καμπύλης.



2.3 Ανασκόπηση υπάρχουσας βιβλιογραφίας

Η υπάρχουσα έρευνα στο πρόβλημα των εξεχόντων περιγραμμάτων έχει επικεντρωθεί σε δύο διαφορετικές κατευθύνσεις: τις συμβατικές και τις βιολογικά εμπνευσμένες μεθόδους (κυρίως από το ΑΟΣ). Οι διαφορές μεταξύ των δύο κατηγοριών είναι σημαντικές τόσο στην προσέγγιση του προβλήματος όσο και στους στόχους τους. Οι συμβατικές μέθοδοι έχουν σαν κύριο στόχο την πρακτική λύση του προβλήματος που παρουσιάστηκε στην ενότητα 2.1, για εφαρμογές σε όραση μηχανής και επεξεργασία εικόνας. Οι προσεγγίσεις τους στο πρόβλημα ποικίλουν και προέρχονται κυρίως από το χώρο των μαθηματικών (π.χ. θεωρία γράφων). Οι βιολογικά εμπνευσμένες μέθοδοι έχουν σαν κύριο στόχο τους τη μαθηματική μοντελοποίηση των δεδομένων που παρουσιάζονται στις νευροφυσιολογικές μελέτες του ΑΟΣ. Ουσιαστική επιδίωξή τους είναι η όσο το δυνατό πιο πιστή αντιγραφή της προσέγγισης που η εξέλιξη έχει επιλέξει για τη λύση του προβλήματος των εξεχόντων περιγραμμάτων. Παρ' όλες τις διαφορές μεταξύ των δύο κατηγοριών, και οι δύο έχουν ένα κοινό βασικό χαρακτηριστικό: ενσωματώνουν τις αρχές Gestalt της ομαλής συνέχειας (Εικόνα 2.6iv) και εγγύτητας (Εικόνα 2.6ii). Η σχολή Gestalt είναι ένας κλάδος της ψυχολογίας που αναπτύχθηκε στην Αυστρία και τη Γερμανία στα τέλη του 19^{ου} αιώνα. Βασική του ιδέα είναι ότι η αντίληψη στο σύνολό της είναι διαφορετική από το άθροισμα των επιμέρους αντιλήψεων του συνόλου των ερεθισμάτων που δέχεται ο εγκέφαλος (Εικόνα 2.6i). Οι ψυχολόγοι της σχολής Gestalt διατύπωσαν ένα σύνολο αρχών που διέπουν την ανθρώπινη αντίληψη. Οι αρχές αυτές εκφράζονται με εμφανή τρόπο και στην οπτική αντίληψη, όπως φαίνεται στην Εικόνα 2.6.



Εικόνα 2.6

i. Το σύνολο είναι διαφορετικό από το άθροισμα των στοιχείων του.

ii. Αρχή της εγγύτητας: Στοιχεία που βρίσκονται κοντά, ομαδοποιούνται.

iii. Αρχή της ομοιότητας: Τα όμοια στοιχεία ομαδοποιούνται.

iv. Αρχή της ομαλής συνέχειας: Αντιλαμβανόμαστε ΑΒ-ΓΔ, και όχι ΑΓ-ΔΒ ή ΑΔ-ΓΒ.

ν. Αρχή της ολοκλήρωσης: Τα ελλιπή περιγράμματα συμπληρώνονται.

vi. Αρχή της απλότητας: Επικρατεί ο απλούστερος συνδυασμός.

Μια από τις πρώτες προσεγγίσεις στην κατηγορία των συμβατικών μεθόδων παρουσιάστηκε από τους Sha'shua και Ullman (Sha'ashua, et al., 1988), οι οποίοι δώσανε για πρώτη φορά μια ολοκληρωμένη διατύπωση για το πρόβλημα των εξεχόντων περιγραμμάτων. Η μέθοδος βασίζεται σε ένα δίκτυο από τοπικά συνδεδεμένες μονάδες επεξεργασίας ανάλογο με το μέγεθος της εικόνας. Για κάθε εικονοστοιχείο της εικόνας υπάρχουν 16 διαφορετικές μονάδες επεξεργασίας, κάθε μια από τις οποίες επεξεργάζεται μια συγκεκριμένη κλίση ακμών στο συγκεκριμένο σημείο. Είσοδος του συστήματος είναι μια δυαδική εικόνα ακμών η οποία ενεργοποιεί εκείνες τις μονάδες επεξεργασίας που η κλίση τους είναι τοπικά ίδια με την κλήση των ακμών. Οι ενεργοποιημένες μονάδες επεξεργασίας αλληλεπιδρούν τοπικά μεταξύ τους βάσει των αρχών Gestalt της συνέχειας

Συμβατικές μέθοδοι

Αρχές Gestalt

και της εγγύτητας. Μετά από την πάροδο ενός αριθμού επαναλήψεων, το δίκτυο συγκλίνει εξάγοντας το πιο εξέχον περίγραμμα της εικόνας. Μια εκτεταμένη ανάλυση της μεθόδου αυτής παρουσιάστηκε από τους Alter και Basri (Alter, et al., 1998). Ένα βασικό μειονεκτήματα της μεθόδου είναι ότι ενώ εξάγει σωστά το πιο εξέχον περίγραμμα της εικόνας, παρουσιάζει προβλήματα στην εξαγωγή των υπόλοιπων εξεχόντων περιγραμμάτων. Αυτό σημαίνει ότι εάν μια εικόνα περιέχει περισσότερα του ενός εξέχοντα περιγράμματα, θα εξαχθεί μόνο ένα. Οι πιο πρόσφατες μέθοδοι αυτής της κατηγορίας εστιάζουν στη χρήση των γράφων. Για κάθε εικόνα ακμών σχηματίζεται ένας γράφος, τα σημεία διακλάδωσής του οποίου συμβολίζουν εικονοστοιχεία ή περιοχές της εικόνας. Οι γραμμές του γράφου που ενώνουν τα σημεία διακλάδωσης δηλώνουν τις ακμές τις εικόνας. Στο πλαίσιο αυτό, η εύρεση ενός περιγράμματος ανάγεται στο να διαχωριστεί ο γράφος σε τμήματα, έτσι ώστε να βελτιστοποιείται μια συνάρτηση κόστους. Οι Williams και Thornber (Williams, et al., 2000) παρουσίασαν μια μέθοδο βασισμένη σε αυτή την προσέγγιση. Για κάθε ακμή της εικόνας, όρισαν ως βαθμό εξέχοντος περιγράμματος της ακμής, τον αριθμό των κλειστών διαδρομών πάνω στο γράφο, οι οποίες διέρχονται από τη συγκεκριμένη ακμή. Όσο περισσότερες κλειστές διαδρομές διέρχονται από τα εικονοστοιχεία μιας ακμής πάνω στον γράφο, τόσο αυξάνεται η πιθανότητα η συγκεκριμένη ακμή να είναι μέρος ενός εξέχοντος περιγράμματος. Σύμφωνα με τους συγγραφείς, τα αποτελέσματα της μεθόδου είναι καλύτερα από αυτά των Sha'shua και Ullman (Sha'ashua, et al., 1988). Ο Mahamud (Mahamud, et al., 2003) παρουσίασε μια εξέλιξη της μεθόδου των Williams και Thornber, στην οποία χρησιμοποιείται ένα παρόμοιο μέτρο για τον καθορισμό του βαθμού συμμετοχής της κάθε ακμής σε εξέχοντα περιγράμματα. Επιπλέον, παρουσιάζεται μια μέθοδος δυαδικοποίησης του βαθμού της κάθε ακμής, έτσι ώστε να εξαχθούν τα τελικά εξέχοντα περιγράμματα. Πιο πρόσφατα, ο Wang (Wang, et al., 2005) παρουσίασε μια νέα μέθοδο εξαγωγής εξεχόντων περιγραμμάτων, βασισμένη και αυτή σε γράφους. Ως είσοδος θεωρείται το αποτέλεσμα ενός εξαγωγέα ακμών, συνήθως της μεθόδου Canny. Η μέθοδος συνδέει διαδοχικά όλες τις ακμές της εικόνας και ελέγχει το αποτέλεσμα βάσει ενός μέτρου το οποίο στηρίζεται στις αρχές Gestalt της ομαλής συνέχειας και εγγύτητας. Οι συγγραφείς παρουσιάζουν βελτιωμένα αποτελέσματα της μεθόδου τους σε σχέση με τις προηγούμενες μεθόδους. Το βασικότερο μειονέκτημα της κατηγορίας αυτής των μεθόδων είναι η πολυπλοκότητα και κατά συνέπεια ο χρόνος εκτέλεσης του αλγορίθμου, λόγω του γεγονότος ότι ανιχνεύουν μεταξύ όλων των πιθανών συνδυασμών των ακμών. Ως αποτέλεσμα, ο χρόνος εκτέλεσης μεταβάλλεται ανάλογα με τον αριθμό των εξεχόντων περιγραμμάτων και σε περίπτωση που τα εξέχοντα περιγράμματα είναι περισσότερα του ενός, ο χρόνος αυτός γίνεται πολύ μεγάλος. Η μέθοδος με το συντομότερο χρόνο εκτέλεσης είναι αυτή των (Mahamud, et al., 2003) η οποία χρειάζεται 10 περίπου δευτερόλεπτα για κάθε εξέχον περίγραμμα σε εικόνα μεγέθους 480×480 εικονοστοιχείων. Οι υπόλοιπες μέθοδοι παρουσιάζουν τα

Βιολογικά εμπνευσμένες μέθοδοι Οι βιολογικά εμπνευσμένες μέθοδοι για την εξαγωγή των εξεχόντων περιγραμμάτων είναι κυρίως μαθηματικά μοντέλα που προσπαθούν να αναπαράγουν ακριβώς τα δεδομένα των νευροφυσιολογικών μελετών του ΑΟΣ. Οι περισσότερες δεν ασχολούνται με πρακτικά θέματα, όπως πολυπλοκότητα και χρόνος εκτέλεσης, ενώ τις περισσότερες φορές τα

αποτελέσματά τους σε εικόνες το πολύ 200×200 εικονοστοιχείων.

αποτελέσματά τους παρουσιάζονται για τεχνητές εικόνες, παρόμοιες με την Εικόνα 2.2. Οι Yen και Finkel (Yen, et al., 1998) παρουσίασαν ένα μοντέλο το οποίο βασίζεται στο συγχρονισμό των νευρώνων κλίσης του πρωτοταγούς οπτικού φλοιού. Το μοντέλο καθορίζει επαναληπτικά το βαθμό εξεχόντων περιγραμμάτων με τη χρήση γενικών κανονικοποιήσεων. Οι συνδέσεις των νευρώνων είναι όμοιες με το πεδίο συσχετίσεως, όπως παρουσιάζεται στην εργασία (Field, et al., 1993). Για τον υπολογισμό τους χρησιμοποιείται η συν-κυκλική (co-circular) προσέγγιση που προτάθηκε από τους Parent και Zucker (Parent, et al., 1989). Σύμφωνα με την προσέγγιση αυτή, δύο ευθύγραμμα τμήματα ανήκουν στην ίδια καμπύλη αν εφάπτονται στην ίδια περιφέρεια κύκλου. Η Li (Li, 1998) παρουσίασε ένα άλλο μοντέλο, βασισμένο και αυτό στο συγχρονισμό των νευρώνων κλίσης, το οποίο όμως προβλέπει ανάδραση από υψηλότερες οπτικές περιοχές, καθώς και εξαγωνικό πλέγμα υπερστηλών. Ο Grigorescu (Grigorescu, et al., 2003), ανέπτυξε μια νέα μέθοδο για εξαγωγή εξεχόντων περιγραμμάτων, η οποία βασίζεται στην πλευρική αναστολή (lateral inhibition) των νευρώνων κλίσης του ΑΟΣ. Χρησιμοποιούνται φίλτρα Gabor 12 διαφορετικών κλίσεων και η έξοδος της μεθόδου υπόκειται σε επιπλέον επεξεργασία για την ομαλοποίηση των περιγραμμάτων του τελικού αποτελέσματος. Ο Mundhenk (Mundhenk, et al., 2003), (Mundhenk, et al., 2005), παρουσίασε ένα νέο μοντέλο για την εξαγωγή των εξεχόντων περιγραμμάτων, το οποίο ονομάζεται CINNIC (Carefully Implemented Neural Network for Integrating Contours – Προσεκτικά Σχεδιασμένο Νευρωνικό Δίκτυο για την Ενοποίηση Περιγραμμάτων). Χρησιμοποιούνται για πρώτη φορά τρεις διαφορετικές χωρικές κλίμακες επεξεργασίας, διαρρέοντες νευρώνες ενοποίησης (leaky integrator neurons) και ο όρος ταχείας πλαστικότητας (fast plasticity term). Το πεδίο συσχετισμού που χρησιμοποιείται είναι αυτό που έχει προταθεί από τον Braun (Braun, 1999). Το CINNIC και η μέθοδος (Grigorescu, et al., 2003) είναι οι μόνες μέθοδοι για τις οποίες παρουσιάζονται εκτεταμένα αποτελέσματα για πραγματικές εικόνες. Οι Grossberg και Mingolla παρουσίασαν σε μια εκτεταμένη σειρά άρθρων (Grossberg, et al., 1985), (Grossberg, 1994), (Grossberg, et al., 1995), (Grossberg, 2004), μια ολοκληρωμένη θεωρία για ενοποίηση περιγραμμάτων και εξαγωγή εξεχόντων ακμών, η οποία ονομάζεται BCS (Boundary Contour System – Σύστημα Περιγραμμάτων Ορίων). Εισάγουν για πρώτη φορά την έννοια των δίπολων κυττάρων (bipole cells) και των Τενώσεων (T-junctions), τα οποία συνδυάζουν εισόδους και από τις δύο κατευθύνσεις του περιγράμματος για την ομαλή εξαγωγή τους. Ο αλγόριθμος μπορεί να εξάγει ψευδαισθητικά περιγράμματα (illusionary contours) όπως αυτά στην Εικόνα 2.6ν. Εντούτοις, δεν παρουσιάζονται εκτεταμένα αποτελέσματα για πραγματικές εικόνες.

2.4 Μοντελοποίηση κυττάρων κλίσης

Ένα βασικό στοιχείο κάθε μεθόδου εμπνευσμένης από βιολογικά οπτικά συστήματα είναι η μοντελοποίηση που ακολουθείται για τα απλά κύτταρα κλίσης του πρωτοταγούς οπτικού φλοιού.

2.4.1 Υπάρχουσα μοντελοποίηση κυττάρων κλίσης

Οι νευροφυσιολογικές έρευνες στον πρωτοταγή οπτικό φλοιό πρωτευόντων θηλαστικών έδειξαν ότι σε κάθε θέση του οπτικού πεδίου υπάρχουν δύο πληθυσμοί νευρώνων που

ανιχνεύουν κλίσεις ακμών (Hubel, et al., 1965), (Hubel, et al., 1968). Ορισμένα από τα υποδεκτικά πεδία των κυττάρων αυτών απεικονίζονται στην Εικόνα 2.7i και 2.7ii. Στην πρώτη κατηγορία ανήκουν υποδεκτικά πεδία εξειδικευμένα να ανιχνεύουν γραμμές μεταβολής έντασης σε διάφορες κλίσεις (συνήθως 12) και σε διαφορετικές αντιθέσεις (φωτεινή γραμμή σε σκοτεινό φόντο και σκοτεινή γραμμή σε φωτεινό φόντο). Στη δεύτερη κατηγορία ανήκουν υποδεκτικά πεδία που ανιχνεύουν μεταβάσεις έντασης από φωτεινό σε σκοτεινό και σε διάφορες κλίσεις (συνήθως 12).



Εικόνα 2.7

i,ii. Υποδεκτικά πεδία ανίχνευσης κλίσης ακμών (απεικονίζονται μόνο 3 από τις 12 κλίσεις).

iii,iv. Μοντελοποίηση της κατηγορίας i από άρτια φίλτρα Gabor και της κατηγορίας ii από περιττά φίλτρα Gabor. Γκρι: μηδενικές τιμές Άσπρο: θετικές τιμές Μαύρο: αρνητικές τιμές

> Τα κύτταρα και των δύο κατηγοριών μοντελοποιούνται από οικογένειες φίλτρων Gabor δύο διαστάσεων (Jones, et al., 1987). Τα φίλτρα Gabor είναι το γινόμενο μιας Γκαουσιανής κατανομής με ένα συνημιτονοειδές, όπως φαίνεται από της εξίσωση (2.1).

$$G_{\lambda,\sigma,\theta,\varphi}(x,y) = e^{-\frac{\tilde{x}^2 + \gamma^2 \tilde{y}^2}{2\sigma^2}} \cos\left(2\pi \frac{\tilde{x}}{\lambda} + \varphi\right)$$

$$\tilde{x} = x \cos\theta + y \sin\theta$$

$$\tilde{y} = -x \sin\theta + y \cos\theta$$

$$(x,y) \in \Omega \subset \mathbb{R}^2$$

$$(2.1)$$

Η τιμή γ είναι μια σταθερά που ονομάζεται λόγος χωρικής όψης (spatial aspect ratio) και καθορίζει την ελλειπτικότητα του υποδεκτικού πεδίου. Η τυπική απόκλιση σ της Γκαουσιανής κατανομής καθορίζει το μέγεθος του υποδεκτικού πεδίου. Η παράμετρος λ είναι το μήκος κύματος και $1/\lambda$ η χωρική συχνότητα του συνημίτονου. Ο λόγος σ/λ καθορίζει το εύρος χωρικής συχνότητας (spatial frequency bandwidth) και κατά συνέπεια τον αριθμό των παράλληλων διεγερτικών και ανασταλτικών ραβδώσεων (οι λευκές και μαύρες λωρίδες στην Εικόνα 2.7iii). Η παράμετρος θ , $\theta \in [0, \pi)$ καθορίζει την κλίση του φίλτρου, ενώ η παράμετρος φ , $\varphi \in (-\pi, \pi]$, καθορίζει τη μετατόπιση φάσης του φίλτρου.

αυτή, αποκρίνεται καλύτερα σε γραμμές μεταβολής έντασης και για το λόγο αυτό χρησιμοποιείται για τη μοντελοποίηση των υποδεκτικών πεδίων της πρώτης κατηγορίας (Εικόνα 2.7i). Για $\varphi = -(\pi/2)$ και $\varphi = (\pi/2)$, είναι αντισυμμετρικό ή περιττό (Εικόνα 2.7iv). Στην περίπτωση αυτή, υπάρχει μεγαλύτερη απόκριση για μεταβάσεις έντασης και κατά συνέπεια, χρησιμοποιείται για τη μοντελοποίηση των υποδεκτικών πεδίων της δεύτερης κατηγορίας (Εικόνα 2.7ii). Όλες οι άλλες τιμές φάσης δίνουν αντισυμμετρική μορφή στο φίλτρο. Η απόκριση $r_{\lambda,\sigma,\theta,\varphi}(x,y)$ ενός απλού κυττάρου κλίσης, που μοντελοποιείται από ένα φίλτρο Gabor $G_{\lambda,\sigma,\theta,\varphi}(x,y)$ σε μια εικόνα εισόδου με κατανομή έντασης f(x, y), υπολογίζεται από την παρακάτω συνέλιξη.

$$r_{\lambda,\sigma,\theta,\varphi}(x,y) = \left(f * G_{\lambda,\sigma,\theta,\varphi}\right)(x,y) = \iint_{\Omega} f(x,y) G_{\lambda,\sigma,\theta,\varphi}(x-u,y-v) du dv$$
(2.2)

Για τη σωστή εξαγωγή της κλίσης των ακμών, θα πρέπει σε κάθε σημείο του οπτικού πεδίου να υπάρχει η δυνατότητα ανίχνευσης τόσο γραμμών όσο και μεταβάσεων και για τις δύο πολικότητες. Τα φίλτρα των Εικόνων 2.7iii και 2.7iv μπορούν να ανιχνεύσουν ακμές συγκεκριμένης μόνο πολικότητας, π.χ. φωτεινή γραμμή σε σκούρο φόντο. Σε περίπτωση αντίθετης πολικότητας ακμών, δίνουν αρνητική τιμή. Για το λόγο αυτό, το αποτέλεσμα της απόκρισης $r_{\lambda,\sigma,\theta,\varphi}(x,y)$ υψώνεται στο τετράγωνο, ώστε να περιλαμβάνονται και οι δύο πολικότητες. Επιπλέον, σε κάθε σημείο εφαρμόζονται ταυτόχρονα ένα άρτιο και ένα περιττό φίλτρο, ώστε να υπάρχει η δυνατότητα για ανίχνευση γραμμών και μεταβάσεων. Η τελική έξοδος των μοντελοποιημένων κυττάρων κλίσης του πρωτοταγούς οπτικού φλοιού δίνεται από την εξίσωση (2.3), η οποία ονομάζεται ενέργεια Gabor.

$$E_{\lambda,\sigma,\theta}(x,y) = \sqrt{r_{\lambda,\sigma,\theta,0}^2(x,y) + r_{\lambda,\sigma,\theta,-\frac{\pi}{2}}^2(x,y)}$$
(2.3)

Όλες οι βιολογικά εμπνευσμένες μέθοδοι, που παρουσιάστηκαν στην προηγούμενη ενότητα, χρησιμοποιούν τη μοντελοποίηση των κυττάρων κλίσης από φίλτρα Gabor. Για να μπορέσουν να εξαχθούν όλες οι ακμές της εικόνας και για τις 12 κλίσεις μιας συγκεκριμένης κλίμακας, θα πρέπει η εικόνα να συνελιχθεί με 24 συνολικά φίλτρα Gabor (12 άρτια και 12 περιττά). Μια τέτοια πράξη είναι αρκετά απαιτητική σε υπολογιστικούς πόρους και ως αποτέλεσμα, ένα σημαντικό μέρος του χρόνου εκτέλεσης των βιολογικά εμπνευσμένων μεθόδων οφείλεται στην πράξη της συνέλιξης.

2.4.2 Προτεινόμενη μοντελοποίηση κυττάρων κλίσης

Συστήματα, που έχουν ως στόχο την αντιγραφή σε κάποιο επίπεδο της λειτουργίας του πρωτοταγούς οπτικού φλοιού, πρέπει να διαθέτουν ένα αμετάβλητο πλέγμα διακριτών θέσεων και κλίσεων, το οποίο θα αντιστοιχεί στις υπερστήλες κλίσης του ΑΟΣ. Αυτό το πλέγμα των διακριτών θέσεων εισάγει ένα σημαντικό πρόβλημα στην εξαγωγή των ακμών, αφού δεν επιτρέπει τη συνέλιξη μιας μάσκας σε όλες τις πιθανές θέσεις της εικόνας παρά μόνο στη θέση κάθε υπερστήλης. Αυτό το πρόβλημα γίνεται πιο φανερό όταν οι διακριτές θέσεις των υπερστηλών είναι μη επικαλυπτόμενες. Για τον περιορισμό των προβλημάτων της πολυπλοκότητας στη συνέλιξη και της ασυμβατότητας με το πλέγμα των διακριτών θέσεων, προτείνεται μια νέα προσέγγιση η οποία είναι συμβατή με την αρχιτεκτονική του πρωτοταγούς οπτικού φλοιού και βελτιώνει το χρόνο εκτέλεσης. Σύμφωνα με την προσέγγιση αυτή, αντί να μετατοπίζεται η μάσκα μιας κλίσης σε όλες τις πιθανές θέσεις της εικόνας, μετατοπίζεται το ευθύγραμμο τμήμα της μάσκας σε κάθε θέση του εμβαδού των διακριτών θέσεων του πλέγματος. Αυτό γίνεται εφικτό με τη χρήση διαφορετικών μασκών ίδιας κλίσης και διαφορετικής θέσης του ευθύγραμμου τμήματος ανίχνευσης. Στη συνέχεια ακολουθεί η περιγραφή του σχεδιασμού αυτών των μασκών.

Το σετ μασκών που θα προκύψει από τη διαδικασία σχεδιασμού, θα πρέπει να διέπεται από τους ακόλουθους βασικούς κανόνες:

- Οι μάσκες πρέπει να είναι τετράγωνες για να ευνοούν την εφαρμογή τους σε μη επικαλυπτόμενες θέσεις και δυαδικές για να μειωθεί η πολυπλοκότητα των υπολογισμών.
- Πρέπει να έχουν αρκετό μέγεθος (ανάλυση), ώστε να περιλαμβάνουν όλες τις πιθανές κλίσεις του ευθύγραμμου τμήματος ανίχνευσης.
- Για κάθε κλίση πρέπει να περιλαμβάνονται όλες οι πιθανές θέσεις του ευθύγραμμου τμήματος μέσα στο μέγεθος της μάσκας.
- Ο συνολικός αριθμός των μασκών, πρέπει να είναι όσο το δυνατό μικρότερος, ώστε να περιοριστεί το υπολογιστικό κόστος της συνέλιξης στο ελάχιστο.

Σημαντικό ρόλο στη λειτουργικότητα του σετ των μασκών παίζει ο λόγος της διάστασης n των μασκών (σημειώνουμε ότι βάσει του πρώτου κανόνα οι μάσκες είναι τετράγωνες με διάσταση $n \times n$ εικονοστοιχεία) και του εύρους e του ευθύγραμμου τμήματος, εκφρασμένα και τα δύο σε εικονοστοιχεία (Εικόνα 2.8). Ο λόγος αυτός επηρεάζει άμεσα το συνολικό αριθμό μασκών M του σετ που απαιτείται για την κάλυψη όλων των πιθανών θέσεων του ευθύγραμμου τμήματος μέσα στη μάσκα. Όσο μεγαλύτερο το M, τόσο μεγαλύτερο θα είναι και το υπολογιστικό κόστος.

Εικόνα 2.8

Τα 3 βασικά μεγέθη στο σχεδιασμό του σετ μασκών.



Για να μελετηθεί η συμπεριφορά του λόγου *n/e* στη λειτουργικότητα του σετ μασκών, δημιουργήθηκαν πειραματικά σετ με διαφορετικούς λόγους *n/e* και διαφορετικές βηματικές γωνίες θ (η γωνία που καθορίζει το βήμα με το οποίο αυξάνονται οι κλίσεις του ευθύγραμμου τμήματος). Τα χαρακτηριστικά αυτών των σετ παρουσιάζονται στον Πίνακα 2.1.

Σετ	Διάσταση <i>n</i>	Εύρος <i>e</i>	Βηματική γωνία <i>θ</i>	Αριθμός μασκών <i>Μ</i>
1	10	2	15°	60
2	10	2	30°	30
3	10	1	15°	120
4	10	1	7,5°	240
5	8	1	9°	160
6	8	1	15°	95
7	8	2	15°	50
8	12	2	15°	70

Πίνακας 2.1

Πειραματικά σετ μασκών για τη μελέτη των αποτελεσμάτων του λόγου n/e στο συνολικό αριθμό απαιτούμενων μασκών.

Η ανάλυση των πειραματικών σετ έγινε με τον έλεγχο του βαθμού ικανοποίησης των αρχικών κανόνων σχεδιασμού. Διαπιστώθηκε ότι τα σετ με λόγους n/e μικρότερους του 3, παρουσιάζουν σημαντική υποβάθμιση στην ποιότητα του ευθύγραμμου τμήματος. Αυτό συμβαίνει γιατί οι διαστάσεις n και e γίνονται συγκρίσιμες, με αποτέλεσμα το ευθύγραμμου τμήμα να διαστρεβλώνεται από το πλέγμα των εικονοστοιχείων της μάσκας. Συνεπώς, όσο μεγαλύτερος είναι ο λόγος n/e, τόσο καλύτερη και η ποιότητα του ευθύγραμμου τμήματος της μάσκας. Είναι ακόμη φανερό ότι το εύρος του ευθύγραμμου τμήματος της μάσκας. Είναι ακόμη φανερό ότι το εύρος του ευθύγραμμου τμήματος της μάσκας. Είναι ακόμη φανερό ότι το εύρος του ευθύγραμμου τμήματος της μάσκας. Είναι ακόμη φανερό ότι το εύρος του ευθύγραμμου τμήματος της μάσκας. Είναι ακόμη φανερό ότι το εύρος του ευθύγραμμου τμήματος n και 7. Αναλόγως, όταν η βηματική γωνία θ μειώνεται στο μισό, το M επίσης διπλασιάζεται (Πίνακας 2.1: σετ 1 και 2, 3 και 4). Αντίθετα, μια αλλαγή στο μέγεθος των μασκών n, μεταβάλει τον απαιτούμενο αριθμό μασκών M κατά τον ίδιο παράγοντα, υποδεικνύοντας ότι τα μεγέθη n και M σχετίζονται γραμμικά (Πίνακας 2.1: σετ 1 και 7, 1 και 8). Τα παραπάνω συμπεράσματα μπορούν να εκφραστούν από την εξίσωση (2.4).

$$M(n, e, \theta) = \begin{cases} \frac{180 \cdot n}{e \cdot \theta} & \epsilon \acute{\alpha} v \frac{n}{e} \ge 3\\ \chi \alpha \mu \eta \lambda \acute{\eta} \pi \sigma i \acute{\sigma} \tau \eta \tau \alpha & \epsilon \acute{\alpha} v \frac{n}{e} < 3 \end{cases}$$
(2.4)

Η σταθερά 180 είναι απαραίτητη για το συσχετισμό των πειραματικών δεδομένων του Πίνακα 2.1 σε μια εξίσωση. Εκτός των άλλων, ο λόγος n/e επηρεάζει επίσης την ελάχιστη τιμή της βηματικής γωνίας θ . Έχοντας σχεδιάσει ένα σετ μασκών με λόγο n/e = 10/2 > 3και κρατώντας σταθερές τις τιμές n = 10 και e = 2, η βηματική γωνία θ μειώθηκε σταδιακά έως ότου οι διαφορετικές κλίσεις του ευθύγραμμου τμήματος δεν ήταν εύκολα διακριτές. Η τιμή αυτή της βηματικής γωνίας θ είναι 15° και αποτελεί την ελάχιστη τιμή της θ για το συγκεκριμένο ζεύγος (n = 10, e = 2). Όμοια, βρέθηκε ότι η ελάχιστη τιμή της θ για το ζεύγος (n = 8, e = 1) είναι 9° και για το ζεύγος (n = 10, e = 1) είναι 7,5°. Διαπιστώνεται λοιπόν ότι ο υποδιπλασιασμός του εύρους e οδηγεί στον υποδιπλασιασμό της βηματικής γωνίας θ , δείχνοντας ότι τα μεγέθη σχετίζονται γραμμικά (Πίνακας 2.1: σετ 1 και 4). Αντίθετα, μια αύξηση του μεγέθους n οδηγεί σε μείωση της βηματικής γωνίας θ , υποδεικνύοντας ότι τα μεγέθη αυτά είναι αντιστρόφως ανάλογα (Πίνακας 2.1: σετ 4 και 5). Τα συμπεράσματα για τη μικρότερη τιμή της βηματικής γωνίας θ σε σχέση με το λόγο n/e, καθώς και τα πειραματικά δεδομένα του Πίνακα 2.1 περιγράφονται από την εξίσωση (2.5).

$$\theta(e,n) \approx \frac{e}{n} \times 75^{\circ}$$
 (2.5)

Ένα λειτουργικό σετ μασκών θα πρέπει να περιέχει όλες τις θέσεις του ευθύγραμμου τμήματος, καθώς και τη μικρότερη δυνατή βηματική γωνία θ , έτσι ώστε να μπορεί να ανιχνεύσει όλες τις κλίσεις ακμών μέσα σε ένα παράθυρο $n \times n$. Αντικαθιστώντας τη γωνία της εξίσωσης (2.5) στην εξίσωση (2.4) απαλείφουμε τη βηματική γωνία θ .

$$(2.4) \xrightarrow{(2.5)} M(n,e) \approx \begin{cases} 2.4 \cdot \left(\frac{n}{e}\right)^2 & \epsilon \acute{\alpha} \nu \frac{n}{e} \ge 3\\ \chi \alpha \mu \eta \lambda \acute{\eta} \pi 0.\acute{o} \tau \eta \tau \alpha \epsilon \acute{\alpha} \nu \frac{n}{e} < 3 \end{cases}$$
(2.6)

Για την πλήρωση του τέταρτου κανόνα σχεδιασμού, πρέπει να επιλεχθεί εκείνος ο λόγος *n/e* που ελαχιστοποιεί την τιμή της εξίσωσης (2.6). Η Εικόνα 2.9 παρουσιάζει την τρισδιάστατη απεικόνιση της εξίσωσης (2.6).



Εικόνα 2.9

Η τρισδιάστατη απεικόνιση της εξίσωσης (2.6) και το ζεύγος τιμών που επιλέχθηκε για το σχεδιασμό του σετ μασκών. Σύμφωνα με αυτή, το ζεύγος που ελαχιστοποιεί τον απαιτούμενο αριθμό μασκών είναι το n = 3 και e = 1 με $M \approx 21$. Παρόλο που ο λόγος n/e του ζεύγους είναι ίσος με 3, εντούτοις η ποιότητα των μασκών είναι οριακά ανεκτή. Επιπλέον, το μέγεθος των μασκών είναι πολύ μικρό με αποτέλεσμα να μη μπορεί να ανιχνεύσει μεταβολές έντασης σε μέγεθος μεγαλύτερο των 3×3 εικονοστοιχείων, κάτι που όμως είναι συνηθισμένο στις πραγματικές εικόνες. Το ζεύγος που επιλέχθηκε για το σετ μασκών είναι το n = 10 και e=2 με Mpprox 60. Το ζεύγος αυτό έχει αρκετά χαρακτηριστικά που είναι σημαντικά στην πρακτική εφαρμογή του. Ο λόγος *n/e* είναι 5 δίνοντας καλής ποιότητας μάσκες. Το μέγεθος n = 10 είναι συμβατό με τα συνήθη μεγέθη των εικόνων, τα οποία είναι πολλαπλάσια του 10. Έτσι, θα μπορεί να χωρίζεται όλη η εικόνα σε μη επικαλυπτόμενες περιοχές 10×10 εικονοστοιχείων. Επιπλέον, το μέγεθός του είναι αρκετά μεγάλο για να ανιχνεύει ακόμα και σταδιακές μεταβολές έντασης. Ακόμη, η βηματική γωνία του ζεύγους είναι θ = 15° και ταιριάζει απόλυτα με τα νευροφυσιολογικά χαρακτηριστικά των νευρώνων κλίσης του ΑΟΣ (Hubel, et al., 1965), (Hubel, et al., 1968). Τέλος, το εύρος του ευθύγραμμου τμήματος e, είναι τέτοιο που δίνει ανοχή στην ανίχνευση ακμών με κλίση που δεν είναι πολλαπλάσια των 15°. Έτσι, για παράδειγμα, μια ακμή 72° θα ανιχνεύεται από τις μάσκες των 75°, ενώ μια ακμή 16° από εκείνες των 15°. Οι πίνακες 2.2 και 2.3 δείχνουν τις 60 μάσκες του σετ σε δύο διαφορετικές μορφές: για την ανίχνευση γραμμών και για την ανίχνευση μεταβάσεων έντασης.

Μορφή 1		1	2	3	4	5
Ομάδα 1	0°					
Ομάδα 2	15°					
Ομάδα 3	30°	می ا	مر ا			
Ομάδα 4	45°	الم	<i>,</i>			
Ομάδα 5	60°	r.	7	1	1	
Ομάδα 6	75°					
Ομάδα 7	90°					
Ομάδα 8	105°					
Ομάδα 9	120°			٦.	٦.	
Ομάδα 10	135°					
Ομάδα 11	150°					
Ομάδα 12	165°					

Πίνακας 2.2

Η δομή των μασκών για την ανίχνευση γραμμών μεταβολής έντασης.

Μορφή 2		1	2	3	4	5
Ομάδα 1	0°					
Ομάδα 2	15°					
Ομάδα 3	30°					
Ομάδα 4	45°					
Ομάδα 5	60°					
Ομάδα 6	75°					
Ομάδα 7	90°					
Ομάδα 8	105°					
Ομάδα 9	120°					
Ομάδα 10	135°				×.	· · ·
Ομάδα 11	150°					
Ομάδα 12	165°					

Πίνακας 2.3

Η δομή των μασκών για την ανίχνευση μεταβάσεων έντασης.

> Η ονοματολογία των μασκών γίνεται με 3 αριθμούς: (αριθμός ομάδας, αριθμός στήλης, αριθμός μορφής). Για παράδειγμα, η πρώτη μάσκα του πίνακα 2.2 της πρώτης γραμμής είναι η (1,1,1). Οι μάσκες επιλέχθηκε να είναι δυαδικές, για τη μείωση του υπολογιστικού κόστους. Όπως θα αποδειχθεί στη συνέχεια, η επιλογή αυτή δεν έχει καμία επίπτωση τελικό αποτέλεσμα. Τα γκρι εικονοστοιχεία των μασκών είναι διεγερτικές περιοχές (θετική τιμή), ενώ τα μαύρα εικονοστοιχεία είναι ανασταλτικές περιοχές (αρνητική τιμή). Σε κάθε σημείο της εικόνας, συνελίσσονται και οι δύο μορφές της κάθε μάσκας, έτσι ώστε να μπορεί να ανιχνευθεί οποιοδήποτε είδος ακμής. Το χαρακτηριστικό των προτεινόμενων μασκών είναι ότι χρησιμοποιούνται κυρίως σαν πλακίδια, παρά ως κλασικές μάσκες ανίχνευσης ακμών. Αυτό συμβαίνει γιατί και οι 120 μάσκες (60×2=120) συνελίσσονται μόνο σε μη επικαλυπτόμενες περιοχές 10×10 εικονοστοιχείων της εικόνας και σε όλες τις πιθανές θέσεις. Αυτή η προσέγγιση έχει δύο κύρια πλεονεκτήματα. Πρώτον, είναι συμβατή με το πλέγμα των διακριτών θέσεων που απαιτεί η αρχιτεκτονική του οπτικού φλοιού. Παρόλο που οι διακριτές θέσεις εφαρμογής των μασκών είναι μη επικαλυπτόμενες, οποιαδήποτε κλίση ακμής σε οποιαδήποτε θέση κάθε 10×10 περιοχής μπορεί να ανιχνευθεί τουλάχιστον από μια μάσκα. Δεύτερο, μειώνεται σημαντικά το υπολογιστικό κόστος σε σχέση με την κλασική συνέλιξη κατά την οποία η μάσκα μετατοπίζεται σε όλες τις θέσεις της εικόνας. Θα συγκρίνουμε τις πράξεις που απαιτούνται τόσο με την κλασική συνέλιξη, όσο και με την προτεινόμενη μέθοδο για την εύρεση ακμών σε μια εικόνα μεγέθους $n \times m$ εικονοστοιχείων με μάσκες μεγέθους 10×10 εικονοστοιχείων και για 12 διαφορετικές κλίσεις. Η κλασική συνέλιξη απαιτεί τη χρήση 24 μασκών (12 κλίσεις × 2 μάσκες – μια άρτια και μια περιττή – σε κάθε κλίση). Έτσι, για τη συνέλιξη των 24 μασκών με όλα τα εικονοστοιχεία της εικόνας και θεωρώντας πλήρωση με μηδενικά (zero padding) στην περιφέρεια της εικόνας, η κλασική συνέλιξη απαιτεί 24 imes n imes m πράξεις. Αντίθετα, η προτεινόμενη μέθοδος με τη χρήση 120 μασκών σε μη επικαλυπτόμενες περιοχές 10×10 εικονοστοιχείων απαιτεί $120 \times (n/10) \times (m/10) = (24 \times n \times m)/20$ πράξεις, 20 φορές

δηλαδή λιγότερες πράξεις από την κλασική συνέλιξη. Επιπλέον, δεν υπάρχουν εξαρτήσεις μεταξύ των μη επικαλυπτόμενων περιοχών και κάθε μια μπορεί να θεωρηθεί σαν μια διαφορετική εικόνα. Αυτό ευνοεί την παράλληλη επεξεργασία.

Έστω (m, n) το πάνω αριστερό εικονοστοιχείο μιας περιοχής C 10×10 εικονοστοιχείων σε μια μονόχρωμη (grayscale) εικόνα P και (k, l, w) μια από τις 60 μάσκες της μορφής w, όπου $k \in [1,12]$, $l \in [1,5]$ και $w \in \{1,2\}$. Η συνολική τοπική διέγερση $local.ex_{k,l}^C$ της μάσκας (k, l) είναι η μέγιστη τιμή μεταξύ των διεγέρσεων $F_{k,l,w}^C$ των δύο μορφών w της μάσκας.

$$local. ex_{k,l}^{C} = max \left[F_{k,l,1}^{C}, F_{k,l,2}^{C} \right]$$
(2.7)

$$F_{k,l,w}^{C} = \left| \sum_{i=m}^{m+9} \sum_{j=n}^{n+9} \left[P(i,j) \times Q_{k,l,w}(i-m,j-n) \right] \right|$$
(2.8)

με

$$Q_{k,l,w}(y,x) = \begin{cases} \frac{1}{s_{k,l,w}} & \forall \, \delta i \epsilon \gamma \epsilon \rho \tau i \kappa \delta \, \epsilon i \kappa o \nu o \sigma \tau o i \chi \epsilon i \delta \, \tau \eta \varsigma \, \mu \dot{\alpha} \sigma \kappa \alpha \varsigma \\ \frac{-1}{100 - s_{k,l,w}} & \forall \, \alpha \nu \alpha \sigma \tau \alpha \lambda \tau i \kappa \delta \, \epsilon i \kappa o \nu o \sigma \tau o i \chi \epsilon i \delta \, \tau \eta \varsigma \, \mu \dot{\alpha} \sigma \kappa \alpha \varsigma \end{cases}$$

 $Q_{k,l,w}(y,x)$ η συνάρτηση της μάσκας (k,l,w) $s_{k,l,w}$ ο αριθμός των γκρι εικονοστοιχείων της μάσκας (k,l,w), (μαύρα= $100 - s_{k,l,w}$), $m \le i \le m + 9$ η κάθετη συντεταγμένη της περιοχής C, $n \le j \le n + 9$ η οριζόντια συντεταγμένη της περιοχής C, $0 \le y \le 9$ η κάθετη συντεταγμένη της μάσκας (k,l,w), $0 \le x \le 9$ η οριζόντια συντεταγμένη της μάσκας (k,l,w), $0 \le P(i,j) \le 255$ η τιμή έντασης της εικόνας P στη θέση (i,j).

Η εξίσωση (2.8) δείχνει ότι η διέγερση της μάσκας (k, l, w) είναι η διαφορά της μέσης έντασης των εικονοστοιχείων της εικόνας P που βρίσκονται στο διεγερτικό τμήμα της μάσκας (γκρι) από τη μέση ένταση των εικονοστοιχείων στο ανασταλτικό τμήμα της μάσκας (μαύρο). Η τελική κλίση της ακμής στην περιοχή C της εικόνας P είναι η κλίση της μάσκας (a, b) με τη μέγιστη τοπική διέγερση, όπως φαίνεται στην ακόλουθη εξίσωση.

$$local. ex_{a,b}^{C} = max \left[local. ex_{k,l}^{C} \right]_{k=1,l=1}^{k=12,l=5}$$
(2.9)

Σε περίπτωση που η εικόνα P είναι έγχρωμη, δηλαδή P_u , όπου $u = \{R, G, B\}$ είναι οι τρεις χρωματικές συνιστώσες της εικόνας, τότε η διέγερση των μασκών γίνεται ξεχωριστά για κάθε συνιστώσα και επιλέγεται η μέγιστη διέγερση μεταξύ των τριών.

$$F_{k,l,w}^{u,c} = \left| \sum_{i=m}^{m+9} \sum_{j=n}^{n+9} \left[P_u(i,j) \times Q_{k,l,w}(i-m,j-n) \right] \right|$$
(2.10)

$$local. ex_{k,l}^{u,C} = max[F_{k,l,1}^{u,C}, F_{k,l,2}^{u,C}]$$
(2.11)

$$local. ex_{k,l}^{C} = max \left[local. ex_{k,l}^{R,C}, local. ex_{k,l}^{G,C}, local. ex_{k,l}^{B,C} \right]$$
(2.12)

Για την τελική απεικόνιση της εικόνα των ακμών, εφόσον αυτό απαιτείται, για κάθε περιοχή *C* της εικόνας τοποθετείται η μάσκα (*a*, *b*) με τη μεγαλύτερη διέγερση. Πρέπει να τονιστεί επίσης ότι τα αποτελέσματα αυτής της απεικόνισης δεν είναι τα βέλτιστα, αφού οι ακμές σχηματίζονται από μεμονωμένα ευθύγραμμα τμήματα. Εντούτοις, η σωστή απεικόνισή τους δεν είναι στόχος της παρούσας προσέγγισης. Κύριος στόχος είναι η γρήγορη και σωστή ανίχνευση της κλίσης των ακμών σε ένα πλέγμα διακριτών θέσεων της εικόνας. Η προσέγγιση αυτή είναι συμβατή με τη δομή του πρωτοταγούς οπτικού φλοιού, στον οποίο οι νευρώνες κλίσης καταλαμβάνουν διακριτές θέσεις του οπτικού πεδίου. Η Εικόνα 2.10 δείχνει τα αποτελέσματα της προτεινόμενης μεθόδου σε σύγκριση με τη συνέλιξη των φίλτρων Gabor.



Εικόνα 2.10

Τα αποτελέσματα της προτεινόμενης μεθόδου για την ανίχνευση της κλίσης ακμών σε πραγματικές εικόνες. Παρατίθενται επίσης τα αποτελέσματα της συνέλιξης με φίλτρα Gabor 12 κλίσεων ως μέτρο σύγκρισης της ορθότητας των αποτελεσμάτων της προτεινόμενης μεθόδου.

2.5 Περιγραφή και αναγνώριση απλών σχημάτων με μάσκες κλίσης

Στην παράγραφο αυτή παρουσιάζεται μια εφαρμογή των προτεινόμενων μασκών κλίσης, για την περιγραφή και την αναγνώριση απλών περιγραμμάτων, που βρίσκονται σε ομοιόμορφο φόντο.

2.5.1 Περιγραφή της μεθόδου

Η Εικόνα 2.11 παρουσιάζει το μπλοκ διάγραμμα της προτεινόμενης μεθόδου.



Εικόνα 2.11

Το μπλοκ διάγραμμα της προτεινόμενης μεθόδου.

Αρχικά τα δύο σετ μασκών που περιγράφηκαν στην παράγραφο 2.4 εφαρμόζονται στην εικόνα που περιέχει το περίγραμμα. Η εκτέλεση των εξισώσεων (2.7)-(2.12) εξασφαλίζει Διέγερση ότι για κάθε περιοχή 10×10 εικονοστοιχείων της εικόνας θα προκύψει μια νικήτρια μάσκα, μασκών η οποία θα περιγράφει την κλίση των ακμών στο συγκεκριμένο σημείο. Γνωρίζοντας τη νικήτρια μάσκα κάθε περιοχής, είναι ταυτόχρονα γνωστή και η κλίση των ακμών του συγκεκριμένου σημείου. Στα σημεία στα οποία δεν υπάρχουν ακμές, η διέγερση όλων των μασκών είναι μηδενική.

Στη συνέχεια ακολουθεί ο περιγραφέας σχήματος, ο οποίος ουσιαστικά είναι η αλυσίδα των γωνιών που σχηματίζει η κλίση κάθε νικήτριας μάσκας με την επόμενη, καθώς κινούμαστε με ωρολογιακή φορά γύρω από το περίγραμμα. Όταν η γωνία που σχηματίζεται μεταξύ της μάσκας που ελέγχεται και της επόμενης είναι στα αριστερά της κατεύθυνσης κίνησης, τότε η γωνία αυτή θεωρείται αρνητική. Στην αντίθετη περίπτωση, όταν η σχηματιζόμενη γωνία είναι δεξιά σε σχέση με την κατεύθυνση κίνησης, τότε θεωρείται θετική. Οι γωνίες υπολογίζονται με βάση τον αριθμό των βηματικών γωνιών κατά τον οποίο διαφέρουν οι κλίσεις τους. Υπενθυμίζουμε ότι τα δύο σετ των χρησιμοποιούμενων μασκών έχουν σχεδιαστεί ώστε να έχουν βηματική γωνία 15°. Κατά συνέπεια, μια γωνία με τιμή '-2' ουσιαστικά σημαίνει «30° αριστερά προς την κατεύθυνση κίνησης». Η Εικόνα 2.12 δείχνει ένα απλό σχήμα, καθώς και την περιγραφή του βάσει των σχηματιζόμενων γωνιών μεταξύ των μασκών κλίσης.

Περιγραφέας σχήματος



Εικόνα 2.12

Ο υπολογισμός της περιγραφής ενός σχήματος από τις γωνίες των μασκών κλίσης.



Για να μπορέσει η περιγραφή του σχήματος να είναι ανεξάρτητη της μάσκας εκκίνησης, χρησιμοποιείται η ακόλουθη κανονικοποίηση. Έστω ότι η περιγραφή ενός σχήματος Α είναι $A: S_1, S_2, \dots, S_M$, όπου $S_M \in \{-11, -10, -9, -8, -7, -6, -5, -4, -3, -2, -1, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11\}$ και $m = 1 \dots M$. Η αλυσίδα Α ελέγχεται κατά ζεύγη γειτονικών γωνιών, έτσι ώστε να ανιχνευθεί η μεγαλύτερη μετάβαση σύμφωνα με την έκφραση max $[|S_n - S_{n+1}|]_{n=1}^{M-1}$. Το σημείο της μέγιστης μετάβασης αποτελεί και το κανονικοποιημένο σημείο έναρξης της αλυσίδας. Για να γίνει αυτό, η αλυσίδα ολισθαίνει κυκλικά έως ότου το σημείο της μέγιστης μετάβασης βρεθεί πρώτο. Σε περίπτωση που υπάρχουν περισσότερα του ενός υποψήφια σημεία έναρξης, ελέγχονται οι μεταβάσεις που τα ακολουθούν και επιλέγεται αυτό που ακολουθείται από τη μεγαλύτερη μετάβαση. Η διαδικασία αυτή μπορεί να ακολουθηθεί έως ότου προκύψει ένα σημείο εκκίνησης. Πρέπει να σημειωθεί ότι η περιγραφή του σχήματος που προκύπτει είναι ανεξάρτητη της περιστροφής (rotation invariant), αφού δεν καταγράφονται οι απόλυτες τιμές των κλίσεων των μασκών, αλλά οι σχετικές γωνίες μεταξύ τους. Επίσης, στην ακολουθία δεν υπάρχουν μηδενικές γωνίες. Αυτό σημαίνει ότι δεν καταγράφονται οι μεταβάσεις μεταξύ της ίδιας κλίσης μασκών. Κατά συνέπεια, καταγράφονται μόνο σχετικές γωνίες και όχι αναλογίες τμημάτων. Το γεγονός αυτό καθιστά την περιγραφή ανεξάρτητη της κλίμακας (scale invariant).

Η κανονικοποιημένη περιγραφή του σχήματος βρίσκεται σε μια μορφή κατάλληλη για *Ταξινομητής* ταξινόμηση. Έτσι, ένα συνηθισμένο νευρωνικό δίκτυο του τύπου Multi-Layer Perceptron (MLP), εκπαιδευμένο με τη μέθοδο Back Propagation, μπορεί να ταξινομήσει με επιτυχία τις περιγραφές των σχημάτων. Η μορφή του νευρωνικού δικτύου επιλέχθηκε μετά από πειραματισμό με τις βασικές παραμέτρους του. Έτσι, επιλέχθηκε ένα κρυφό επίπεδο νευρώνων και σιγμοειδής λογαριθμική συνάρτηση για την ενεργοποίησή τους. Ο αριθμός των νευρώνων του κρυφού επιπέδου εξαρτάται από το σύνολο των κλάσεων στις οποίες επιθυμούμε να εκπαιδεύσουμε το νευρωνικό δίκτυο. Κάθε τμήμα της αλυσίδας εισάγεται σε έναν ξεχωριστό νευρώνα. Για το λόγο αυτό, ο αριθμός των νευρώνων στο επίπεδο εισόδου πρέπει να είναι ίσος με τη μεγαλύτερη αλυσίδα που μπορεί να συναντηθεί στο δοθέν σύνολο σχημάτων. Το επίπεδο των νευρώνων εξόδου πρέπει να έχει τόσους νευρώνες όσες και οι κλάσεις των σχημάτων. Αυτοί οι νευρώνες έχουν εύρος εξόδου στο διάστημα [0,1], το οποίο χαρακτηρίζει το ποσοστό συμμετοχής του εκάστοτε σχήματος στην κάθε κλάση.

2.5.2 Παράδειγμα ταξινόμησης

Στο παράδειγμα που ακολουθεί χρησιμοποιήθηκε ένα νευρωνικό δίκτυο με 20 νευρώνες στο κρυφό του επίπεδο, το οποίο εκπαιδεύτηκε ώστε να ταξινομεί τέσσερεις κλάσεις σχημάτων. Το σετ εκπαίδευσης που χρησιμοποιήθηκε παρουσιάζεται στην Εικόνα 2.13. Μετά την εκπαίδευση του νευρωνικού δικτύου, διάφορα σχήματα τα οποία αποτελούσαν παραλλαγές σε σχέση με εκείνα του σετ εκπαίδευσης, τροφοδοτήθηκαν στο νευρωνικό δίκτυο. Τα αποτελέσματα αυτής της ταξινόμησης παρουσιάζονται στον Πίνακα 2.4.

Όπως φαίνεται και από τα αποτελέσματα, επιτυγχάνεται ανοχή στην περιστροφή και στην κλίμακα, αφού όλα τα σχήματα του Πίνακα 2.4 διαφέρουν στη θέση, το μέγεθος και την περιστροφή σε σχέση με αυτά του σετ εκπαίδευσης. Σημαντικό ενδιαφέρον παρουσιάζουν τα σχήματα a, b και c, τα οποία αποτελούν διαφορετικές εκδοχές εκείνου για το οποίο εκπαιδεύτηκε το δίκτυο, αφού έχουν διαφορετικές αναλογίες. Επιπλέον, το σχήμα c αποτελεί σχεδίαση με το χέρι, και κατά συνέπεια έχει αρκετές διακυμάνσεις στο περίγραμμά του. Εντούτοις, ταξινομούνται στη σωστή κατηγορία. Τα σχήματα d και e έχουν προέλθει από συμπίεση στον άξονα x και στον άξονα y του αντίστοιχου αρχικού σχήματος της κλάσης. Το νευρωνικό δίκτυο με τη χρήση της προτεινόμενης περιγραφής εξακολουθεί να τα ταξινομεί στη σωστή κλάση.

Κλάση 1	Κλάση 2	Κλάση 3	Κλάση 4
		\bigcirc	
	\bigcirc	\bigcirc	
		\mathcal{O}	\swarrow
	\bigcirc	$\left(\right)$	

Εικόνα 2.13

Το σετ εκπαίδευσης του νευρωνικού δικτύου.
	Σχήματα εισόδου		Κλάσεις			
			Κλάση 1	Κλάση 2	Κλάση 3	Κλάση 4
	а	$\sum_{i=1}^{n}$	0,954	0	0,014	0,102
Πίνακας 2.4 Αποτελέσματα του παραδείγματος ταξινόμησης σχημάτων.	b		0,954	0	0,014	0,101
	с	\sum	0,963	0	0,015	0,104
	d		0,957	0	0,015	0,102
	е		0,954	0	0,014	0,101
	f	\bigcirc	0	0,98	0,019	0,001
	g	\bigcap	0,003	0,053	0,889	0
	h		0,02	0,001	0,091	0,939
	i	M	0,037	0,001	0,067	0,924

Σωστή είναι επίσης και η ταξινόμηση των σχημάτων h και i, τα οποία έχουν υποστεί αρκετές ανομοιόμορφες μεταβολές σε σχέση με το αρχικό σχήμα της εκπαίδευσης. Επίσης, το σχήμα g παριστάνει μια έλλειψη με μεγαλύτερη εκκεντρότητα σε σχέση με αυτή των σχημάτων εκπαίδευσης. Εντούτοις, η ταξινόμησή του είναι σωστή. Γενικότερα, η προτεινόμενη προσέγγιση φαίνεται να έχει αρκετή ανοχή στη μεταβολή των γωνιών των σχημάτων, καθώς και ορισμένα χρήσιμα χαρακτηριστικά γενίκευσης.

2.6 Μέθοδος εξαγωγής εξεχόντων περιγραμμάτων

Στην παράγραφο αυτή παρουσιάζεται η προτεινόμενη μέθοδος για την εξαγωγή εξεχόντων περιγραμμάτων από εικόνες που περιέχουν αντικείμενα σε μη ομοιόμορφο φόντο. Η μέθοδος αυτή βασίζεται στα δύο σετ μασκών που παρουσιάστηκαν στην ενότητα 2.4, καθώς και στα χαρακτηριστικά αλληλεπίδρασης των νευρώνων κλίσης του πρωτοταγούς οπτικού φλοιού.

2.6.1 Δομή του δικτύου

Στην προτεινόμενη προσέγγιση, οι 60 μάσκες των δύο σετ είναι ομαδοποιημένες σε υπερστήλες. Η εικόνα διαιρείται σε μη επικαλυπτόμενες περιοχές, μεγέθους 10×10 εικονοστοιχείων και κάθε περιοχή έχει τη δική της υπερστήλη. Κάθε υπερστήλη είναι μια ανεξάρτητη υπολογιστική μονάδα της κλίσης των ακμών που βρίσκονται στην περιοχή της. Κάθε μάσκα της υπερστήλης είναι επίσης μια ανεξάρτητη υπολογιστική μονάδα μιας συγκεκριμένης μόνο κλίσης ακμών για τη συγκεκριμένη περιοχή. Η έξοδος κάθε υπερστήλης εξαρτάται άμεσα από τη μάσκα με τη μεγαλύτερη διέγερση.

Σύμφωνα με τις νευροφυσιολογικές μελέτες του ΑΟΣ, η εξαγωγή των εξεχόντων περιγραμμάτων βασίζεται στην αλληλεπίδραση των νευρώνων κλίσης από διαφορετικές υπερστήλες. Στην προτεινόμενη προσέγγιση, αυτές οι αλληλεπιδράσεις εκτείνονται σε μια γειτονιά 5×5 υπερστηλών. Αυτό σημαίνει ότι κάθε μάσκα της υπερστήλης (i, j), επηρεάζεται από όλες τις μάσκες, όλων των υπερστηλών που βρίσκονται στην τετράγωνη περιοχή που ορίζουν οι υπερστήλες (i - 2, j - 2) και (i + 2, j + 2). Η Εικόνα 2.14 δείχνει αυτή τη γειτονιά των αλληλεπιδράσεων.



Εικόνα 2.14

Η 5×5 γειτονιά υπερστηλών, μέσα στην οποία όλες οι μάσκες της υπερστήλης (i, j) επηρεάζονται από όλες τις μάσκες, όλων των άλλων υπερστηλών. Έτσι, κάθε μάσκα της υπερστήλης (*i, j*) συνδέεται με 1440 άλλες μάσκες (24 υπερστήλες × 60 μάσκες) στη γειτονιά της. Δε χρησιμοποιείται πλήρωση με μηδενικά (zero padding) στα όρια της εικόνας. Η γειτονιά 5×5 επιλέχθηκε γιατί οι νευροφυσιολογικές μελέτες υποστηρίζουν ότι οι πλευρικές συνδέσεις των νευρώνων εκτείνονται σε αποστάσεις 2-4 φορές το μέγεθος του κλασικού υποδεκτικού τους πεδίου (Sincich, et al., 2001). Η γειτονιά 5×5 επιτρέπει στην κεντρική μάσκα να αλληλεπιδρά με άλλες μάσκες σε αποστάσεις διπλάσιες του μεγέθους της. Η διασύνδεση κάθε μάσκας πραγματοποιείται από 1440 σταθερά βάρη, τα οποία είναι ίδια για όλες τις μάσκες k κάθε υπερστήλης. Σε αντίθεση με προηγούμενες μεθόδους οι οποίες χρησιμοποιούσαν ανασταλτικούς ενδιάμεσους νευρώνες για την αναστολή των μη συγγραμμικών κλίσεων (Li, 1998), (Grossberg, et al., 1985), (Grossberg, 2004), τα βάρη στην προτεινόμενη μέθοδο είναι θετικά για τις διεγερτικές συνδέσεις και αρνητικά για τις ανασταλτικές. Χρησιμοποιούνται με το ίδιο τρόπο όπως και στα κλασικά νευρωνικά δίκτυα με τη διαφορά ότι δε μεταβάλλονται και παραμένουν σταθερά για όλες τις εικόνες. Από τη στιγμή που τα βάρη θα υπολογιστούν μια φορά, παραμένουν για πάντα σταθερά, χωρίς να απαιτείται κάποια επιπλέον τροποποίηση στις τιμές του δικτύου.

Η προτεινόμενη μέθοδος εισάγει ένα νέο τρόπο διασύνδεσης των νευρώνων κλίσης, ο οποίος παρουσίασε καλύτερα αποτελέσματα για το πλαίσιο της προτεινόμενης μεθόδου. Οι προηγούμενες βιολογικά εμπνευσμένες μέθοδοι (Yen, et al., 1998), (Li, 1998), (Grossberg, et al., 1985) υιοθετούσαν διάφορες παραλλαγές της κλασικής συγκυκλικής (cocircular) προσέγγισης (Parent, et al., 1989), (Herault, et al., 1993), σύμφωνα με την οποία δύο στοιχεία κλίσης ανήκουν στην ίδια καμπύλη αν είναι εφαπτόμενα στον ίδιο κύκλο. Η προτεινόμενη προσέγγιση βασίζεται στη συνεκθετικότητα (co-exponentiality), εννοώντας ότι δύο στοιχεία κλίσης ανήκουν στην ίδια καμπύλη αν είναι εφαπτόμενα στην ίδια εκθετική καμπύλη. Η Εικόνα 2.15 παρουσιάζει το προτεινόμενο μοτίβο διασύνδεσης σε σχέση με το κυκλικό.

Εικόνα 2.15

i.Το προτεινόμενο εκθετικό μοτίβο διασύνδεσης.

ii. Το κυκλικό μοτίβο διασύνδεσης.

Η σχεδίαση αναφέρεται για κλίση 0°. Για λόγους απεικόνισης, δεν είναι σχεδιασμένες όλες οι πιθανές καμπύλες.



Στην Εικόνα 2.15i κάθε καμπύλη είναι η γραφική παράσταση της εξίσωσης $y = e^{ax}$, με $0 \le \alpha \le 0.14$, ενώ στην Εικόνα 2.15 κάθε καμπύλη είναι τμήμα της περιφέρειας κύκλου, με ακτίνα $25 \le r \le 10000$ εικονοστοιχεία. Τα όρια $a_{max} = 0.14$ και $r_{min} = 25$ επιλέχθηκαν γιατί δίνουν τέτοια καμπυλότητα που μπορεί να προσεγγιστεί ικανοποιητικά από τα ευθύγραμμα τμήματα των μασκών στο 5×5 πλέγμα της γειτονιάς των υπερστηλών. Σε περίπτωση που τα όρια ήταν διαφορετικά (μεγαλύτερο για τις εκθετικές καμπύλες και μικρότερο για τους κύκλους), η καμπυλότητα θα ήταν μεγάλη και η προσέγγισή της από μάσκες κλίσης μεγέθους 10×10 εικονοστοιχείων δε θα ήταν ικανοποιητική. Τα όρια $a_{min} = 0$ και $r_{max} = 10000$ εικονοστοιχεία επιλέχθηκαν, γιατί στη γειτονιά των 5×5 υπερστηλών προσεγγίζουν την ευθεία γραμμή (μηδενική καμπυλότητα).

Τα βάρη αλληλεπίδρασης των 60 μασκών θα πρέπει να συμβάλουν στο σχηματισμό του εκθετικού μοτίβου σύνδεσης στην 5×5 γειτονιά κάθε μάσκας. Για τον υπολογισμό των βαρών μιας μάσκας (k, l), η μάσκα τοποθετείται στο κέντρο του 5×5 πλέγματος. Το εκθετικό μοτίβο περιστρέφεται και μετατοπίζεται ώστε να έχει την ίδια κλίση με τη μάσκα (k, l) και το κέντρο του να συμπίπτει με το κέντρο της. Η Εικόνα 2.16i προυσιάζει το μοτίβο διασύνδεσης της μάσκας (8,4).



ii. Η προσέγγιση του δεξιού τμήματος της εκθετικής καμπύλης α = 0.13 για τη

iii, iv. Οι διεγερτικές συνδέσεις της μάσκας (8,4) με τη μάσκα καμπύλη α = 0.13. Οι γκριμάσκες δείχνουν τις θέσεις στις οποίες υπάρχει αναστολή (αρνητικό βάρος σύνδεσης) επειδή δεν ταιριάζει με

Αφού η μάσκα τοποθετηθεί στο κέντρο της 5×5 γειτονιάς, όλες οι εκθετικές καμπύλες $(\alpha \in [0.14, 0])$ προσεγγίζονται μεμονωμένα από τα ευθύγραμμα τμήματα των μασκών. Η Εικόνα 2.16
ίι παρουσιάζει την προσέγγιση του δεξιού τμήματος της καμπύλη
ς $\alpha = 0.13$ για τη μάσκα (8,4). Κάθε μάσκα (c,d), που χρησιμοποιείται σε αυτήν την προσέγγιση, καθορίζει μια διεγερτική σύνδεση (θετικό βάρος σύνδεσης) της υπό εξέταση κεντρικής μάσκας (k, l), με τη μάσκα (c, d) στο συγκεκριμένο σημείο της 5×5 γειτονιάς. Οι Εικόνες 2.16iii και 2.16iv δείχνουν τις διεγερτικές συνδέσεις της μάσκας (8,4) με τη μάσκα (5,5). Όπως φαίνεται στις εικόνες, η μάσκα (8,4) συνδέεται με θετικό βάρος με τη μάσκα (5,5) μόνο στις θέσεις (i + 2, j) και (i - 2, j) της 5×5 γειτονιάς. Σε όλες τις άλλες θέσεις η σύνδεση είναι ανασταλτική (αρνητικό βάρος). Αυτό συνεπάγεται ότι η μάσκα (8,4) οποιασδήποτε υπερστήλης (i, j) σε οποιοδήποτε σημείο της εικόνας διεγείρει τις μάσκες (5,5) των γειτονικών της υπερστηλών (i + 2, j) και (i - 2, j), ενώ τις αναστέλλει σε όλες τις άλλες υπερστήλες της 5×5 γειτονιάς. Αυτές είναι 24 από τις 1440 συνολικά συνδέσεις της μάσκας (8,4). Η ίδια διαδικασία ακολουθείται για όλα τα βάρη όλων των υπολοίπων μασκών. Η Εικόνα 2.17 απεικονίζει ένα εποπτικό παράδειγμα λειτουργίας της προτεινόμενης δομής.



Οι τιμές των βαρών είναι αριθμοί στο διάστημα [-1,1]. Τα αρνητικά βάρη, τα οποία αντιστοιχούν σε αναστολή, έχουν σταθερές αρνητικές τιμές. Οι περισσότερες βιολογικές

μέθοδοι (Grossberg, et al., 1985), (Li, 1998), (Yen, et al., 1998), (Mundhenk, et al., 2005), καθορίζουν την αναστολή μεταξύ δύο στοιχείων κλίσης, βάσει της γωνίας που σχηματίζουν, με τη μεγαλύτερη αναστολή να υφίσταται για κάθετες γωνίες. Στην προτεινόμενη μέθοδο χρησιμοποιείται μια σταθερή αρνητική τιμή για όλες τις γωνίες. Αυτή η τιμή ,η οποία εξαρτάται από πολλούς παράγοντες, όπως ο αριθμός των μασκών σε κάθε υπερστήλη, η συνάρτηση συσχέτισης του δικτύου, η αρχιτεκτονική του δικτύου και οι τιμές των θετικών βαρών, επηρεάζει σημαντικά τη συνολική απόδοση του δικτύου. Μια μεγάλη αρνητική τιμή θα είχε σαν αποτέλεσμα την ισχυρή αναστολή μεταξύ των μασκών, με αποτέλεσμα να μην είναι δυνατό να ξεχωρίσουν οι μάσκες των εξεχόντων περιγραμμάτων. Αντίθετα, μια μικρή αρνητική τιμή θα είχε σαν αποτέλεσμα τη μη αναστολή των μασκών που δεν αντιστοιχούν σε εξέχοντα περιγράμματα, με συνέπεια λανθασμένο τελικό αποτέλεσμα. Μεταξύ αυτών των δύο ακραίων καταστάσεων, υπάρχει μια αρνητική τιμή για την οποία το δίκτυο ισορροπεί μεταξύ του συνόλου των αναστολών και του συνόλου των διεγέρσεων. Η τιμή αυτή υπολογίστηκε μετά από εκτεταμένες δοκιμές. Για την παρούσα δομή του δικτύου, η τιμή αυτή βρέθηκε να είναι -0.06. Αυτό είναι λογικό, εάν λάβουμε υπόψη ότι οι ανασταλτικές συνδέσεις είναι κατά πολύ περισσότερες από τις διεγερτικές.

Σε αντίθεση με τα αρνητικά βάρη, τα θετικά έχουν μεταβλητές τιμές, οι οποίες εξαρτώνται από την ανάδραση που δέχεται το σύστημα από υψηλότερες οπτικές περιοχές, με σκοπό να ευνοηθεί η εξαγωγή συγκεκριμένων βαθμών καμπυλότητας. Αυτό σημαίνει ότι τα θετικά βάρη κυμαίνονται στο διάστημα [0,1] και καθορίζονται ανάλογα με τη γωνία θ που σχηματίζεται από τα ευθύγραμμα τμήματα των μασκών που συνδέονται με διεγερτική σύνδεση (Εικόνα 2.16iv). Εάν επιθυμείται η εξαγωγή εξεχόντων περιγραμμάτων χαμηλής καμπυλότητας, τα βάρη των διεγερτικών συνδέσεων με γωνίες θ κοντά στις 180° θα πρέπει να έχουν τιμές 1. Την ίδια στιγμή, οι συνδέσεις με γωνίες κοντά στις 105° θα πρέπει να έχουν τιμές γύρω στο 0 (το εκθετικό μοτίβο σύνδεσης περιορίζει τις θετικές συνδέσεις στο διάστημα $105^{\circ} \le \theta \le 180^{\circ}$). Αντίθετα, αν επιθυμείται η εξαγωγή περιγραμμάτων μεγάλης καμπυλότητας, τα βάρη των διεγερτικών συνδέσεων με γωνίες θ κοντά στις 180° θα πρέπει να έχουν τιμές 0, ενώ οι συνδέσεις με γωνίες κοντά στις 105° θα πρέπει να έχουν τιμές 1. Τα μεταβλητά θετικά βάρη μπορούν να ρυθμίζουν την απόδοση του δικτύου, έτσι ώστε να τονίζονται εξέχοντα περιγράμματα συγκεκριμένης καμπυλότητας, εάν υπάρχει προηγούμενη γνώση των χαρακτηριστικών της καμπυλότητας που αναζητείται. Η Εικόνα 2.18 παρουσιάζει τα βάρη που χρησιμοποιούνται για τον τονισμό τεσσάρων διαφορετικών βαθμών καμπυλότητας. Στην περίπτωση που δεν υπάρχει προηγούμενη γνώση του βαθμού καμπυλότητας, όλα τα θετικά βάρη έχουν τιμή 1.

Εικόνα 2.18

Οι κατανομές των βαρών για την ανίχνευση συγκεκριμένων καμπυλοτήτων στα εξέχοντα περιγράμματα.



2.6.2 Κανόνες διέγερσης

Ο κύριος στόχος του προτεινόμενου δικτύου είναι να αλλάξει επαναληπτικά την αρχική τοπική διέγερση των μασκών, έτσι ώστε να διευρύνει τις διαφορές τιμών μεταξύ αυτών που ανήκουν σε εξέχοντα περιγράμματα και αυτών που ανήκουν στις υπόλοιπες ακμές. Η έξοδος κάθε μάσκας είναι συνδυασμός δύο παραγόντων: της τοπικής διέγερσης λόγω αντίθεσης έντασης στο συγκεκριμένο σημείο της εικόνας – εξίσωση (2.7) – και της επιρροής των 1440 μασκών με τα οποία είναι συνδεδεμένη στη 5×5 γειτονιά της. Η επιρροή αυτή μπορεί να είναι είτε διεγερτική είτε ανασταλτική.

2.6.2.1 Αναστολή

Η συνολική αναστολή που δέχεται η μάσκα (k, l) της υπερστήλης (i, j) από τις μάσκες στη γειτονιά της είναι:

$$inh_{(k,l)(i,j)}(t) = \begin{cases} \sum_{m=1}^{12} \sum_{n=1}^{5} \sum_{y=i-2}^{i+2} \sum_{x=j-2}^{j+2} out_{(m,n)(y,x)}(t-1) \times weight_{(m,n)(y,x)\to(k,l)(i,j)} \forall \mathcal{W} \\ 0 \forall \mathcal{W}' \\ \mathcal{W}: out_{(k,l)(i,j)}(t-1) > 0, weight_{(m,n)(y,x)\to(k,l)(i,j)} < 0, y \neq i, x \neq j \end{cases}$$

$$(2.13)$$

όπου t είναι ο αριθμός της επανάληψης, (m, n) είναι η μάσκα που αναστέλλει την (k, l), (y, x) είναι οι συντεταγμένες της μάσκας (m, n) στην 5×5 γειτονιά και $weight_{(m,n)(y,x)\to(k,l)(i,j)}$ είναι το βάρος της σύνδεσης από «τη μάσκα (m, n) της θέσης (y, x) στη μάσκα (k, l) της θέσης (i, j)».

2.6.2.2 Διέγερση των δύο λοβών της μάσκας

Η συνολική θετική διέγερση, που δέχεται μια μάσκα από τις υπόλοιπες 1440 της γειτονιάς της, συγκεντρώνεται σε δύο διαφορετικά σύνολα: στο λοβό 1, που αντιστοιχεί στο ένα άκρο του ευθύγραμμου τμήματος της μάσκας και στο λοβό 2, που αντιστοιχεί στο άλλο άκρο. Η προσέγγιση των λοβών παρουσιάστηκε αρχικά από τον Grossberg (Grossberg, et al., 1985), (Grossberg, et al., 1995), (Grossberg, 1994), (Grossberg, 2004), με τη μορφή των διπολικών κυττάρων (bipole cells) και έχει επίσης υιοθετηθεί από στην εργασία (Williams, et al., 2000). Η Εικόνα 2.15i δείχνει τους δύο λοβούς του εκθετικού μοτίβου διασύνδεσης. Η Εικόνα 2.16i δείχνει τους δύο λοβούς της μάσκας (8,4). Η Εικόνα 2.16ii δείχνει τους δύο λοβούς της μάσκας (5,5) στον πρώτο λοβό της μάσκας (8,4), ενώ η Εικόνα 2.16iν τη διέγερση που προκαλεί ο πρώτος λοβός της μάσκας (5,5) στο δεύτερο λοβό της μάσκας (8,4). Η συνολική διέγερση που δέχεται ο λοβός N, όπου $N \in \{1,2\}$, της μάσκας (k,l) στη θέση (i,j) της 5×5 γειτονιάς δίνεται από την ακόλουθη εξίσωση:

$$Lobe \ N_{(k,l)(i,j)}(t) = \begin{cases} \sum_{m=1}^{12} \sum_{n=1}^{5} \sum_{y=i-2}^{i+2} \sum_{x=j-2}^{j+2} out_{(m,n)(y,x)}(t-1) \times weight_{(m,n)(y,x) \to (k,l)(i,j),N} \forall \ Z \\ 0 \ \forall \ Z' \\ Z: out_{(k,l)(i,j)}(t-1) > 0, weight_{(m,n)(y,x) \to (k,l)(i,j),N} > 0, y \neq i, x \neq j \end{cases}$$
(2.14)

όπου t είναι ο αριθμός της επανάληψης, (m, n) είναι η μάσκα που διεγείρει την (k, l), (y, x) είναι οι συντεταγμένες της μάσκας (m, n) στην 5×5 γειτονιά και $weight_{(m,n)(y,x)\to(k,l)(i,j),N}$ είναι το βάρος της σύνδεσης από «τη μάσκα (m, n) της θέσης (y, x) στο N λοβό της μάσκα (k, l) στης θέσης (i, j)». Οι εξισώσεις (2.13) και (2.14) διασφαλίζουν ότι οι μάσκες με μεγάλες τιμές έχουν μεγαλύτερη επιρροή από εκείνες με χαμηλές τιμές. Η ισχύς αυτής της επιρροής καθορίζεται από τα βάρη της σύνδεσης των δύο μασκών.

2.6.2.3 Συνολική διέγερση μάσκας

Σύμφωνα με προηγούμενες μεθόδους, (Grossberg, et al., 1985), (Grossberg, et al., 1995), (Grossberg, 1994), (Grossberg, 2004), κάθε στοιχείο κλίσης ενεργοποιείται μόνο εάν δέχεται διέγερση και στους δύο λοβούς του. Αυτό διασφαλίζει ότι μόνο τα κύτταρα που είναι μέρος μιας αλυσίδας ενεργοποιούνται. Επιπλέον αποτρέπει την ανάπτυξη περιγραμμάτων σε λανθασμένες κατευθύνσεις. Στην προτεινόμενη μέθοδο η προσέγγιση αυτή επεκτείνεται. Πιο συγκεκριμένα, εισάγεται μια νέα συνάρτηση συσχέτισης (affinity function) η οποία ευνοεί περισσότερο την ομαλή συνέχεια των περιγραμμάτων.

$$ex_{(k,l)(i,j)}(t) = \begin{cases} \frac{lobe1_{(k,l)(i,j)}(t) + lobe2_{(k,l)(i,j)}(t)}{2(1 + |lobe1_{(k,l)(i,j)}(t) - lobe2_{(k,l)(i,j)}(t)|)} \forall \mathcal{U} \\ 0 \forall \mathcal{U}' \\ \mathcal{U}: lobe1_{(k,l)(i,j)}(t) > 0, \quad lobe2_{(k,l)(i,j)}(t) > 0 \end{cases}$$
(2.15)

όπου $ex_{(k,l)(i,j)}(t)$ είναι η συνολική διέγερση της μάσκας (k,l) στη θέση (i,j) κατά την επανάληψη t. Η καινοτομία της εξίσωσης (2.15) είναι ο όρος ομοιότητας μεταξύ των δύο λοβών. Η συνολική διέγερση, που μια μάσκα δέχεται από τις γειτονικές της, είναι συνάρτηση της ομοιότητας των διεγέρσεων που δέχονται οι λοβοί της. Μια μάσκα που συμμετέχει σε ένα ομαλό περίγραμμα, θα δεχθεί περίπου ίδια διέγερση και στους δυο λοβούς της από τις υπόλοιπες μάσκες της αλυσίδας. Κατά συνέπεια, ο όρος $|lobe1_{(k,l)(i,j)}(t) - lobe2_{(k,l)(i,j)}(t)|$ του παρονομαστή θα ελαχιστοποιείται για μάσκες που ανήκουν σε μια ομαλή αλυσίδα. Αυτό θα μεγιστοποιήσει τη διέγερση της μάσκας, δίνοντάς της το μέσο όρο των διεγέρσεων των δύο λοβών. Αντίθετα, μάσκες που δεν είναι μέρος μιας ομαλής αλυσίδας, θα δεχτούν ανόμοια διέγερση στους δύο λοβούς. Έτσι, ο όρος του παρονομαστή δε θα ελαχιστοποιηθεί και η μάσκα θα δεχθεί μόνο ένα κλάσμα του μέσου όρου της διέγερσης των δύο λοβών. Γενικότερα, όσο μεγαλύτερη είναι η ομοιότητα της διέγερσης των δύο λοβών τόσο μεγαλύτερη είναι και η συνολική διέγερση της μάσκας. Η Εικόνα 2.19 παρουσιάζει την τρισδιάστατη γραφική παράσταση της εξίσωσης (2.15). Είναι φανερό ότι η μεγαλύτερη συνολική διέγερση συμβαίνει όταν οι δύο λοβοί έχουν παρόμοια διέγερση.



2.6.2.4 Έξοδος μάσκας

Η συνολική έξοδος μιας μάσκας (k, l) που βρίσκεται στην υπερστήλη (i, j) τη χρονική στιγμή t δίνεται από την ακόλουθη εξίσωση:

$$out_{(k,l)(i,j)}(t) = \begin{cases} log(q) & \epsilon \acute{\alpha} v \ q > 0\\ 0 & \epsilon \acute{\alpha} v \ q \le 0 \end{cases}$$

$$q = out_{(k,l)(i,j)}(t-1) + ex_{(k,l)(i,j)}(t) + inh_{(k,l)(i,j)}(t) + hcol. out_{(i,j)}(t-1)$$

$$(2.16)$$

όπου hcol.out_{(k,l)(i,j)}(t) είναι η έξοδος της υπερστήλης (i, j) η οποία αναλύεται στη συνέχεια. Το άθροισμα q αυξάνεται εκθετικά σε κάθε επανάληψη και φτάνει σε πολύ μεγάλες τιμές ακόμη και για μικρό αριθμό επαναλήψεων. Για το λόγο αυτό εισάγεται ο λογάριθμος στην εξίσωση (2.16), ο οποίος γραμμικοποιεί την αύξηση του αθροίσματος q

και ταυτόχρονα διατηρεί τις τιμές του σε χαμηλά επίπεδα. Η αρχική τιμή της μάσκας (k, l) στη θέση (i, j) για την πρώτη επανάληψη (t = 0) της εξίσωσης (2.16) είναι η τοπική της διέγερση:

$$out_{(k,l)(i,j)}(0) = local. ex_{(k,l)}^{(i,j)}$$
(2.17)

2.6.2.5 Εσωτερικός ανταγωνισμός υπερστήλης

Τα βιολογικά μοντέλα (Grossberg, et al., 1985), (Li, 1998), (Ross, et al., 1998), έχουν χρησιμοποιήσει διάφορες μορφές εσωτερικού ανταγωνισμού σε μια υπερστήλη (intracolumnar competition) μεταξύ των κυττάρων κλίσης. Σύμφωνα με αυτό τον ανταγωνισμό, κάθε κύτταρο κλίσης αναστέλλει όλα τα άλλα κατά ένα παράγοντα ανάλογο τις διαφοράς της γωνίας που σχηματίζουν. Η μέγιστη αναστολή υφίσταται για κύτταρα με κάθετες κλίσεις. Η προτεινόμενη μέθοδος υιοθετεί την προσέγγιση «ο νικητής τα παίρνει όλα» (winner takes all). Αυτό σημαίνει ότι η μάσκα με τη μεγαλύτερη έξοδο σε κάθε υπερστήλη (*i*, *j*), η οποία ονομάζεται νικήτρια μάσκα *m*. *out*_(*i*,*j*)(*t*), μηδενίζει τις τιμές όλων των άλλων μασκών της υπερστήλης. Η προσέγγιση αυτή εξασφαλίζει ότι ο συνολικός αριθμός των ενεργών μασκών παραμένει ο ελάχιστος (1 το πολύ για κάθε υπερστήλη). Ως αποτέλεσμα, βελτιώνεται κατά πολύ ο χρόνος εκτέλεσης της μεθόδου. Η εξίσωση (2.18) εκφράζει τον παραπάνω κανόνα.

$$out_{(k,l)(i,j)}(t) = 0 \quad \epsilon \dot{\alpha} \vee \quad out_{(k,l)(i,j)}(t) < m. \quad out_{(i,j)}(t)$$

$$m. \quad out_{(i,j)}(t) = max \left[out_{(m,n)(i,j)}(t) \right]_{m=1,n=1}^{m=12,n=5}$$
(2.18)

2.6.2.6 Προσαρμοσμένη διαρροή δυναμικού

Σε προηγούμενες μεθόδους έχουν χρησιμοποιηθεί τεχνικές διαρροών (leaking) για να προσομοιώσουν ορισμένα χαρακτηριστικά των βιολογικών νευρώνων (Mundhenk, et al., 2003), (Mundhenk, et al., 2005). Αυτό σημαίνει ότι οι νευρώνες χάνουν κάποιο σταθερό δυναμικό σε κάθε επανάληψη. Η προτεινόμενη μέθοδος επεκτείνει αυτήν την προσέγγιση προσαρμόζοντας το δυναμικό διαρροής ανάλογα με τη συνολική διέγερση της νικήτριας μάσκας *m. out*_(*i*,*j*)(*t*). Οι νικήτριες μάσκες, που είναι μέρος μιας αλυσίδας και κατά συνέπεια έχουν μεγάλη τιμή συνολικής διέγερσης *m. ex*_{(*k.l*)(*i*,*j*)}(*t*), χάνουν ελάχιστα από την τιμή τους, ενώ αυτές που δεν είναι μέρος κάποιας αλυσίδας και έχουν μικρή συνολική διέγερση της προσαρμοσμένη διαρροή είναι η ακόλουθη:

$$ad. leak_{(i,j)}(t) = \frac{1}{1 + m. ex_{(i,j)}(t)}$$
(2.19)





Η Εικόνα 2.20 δείχνει την τρισδιάστατη απεικόνιση της εξίσωσης (2.19). Είναι φανερό ότι, όταν η διέγερση των δύο λοβών είναι άνιση, τότε υπάρχει αύξηση της προσαρμοσμένης διαρροής. Η προσέγγιση που εισάγουν οι εξισώσεις (2.15) και (2.19), διασφαλίζει ότι η τοπική διέγερση των μασκών, που δεν αποτελούν τμήμα εξεχόντων περιγραμμάτων, σταδιακά μειώνεται. Αυτή η μείωση ευρύνει τις διαφορές των τιμών σε σχέση με τις μάσκες των εξεχόντων περιγραμμάτων.

2.6.2.7 Τελική έξοδος του δικτύου

Η τελική έξοδος του δικτύου σε κάθε χρονική στιγμή t είναι η έξοδος της κάθε υπερστήλης. Η έξοδος της υπερστήλης (i, j) τη χρονική στιγμή t εξαρτάται από τα χαρακτηριστικά της νικήτριας μάσκας. Αυτή η σχέση περιγράφεται στην ακόλουθη εξίσωση:

$$hcol. out_{(i,j)}(t) = \begin{bmatrix} hcol. out_{(i,j)}(t-1) + m. out_{(i,j)}(t) \\ +m. inh_{(i,j)}(t) - ad. leak_{(i,j)}(t) \end{bmatrix}^{+}$$
(2.20)
$$[x]^{+} = max[x, 0]$$

όπου $m.inh_{(i,j)}(t)$ η συνολική αναστολή που δέχεται η νικήτρια μάσκα $m.out_{(i,j)}(t)$ και $ad.leak_{(i,i)}(t)$ η προσαρμοσμένη διαρροή δυναμικού της νικήτριας μάσκας. Στην εξίσωση (2.20), η έξοδος της υπερστήλης συσσωρεύει τις εξόδους της νικήτριας στήλης, το σύνολο των αναστολών που δέχεται, καθώς και την προσαρμοσμένη διαρροή. Για τις υπερστήλες των εξεχόντων περιγραμμάτων, η έξοδος της νικήτριας μάσκας είναι μεγαλύτερη από το σύνολο των αναστολών και της προσαρμοσμένης διαρροής. Έτσι, η έξοδος της υπερστήλης είναι θετική και ανάλογη με την έξοδο της νικήτριας μάσκας. Στην περίπτωση υπερστηλών που δεν είναι μέρος εξεχόντων περιγραμμάτων, το σύνολο των αναστολών και η προσαρμοσμένη διαρροή ξεπερνούν την έξοδο της νικήτριας μάσκας και κατά συνέπεια, η έξοδός τους είναι μηδενική. Η αρχική έξοδος μιας υπερστήλης (*i,j*) για την πρώτη επανάληψη (t = 0) είναι μηδέν.

$$hcol.out_{(i,j)}(0) = 0$$
 (2.21)







Εικόνα 2.21

i.Μια πραγματική εικόνα στην οποία φαίνονται οι τοπικές διεγέρσεις των μασκών.
Παρουσιάζεται η γειτονιά δύο διαφορετικών μασκών: της (1,3) που ανήκει σε εξέχων
περίγραμμα, και της (7,4) που δεν ανήκει.
Και οι δύο μάσκες έχουν ίδια σχεδόν αρχική διέγερση.

ii. Οι τιμές των μεγεθών της μάσκας (1,3) για 10 επαναλήψεις .

iii. Οι τιμές των μεγεθών της μάσκας (7,4) για 10 επαναλήψεις.

Σε κάθε επανάληψη, οι υπερστήλες που βρίσκονται πάνω από εξέχοντα περιγράμματα τείνουν να αυξάνουν την έξοδό τους, ενώ οι υπόλοιπες υπερστήλες σταδιακά τη μειώνουν.

Ο τονισμός των εξεχόντων περιγραμμάτων έρχεται σαν απόρροια των όλο και διευρυνόμενων διαφορών στις εξόδους των υπερστηλών δύο κατηγοριών. Αυτό φαίνεται ξεκάθαρα στην Εικόνα 2.21. Η Εικόνα 2.21i απεικονίζει τις τοπικές διεγέρσεις των μασκών για μια πραγματική εικόνα. Σε αυτήν την εικόνα έχουν επιλεχθεί δύο μάσκες, μια που είναι μέρος ενός εξέχοντος περιγράμματος (1,3) και μια που δεν ανήκει σε εξέχον περίγραμμα (7,4). Οι Εικόνες 2.21ii και 2.21iii, δείχνουν τις τιμές των εξισώσεων (2.13), (2.14), (2.15), (2.16), (2.19) και (2.20) για τις δύο μάσκες. Και οι δύο μάσκες έχουν σχεδόν ίδια αρχική διέγερση. Αυτό σημαίνει ότι η τοπική αντίθεση έντασης στο σημείο τους είναι ίδια. Εντούτοις, λόγω των διαφορετικών γειτονικών αλληλεπιδράσεών τους, οι τιμές των μεγεθών τους έχουν τελείως διαφορετική συμπεριφορά, ακόμη και αν ξεκινάνε με την ίδια διέγερση. Πιο συγκεκριμένα, η μάσκα (7,4) παρουσιάζει συνεχώς ελαττούμενη έξοδο της υπερστήλης της, με αποτέλεσμα στην 8^η επανάληψη να μηδενίζεται. Αντίθετα, στο τέλος των 10 επαναλήψεων η έξοδος της υπερστήλης για τη μάσκα (1,3) γίνεται σχεδόν διπλάσια της τοπικής της διέγερσης. Αυτό είναι αποτέλεσμα τριών παραγόντων: της αναστολής από τις γειτονικές μάσκες, της διέγερσης των δύο λοβών και της προσαρμοσμένης διαρροής. Είναι ξεκάθαρο από τα διαγράμματα, ότι η μάσκα (7,4) δέχεται στη 2^η επανάληψη σχεδόν δεκαπλάσια αναστολή από τις γειτονικές της μάσκες σε σχέση με την (1,3). Αυτό συμβαίνει, γιατί στη γειτονιά της μάσκας (7,4) δεν υπάρχουν χωρικές σχέσεις που να αναδεικνύουν κάποιο εξέχον περίγραμμα. Ως αποτέλεσμα, η έξοδος των δύο υπερστηλών ήδη από τη 2^η επανάληψη διαφέρει σημαντικά. Η μάσκα (1,3) λαμβάνει επιπλέον μεγαλύτερη και πιο ομοιόμορφη διέγερση στους δύο λοβούς της, με αποτέλεσμα να παρουσιάζει μεγαλύτερη συνολική διέγερση. Αυτό με τη σειρά του επηρεάζει την προσαρμοσμένη διαρροή, η οποία είναι μικρότερη για τις μάσκες με μεγάλη συνολική διέγερση. Έτσι, η μάσκα (7,4) μειώνει συνεχώς την τιμή εξόδου της υπερστήλης στην οποία ανήκει.

Ένα σημαντικό χαρακτηριστικό των εξισώσεων της προτεινόμενης μεθόδου είναι ότι δεν παρουσιάζουν συσχετισμούς μεταξύ των διαφόρων μασκών. Πιο συγκεκριμένα, σε κάθε επανάληψη, τα χαρακτηριστικά κάθε μάσκας μπορούν να υπολογιστούν ανεξάρτητα από όλες τις άλλες, αφού τόσο οι μάσκες όσο και οι υπερστήλες λειτουργούν σαν ανεξάρτητα στοιχεία. Από τα παραπάνω, είναι φανερό ότι η πολυπλοκότητα της μεθόδου είναι ανάλογη του μεγέθους της εικόνας στην οποία εφαρμόζεται, δηλαδή Ο(*N*).

Η τελική δυαδικοποίηση των αποτελεσμάτων του δικτύου γίνεται με τη μέθοδο (Otsu, 1979), η οποία θα περιγραφεί αναλυτικά στο κεφάλαιο 4. Με τη μέθοδο αυτή καθορίζεται το κατώφλι δυαδικοποίησης από τις τελικές τιμές των μασκών του δικτύου. Το κατώφλι αυτό τμηματοποιεί τις μάσκες σε δύο κλάσεις: σε αυτές που θα αποτελέσουν το φόντο και σε αυτές των οποίων τα ευθύγραμμα τμήματα θα αποτελέσουν τα εξέχοντα περιγράμματα.

2.7 Πειραματικά αποτελέσματα – συγκρίσεις

Το σύστημα, που περιγράφηκε στην προηγούμενη ενότητα, υλοποιήθηκε σε γλώσσα προγραμματισμού C και εκτελέστηκε από έναν επεξεργαστή Intel Pentium στα 3.0 MHz με 512MB μνήμης σε λειτουργικό σύστημα Windows XP. Σαν είσοδοι επιλέχθηκαν συνθετικές

αλλά και αρκετές πραγματικές εικόνες. Από τις πραγματικές εικόνες που χρησιμοποιήθηκαν, ορισμένες προέρχονται από άλλες υπάρχουσες μεθόδους, έτσι ώστε να υπάρξει σύγκριση των αποτελεσμάτων. Πρέπει να τονιστεί επίσης ότι κύριος σκοπός της προτεινόμενης μεθόδου είναι η επιτυχής εξαγωγή των εξεχόντων περιγραμμάτων σε πραγματικές εικόνες και σε χρόνους εκτέλεσης λίγων δευτερολέπτων.

2.7.1 Αξιολόγηση απόδοσης

Για την αξιολόγηση των αποτελεσμάτων της διαδικασίας εξαγωγής εξεχόντων περιγραμμάτων είναι αναγκαίο να οριστεί μια μετρική, η οποία θα ποσοτικοποιεί την απόδοση κάθε μεθόδου. Προηγούμενες έρευνες έχουν χρησιμοποιήσει τέτοιες μετρικές για συνθετικές κυρίως εικόνες. Ο Braun (Braun, 1999), χρησιμοποίησε συνθετικές εικόνες από φίλτρα Gabor παρόμοιες με την Εικόνα 2.2 και τις παρουσίασε σε ανθρώπινους παρατηρητές για να αξιολογήσει το μηχανισμό εξαγωγής εξεχόντων ακμών του ΑΟΣ. Παρόμοια προσέγγιση ακολουθήθηκε επίσης στην εργασία (Williams, et al., 2000), στην οποία χρησιμοποίησαν συνθετικές εικόνες από απλά φυσικά περιγράμματα τοποθετημένα πάνω σε θορυβώδη (μη ομοιόμορφο) φόντο. Στην περίπτωση των συνθετικών εικόνων, το σήμα (εξέχοντα περιγράμματα) και ο θόρυβος (φόντο) είναι εκ των προτέρων γνωστά και η απόδοση της μεθόδου είναι το ποσοστό του σήματος το οποίο περιλαμβάνεται στην έξοδο του συστήματος.

Η αξιολόγηση της εξαγωγής εξεχόντων περιγραμμάτων σε πραγματικές εικόνες αναπόφευκτα απαιτεί την ύπαρξη της ιδανική εικόνας ακμών (ground truth), η οποία είναι η εικόνα των ακμών, έτσι όπως την αντιλαμβάνεται ο τυπικός ανθρώπινος παρατηρητής. Έχοντας την ιδανική εικόνα, ορίζεται ποιο είναι το σήμα και ποιος ο θόρυβος, με αποτέλεσμα να μπορούν να αξιολογηθούν τα αποτελέσματα των αλγορίθμων. Αυτή η προσέγγιση έχει ακολουθηθεί στην εργασία (Grigorescu, et al., 2003) και χρησιμοποιείται για την αξιολόγηση των αποτελεσμάτων της προτεινόμενης μεθόδου. Έστω E_{GT} και B_{GT} το σύνολο των εικονοστοιχείων της ιδανικής εικόνα ακμών, τα οποία αντιστοιχούν στις ακμές και στο φόντο. Παρομοίως, έστω E_D και B_D το σύνολο των εικονοστοιχείων του τελικού αποτελέσματος μιας μεθόδου εξαγωγής περιγραμμάτων, τα οποία αντιστοιχούν στις ακμές και στο φόντο. Τότε, το σύνολο των σωστών εξεχόντων περιγραμμάτων θα δίνεται από τη σχέση $E = E_D \cap E_{GT}$. Ανάλογα, το σύνολο των σωστών περιγραμμάτων, τα οποία δεν εξήχθησαν από την υπό εξέταση μέθοδο, θα δίνεται από τη σχέση $E_{FN} = B_D \cap E_{GT}$. Τέλος, το σύνολο των περιγραμμάτων, τα οποία λανθασμένα εξήχθησαν από την υπό εξέταση μέθοδο, θα δίνεται από τη σχέση $E_{FP} = E_D \cap B_{GT}$. Έχοντας τα τρία αυτά σύνολα, η μετρική, που καθορίζει την ποιότητα του εξαγόμενου αποτελέσματος της μεθόδου, είναι η ακόλουθη:

$$M = \frac{num(E)}{num(E) + num(E_{FP}) + num(E_{FN})}$$
(2.22)

όπου *num*(*X*) είναι ο αριθμός των στοιχείων του συνόλου *X*. Η μετρική που περιγράφεται στην εξίσωση (2.22) λαμβάνει τιμές στο διάστημα [0,1]. Όταν όλα τα εικονοστοιχεία των εξεχόντων περιγραμμάτων εξαχθούν σωστά και δεν υπάρχουν εικονοστοιχεία φόντου, τα

οποία λανθασμένα εξήχθησαν σαν εξέχοντα περιγράμματα, τότε η τιμή της μετρικής γίνεται 1. Σε όλες τις άλλες περιπτώσεις, η τιμή της μετρικής γίνεται μικρότερη της μονάδας.

2.7.2 Αποτελέσματα για πραγματικές εικόνες

Η Εικόνα 2.22 παρουσιάζει μερικά από τα αποτελέσματα της μεθόδου για πραγματικές εικόνες. Η πρώτη στήλη δείχνει την αρχική εικόνα. Η δεύτερη τις ακμές με τη μέθοδο Canny έτσι όπως εξάγονται από το Matlab (R13). Η τρίτη στήλη περιέχει την τοπική διέγερση των υπερστηλών, που είναι η αρχική τους τιμή κατά την πρώτη επανάληψη, ενώ η τελευταία περιλαμβάνει το τελικό αποτέλεσμα του δικτύου.



Εικόνα 2.22

Αποτελέσματα της μεθόδου για πραγματικές εικόνες. Η τοπική διέγερση δείχνει την αρχική κατάσταση των μασκών κατά την έναρξη των επαναλήψεων, ενώ η έξοδος, τα τελικά περιγράμματα. Όπως είναι φανερό από την τελευταία στήλη, το προτεινόμενο δίκτυο εξάγει μερικά από τα πιο εξέχοντα περιγράμματα της εικόνας, ενώ δεν επηρεάζεται από τις ακμές ισχυρής αντίθεσης που μπορεί να υπάρχουν στην αρχή των επαναλήψεων (στήλη 3). Σε όλες τις περιπτώσεις, η προτεινόμενη μέθοδος έχει απλοποιήσει ιδιαίτερα την αρχική εικόνα των ακμών, σχηματίζοντας το βασικό της σκιαγράφημα (primal sketch). Πρέπει να σημειωθεί ότι σε ορισμένες περιπτώσεις, ικανοποιητικά αποτελέσματα μπορεί να εξάγει και ο αλγόριθμος Canny. Για να γίνει όμως αυτό, θα πρέπει να ρυθμιστούν κατάλληλα οι 3 παράμετροί του, κάτι που δεν είναι σταθερό για όλες τις εικόνες και απαιτεί χρονοβόρες δοκιμές. Αντίθετα, τα αποτελέσματα της προτεινόμενης μεθόδου λαμβάνονται χωρίς καμία ρύθμιση παραμέτρων, με τα ίδια αρχικά βάρη συνδέσεων.

2.7.3 Αποτελέσματα για συνθετικές εικόνες

Η Εικόνα 2.23 παρουσιάζει τα αποτελέσματα της προτεινόμενης μεθόδου για συνθετικές εικόνες. Οι εικόνες αυτές αποτελούνται από ευθύγραμμα τμήματα τυχαίου μεγέθους, κλίσης και θέσης, ορισμένα από τα οποία σχηματίζουν εξέχοντα περιγράμματα. Ανάλογες εικόνες έχουν χρησιμοποιηθεί και από τους (Sha'ashua, et al., 1988), (Alter, et al., 1998), (Yen, et al., 1998), (Li, 1998), (Choe, και συν., 2004), (Mundhenk, et al., 2005). Όπως και στις πραγματικές εικόνες, η μέθοδος τονίζει τα εξέχοντα περιγράμματα χωρίς να επηρεάζεται από τις υπόλοιπες ακμές.





Αποτελέσματα της προτεινόμενης μεθόδου για συνθετικές εικόνες.

2.7.4 Τονισμός συγκεκριμένων καμπυλοτήτων

Η προτεινόμενη μέθοδος μπορεί να ρυθμίσει την επιλεκτικότητά της ως προς συγκεκριμένες καμπυλότητες των εξεχόντων περιγραμμάτων. Αυτό μπορεί να επιτευχθεί ρυθμίζοντας τα βάρη των διεγερτικών συνδέσεων των μασκών, ανάλογα με τη γωνία θ που σχηματίζουν οι διευθύνσεις τους. Η Εικόνα 2.24 απεικονίζει τα αποτελέσματα της μεθόδου για μια πραγματική και μια συνθετική εικόνα και για διάφορες τιμές καμπυλότητας. Τα βάρη των συνδέσεων που χρησιμοποιήθηκαν είναι αυτά που παρουσιάζονται στην Εικόνα 2.18. Η συνθετική εικόνα περιέχει κύκλους διαφορετικής ακτίνας. Η μικρές ακτίνες σχηματίζουν μεγάλη καμπυλότητα, ενώ οι μεγάλες ακτίνες μικρή καμπυλότητα. Όταν δεν τονίζεται κάποια συγκεκριμένη καμπυλότητα (2^η στήλη), τότε όλοι οι κύκλοι τονίζονται το ίδιο. Αντίθετα, καθώς τονίζονται όλο και μεγαλύτερες καμπυλότητες πηγαίνοντας προς τα δεξιά, εμφανίζονται περισσότερο οι μικροί κύκλοι, ενώ οι μεγαλύτεροι σβήνουν. Το ίδιο παρατηρείται και για την πραγματική εικόνα. Στις χαμηλές καμπυλότητες τονίζεται η ευθεία γραμμή και όχι οι κύκλοι. Σε υψηλότερες καμπυλότητες τονίζονται σταδιακά κύκλοι όλο και μικρότερου μεγέθους.



2.7.5 Εκθετικό και κυκλικό μοτίβο διασύνδεσης

Για να μπορέσει να συγκριθεί η απόδοση του κυκλικού (Parent, et al., 1989) και του προτεινόμενου εκθετικού μοτίβου διασύνδεσης, υλοποιήθηκε μια δεύτερη μορφή της μεθόδου, η οποία χρησιμοποιεί βάρη του κυκλικού μοτίβου. Τα αποτελέσματα αυτής της σύγκρισης φαίνονται στην Εικόνα 2.25. Όπως φαίνεται και από τις τιμές της μετρικής, το εκθετικό μοτίβο διασύνδεσης δίνει πιο ικανοποιητικά αποτελέσματα. Πιο συγκεκριμένα, τα περιγράμματα του εκθετικού μοτίβου έχουν καλύτερη συνέχεια, ενώ αυτά του κυκλικού μοτίβου εμφανίζουν περισσότερες ασυνέχειες. Αυτό είναι ιδιαίτερα φανερό στις Εικόνες 2.25.1 και 2.25.2. Επιπλέον, το κυκλικό μοτίβο διασύνδεσης τείνει να εξάγει περισσότερα λανθασμένα περιγράμματα σε σχέση με το εκθετικό, το οποίο έχει σημαντική επίπτωση στη μετρική. Αυτό φαίνεται ιδιαίτερα στις Εικόνες 2.25.3 και 2.25.6, όπου ορισμένες ακμές, οι οποίες δεν αποτελούν τμήμα εξεχόντων περιγραμμάτων, εμφανίζονται στο τελικό



Εικόνα 2.25

Σύγκριση των αποτελεσμάτων της προτεινόμενης μεθόδου για το εκθετικό και το κυκλικό μοτίβο διασύνδεσης.

2.7.6 Σύγκριση με υπάρχουσες μεθόδους

Η προτεινόμενη μέθοδος εκτελέστηκε σε εικόνες που χρησιμοποιήθηκαν στην παρουσίαση άλλων μεθόδων για την εξαγωγή των εξεχόντων περιγραμμάτων. Οι αλγόριθμοι αυτοί είναι οι (Wang, et al., 2005), CINNIC (Mundhenk, et al., 2005), (Mundhenk, 2005) και

(Grigorescu, et al., 2003). Οι μέθοδοι αυτοί επιλέχθηκαν, γιατί είναι οι μόνες που παρουσιάζουν εκτεταμένα αποτελέσματα για πραγματικές εικόνες. Ο πρώτος αλγόριθμος είναι βασισμένος στη θεωρία των γράφων, χωρίς κάποια σχέση με το ΑΟΣ. Επίσης, είναι σχεδιασμένος να εξάγει μόνο κλειστά εξέχοντα περιγράμματα. Αυτό σημαίνει ότι σε περίπτωση που κάποιο περίγραμμα δεν είναι ολοκληρωμένο, π.χ. λόγω επικάλυψης κάποιου αντικειμένου, τότε το περίγραμμα δεν περιλαμβάνεται στο αποτέλεσμα. Αντίθετα, οι δύο άλλοι αλγόριθμοι, είναι βασισμένοι σε παρόμοιες αρχές με την προτεινόμενη μέθοδο και κατά συνέπεια έχουν και τις περισσότερες ομοιότητες.



Εικόνα 2.26

Σύγκριση των αποτελεσμάτων της προτεινόμενης μεθόδου και της (Wang, et al., 2005).

Σύγκριση με (Wang, et al., 2005)

Η Εικόνα 2.26 δείχνει τη σύγκριση της προτεινόμενης μεθόδου με την (Wang, et al., 2005). Είναι φανερό ότι τα αποτελέσματα των δύο μεθόδων είναι κατά μεγάλο ποσοστό ίδια. Επιπλέον, λόγω του γεγονότος ότι η μέθοδος αυτή βρίσκει ένα κλειστό περίγραμμα κάθε φορά, σε εικόνες, όπως η Εικόνα 2.26.4 και 2.26.6, βρίσκουν ένα μόνο περίγραμμα. Αντίθετα, η προτεινόμενη μέθοδος εξάγει τα εξέχοντα περιγράμματα ακόμη και αν αυτά δεν είναι κλειστά. Η μεγάλη διαφορά μεταξύ των δύο μεθόδων βρίσκεται στους χρόνους εκτέλεσης. Οι χρόνοι της μεθόδου (Wang, et al., 2005) αναφέρονται στην εκτέλεση από έναν επεξεργαστή Intel Xeon στα 3.06 GHz. Παρόλο που ο συγκεκριμένος επεξεργαστής είναι ισχυρότερος του επεξεργαστή στον οποίο εκτελέστηκε η προτεινόμενη μέθοδος, οι χρόνοι εκτέλεσής της είναι πιο αργοί κατά περίπου μια με δύο τάξεις μεγέθους. Αυτό επίσης δικαιολογείται από την πολυπλοκότητα της μεθόδου (Wang, et al., 2005), η οποία είναι Ο(N^{7/4}).

Η Εικόνα 2.27 παρουσιάζει τη σύγκριση της προτεινόμενης μεθόδου με τη CINNIC (Mundhenk, et al., 2005). Η μέθοδος αυτή έχει πολλά κοινά με την προτεινόμενη μέθοδο, αφού και οι δύο βασίζονται σε ίδιες ιδιότητες του ΑΟΣ. Εντούτοις, δεν εξάγει δυαδικές ακμές, αλλά μια εικόνα σε αποχρώσεις του γκρι, στην οποία οι σκούρες αποχρώσεις δηλώνουν εξέχοντα σημεία ή περιγράμματα. Για να μπορέσει να χρησιμοποιηθεί όμως η μετρική, ώστε να εξαχθούν αριθμητικά αποτελέσματα, απαιτείται η ύπαρξη δυαδικών ακμών. Κάτι τέτοιο θα απαιτούσε την επιπλέον χρήση ενός αλγορίθμου δυαδικοποίησης, ο οποίος όμως θα άλλαζε σημαντικά τα αποτελέσματα της συγκεκριμένης μεθόδου. Για το λόγο αυτό, η σύγκριση με τη CINNIC περιορίζεται σε ποιοτικό επίπεδο. Σύμφωνα με τα αποτελέσματα, η μέθοδος CINNIC φαίνεται να έπεται σημαντικά σε σχέση με την προτεινόμενη μέθοδο, ιδιαίτερα όταν υπάρχει ισχυρά ανομοιόμορφο φόντο. Η προτεινόμενη μέθοδος καταφέρνει να εξάγει τα σωστά περιγράμματα, ενώ η CINNIC όχι. Αυτό είναι ιδιαίτερα εμφανές στις Εικόνες 2.27.3, 2.27.5 και 2.27.6. Επιπλέον, ο χρόνος εκτέλεσής τους διαφέρει κατά δύο τάξεις μεγέθους.

Η Εικόνα 2.28 δείχνει την ποιοτική σύγκριση της προτεινόμενης μεθόδου με τη μέθοδο (Grigorescu, et al., 2003), η οποία είναι εμπνευσμένη από το ΑΟΣ. Όπως είναι φανερό από τα αποτελέσματα, η προτεινόμενη μέθοδος δίνει γενικότερα καλύτερα αποτελέσματα. Η βασικότερη διαφορά μεταξύ των δυο μεθόδων είναι το γεγονός ότι η μέθοδος (Grigorescu, et al., 2003) επηρεάζεται από περιοχές με ισχυρή υφή. Στις περιοχές αυτές έχει την τάση να εξάγει λανθασμένα ακμές οι οποίες ανήκουν στην υφή και οι οποίες κανονικά δεν αποτελούν τμήμα των εξεχόντων περιγραμμάτων. Αυτός είναι και ο κύριος λόγος για τον οποίο υποβαθμίζεται η ποιότητα των αποτελεσμάτων της μεθόδου αυτής. Αντίθετα, ο χωρικός ανταγωνισμός που υλοποιείται στην προτεινόμενη μέθοδο, σε συνδυασμό με το μοτίβο της εκθετικής σύνδεσης, αναστέλλουν τις ακμές οι οποίες δεν παρουσιάζουν χωρικούς σχηματισμούς που να οδηγούν σε ομαλή συνέχεια. Έτσι, εξάγονται πολύ λιγότερα λανθασμένα περιγράμματα στις περιοχές με έντονη υφή.

Σύγκριση με (Mundhenk, et al., 2005)

Σύγκριση με (Grigorescu, et al., 2003)

Αρχική εικόνα	CINNIC	Προτεινόμενη μέθοδος	
	A		
1. 254×247	≈ 2 min	254×247: 0.2 sec	
2.254×245	≈ 2 min	254×245: 0.2 sec	
PRESERVED STATU 1820-1807			
3. 254×246	≈ 2 min	254×246: 0.3 sec	
	adverse a		
4. 255×244	≈ 2 min	255×244: 0.3 sec	
5. 253×245	≈ 2 min	253×245: 0.3 sec	
6. 253×250	≈ 2 min	253×250: 0.2	

Εικόνα 2.27

Σύγκριση των αποτελεσμάτων της προτεινόμενης μεθόδου και της (Mundhenk, et al., 2005).





Σύγκριση των αποτελεσμάτων της προτεινόμενης μεθόδου και της (Grigorescu, et al., 2003).

Η μέθοδος (Grigorescu, et al., 2003) εξάγει δυαδικές ακμές στο τελικό της αποτέλεσμα. Για το λόγο αυτό, μπορεί να εφαρμοστεί η μετρική, ώστε να δοθεί μια αριθμητική περιγραφή της ποιότητας των αποτελεσμάτων. Για την αριθμητική σύγκριση με την προτεινόμενη μέθοδο, χρησιμοποιείται η ίδια βάση εικόνων που χρησιμοποιήθηκε και στην εργασία (Grigorescu, et al., 2003). Η βάση αυτή είναι ιδανική για την περίπτωσή μας, αφού περιέχει εικόνες με ζώα στο φυσικό τους περιβάλλον, το οποίο αναπόφευκτα περιέχει αρκετές ακμές προερχόμενες από υφή. Επιπλέον, για κάθε εικόνα περιλαμβάνεται και η αντίστοιχη

ιδανική εικόνα ακμών, η οποία σχεδιάστηκε από ανθρώπινο παρατηρητή. Η αρχική εικόνα αναλύεται από τους δύο αλγορίθμους και εξάγονται τα εξέχοντα περιγράμματα. Τα αποτελέσματα στη συνέχεια συγκρίνονται με την ιδανική εικόνα ακμών και υπολογίζεται η εξίσωση (2.22) η οποία χαρακτηρίζει αριθμητικά την ποιότητα κάθε αποτελέσματος. Η Εικόνα 2.29 παρουσιάζει μια από τις εικόνες του σετ σύγκρισης, μαζί με τα αποτελέσματα των δύο αλγορίθμων και την ιδανική εικόνα ακμών.



Η Εικόνα 2.30 παρουσιάζει τα αποτελέσματα της μετρικής για το σύνολο των εικόνων της βάσης. Όπως φαίνεται, η προτεινόμενη μέθοδος επιτυγχάνει μεγαλύτερη τιμή στη μετρική για όλες σχεδόν τις εικόνες, κάτι που επιβεβαιώνει τα ποιοτικά αποτελέσματα της Εικόνας 2.28.



Η Εικόνα 2.31 παρουσιάζει το σύνολο των εικονοστοιχείων $num(E_{FP})$ που ανήκουν σε λανθασμένα περιγράμματα και τα οποία εξήχθησαν από τις δύο μεθόδους. Στις περισσότερες εικόνες η προτεινόμενη μέθοδος εξήγαγε μικρότερο αριθμό λανθασμένων περιγραμμάτων από την μέθοδο (Grigorescu, et al., 2003). Αυτό επιβεβαιώνει τη

διαπίστωση που έγινε για τα αποτελέσματα της Εικόνας 2.28, στην οποία φαίνεται ξεκάθαρα ότι η προτεινόμενη μέθοδος είναι καταλληλότερη για εικόνες που περιέχουν περιγράμματα σε φόντο με έντονη υφή.





Το σύνολο των λανθασμένων εικονοστοιχείων που εξήχθησαν από τις δύο μεθόδους.

2.8 Συμπεράσματα

Στο κεφάλαιο αυτό προτάθηκε μια νέα μέθοδος εξαγωγής εξεχόντων περιγραμμάτων, βασισμένη στα χαρακτηριστικά του ΑΟΣ. Η μέθοδος αυτή εισάγει μια σειρά από νέα χαρακτηριστικά:

- Ένα σύνολο δυαδικών μασκών και μια νέα προσέγγιση για την εξαγωγή της κλίσης των ακμών μιας εικόνας. Η νέα αυτή προσέγγιση απαιτεί 20 φορές λιγότερους υπολογισμούς από την κλασική συνέλιξη με φίλτρα Gabor, ενώ έχει πολύ καλά αποτελέσματα. Επιπλέον, δίνει τη δυνατότητα για πλήρη παραλληλία στην επεξεργασία τμημάτων της εικόνας και είναι συμβατό με τη δομή του πρωτοταγούς οπτικού φλοιού.
- 2. Έναν καινούριο περιγραφέα σχήματος, ο οποίος βασίζεται στις σχετικές γωνίες που σχηματίζουν οι μάσκες κλίσης ενός σχήματος. Όταν ο συγκεκριμένος περιγραφέας συνδυαστεί με έναν ταξινομητή, έχει την ικανότητα να ταξινομεί τις περιγραφές απλών σχημάτων, παρουσιάζοντας ανοχή στην περιστροφή, στην κλιμάκωση και στις στρεβλώσεις. Επίσης, το γεγονός ότι δεν ενεργεί στο επίπεδο των εικονοστοιχείων της εικόνας, αλλά στο ανώτερο επίπεδο επεξεργασίας των μασκών κλίσης, τον καθιστά ανεκτικό σε διακυμάνσεις που προέρχονται από θόρυβο.

- 3. Μια νέα αρχιτεκτονική δικτύου, η οποία εκμεταλλεύεται τα χαρακτηριστικά του προτεινόμενου σετ μασκών. Το δίκτυο είναι πλήρως παράλληλο και μπορεί να επεξεργάζεται ταυτόχρονα διαφορετικά σημεία της εικόνας.
- 4. Ένα νέο μοτίβο διασύνδεσης των κυττάρων κλίσης (μάσκες), το οποίο βασίζεται στις εκθετικές συναρτήσεις και δίνει καλύτερα αποτελέσματα από το μοτίβο που βασίζεται στην κυκλική διασύνδεση.
- 5. Μια νέα συνάρτηση διέγερσης των κυττάρων κλίσης, η οποία βασίζεται στην ομοιότητα της διέγερσης των δύο λοβών τους. Η συνάρτηση αυτή ευνοεί τη σωστή συνέχεια των εξεχόντων περιγραμμάτων και μειώνει δραστικά τη διέγερση των μασκών που δεν ανήκουν σε εξέχοντα περιγράμματα.
- 6. Μια νέα προσαρμοσμένη μορφή διαρροής του δυναμικού των νευρώνων κλίσης, βάσει της οποίας, οι νευρώνες με μικρή διέγερση χάνουν το μεγαλύτερο ποσοστό δυναμικού. Αυτό ευνοεί τη γρήγορη μείωση της τιμής των μασκών που δεν ανήκουν σε εξέχοντα περιγράμματα.
- 7. Μια νέα μορφή ανταγωνισμού στις υπερστήλες, σύμφωνα με την οποία επικρατεί μόνο η μάσκα με τη μεγαλύτερη τιμή. Η προσέγγιση αυτή διασφαλίζει τον ελάχιστο αριθμό ενεργών μασκών, βελτιώνοντας το χρόνο εκτέλεσης της μεθόδου.
- 8. Το δίκτυο της προτεινόμενης μεθόδου έχει τη δυνατότητα να τονίζει συγκεκριμένους βαθμούς καμπυλότητας και κατά συνέπεια, μπορεί να χρησιμοποιηθεί για την ανίχνευση συγκεκριμένων χαρακτηριστικών σχημάτων, όπως ευθείες ή κύκλους.
- 9. Η εξαγωγή των εξεχόντων περιγραμμάτων έρχεται σαν επακόλουθο της εφαρμογής των κανόνων Gestalt, σε συνδυασμό με τα χαρακτηριστικά των νευρικών συνδέσεων και, έτσι, δεν υπάρχουν περιορισμοί στον αριθμό των περιγραμμάτων που μπορούν να εξαχθούν ή η απαίτηση να είναι κλειστά.
- 10. Τέλος, η μέθοδος παρουσιάζει ικανοποιητικά αποτελέσματα για μια ευρεία γκάμα πραγματικών και συνθετικών εικόνων, σε συνδυασμό πάντα με τους καλύτερους χρόνους εκτέλεσης που έχουν αναφερθεί στην υπάρχουσα βιβλιογραφία.

Τα παραπάνω στοιχεία αναδεικνύουν ότι η προτεινόμενη μέθοδος μπορεί να χρησιμοποιηθεί σε εφαρμογές, που απαιτούν την εξαγωγή των εξεχόντων περιγραμμάτων και γενικότερα του αδρού σκιαγραφήματος της εικόνας, σε μικρούς χρόνος εκτέλεσης. Τέτοιες εφαρμογές μπορεί να είναι η ρομποτική όραση και η πλοήγηση αυτόνομων οχημάτων.



- 128 Συμπεράσματα
- 114 Συγκρίσεις με υπάρχουσες μεθόδους
- 100 Προτεινόμενη μέθοδος
- 95 Νευροφυσιολογικά στοιχεία του ΑΟΣ
- 83 Ανασκόπηση υπάρχουσας βιβλιογραφίας
- 80 Διατύπωση του προβλήματος

3

Βελτίωση αποχρώσεων σε σκηνές μεγάλου δυναμικού εύρους

Το κεφάλαιο αυτό αναφέρεται στο πρόβλημα της υποβάθμισης της ποιότητας των εικόνων, ως αποτέλεσμα της υποέκθεσης ή της υπερέκθεσης του ειδώλου, σε περιπτώσεις όπου το δυναμικό εύρος της σκηνής ξεπερνάει το δυναμικό εύρος της κάμερας. Δίνεται ο αναλυτικός ορισμός του προβλήματος και παρουσιάζεται ο σημαντικός του ρόλος στα προβλήματα όρασης μηχανής. Γίνεται μια σύντομη περιγραφή των μεθόδων που έχουν αναπτυχθεί μέχρι τώρα για την επίλυση αυτού του προβλήματος και παρουσιάζονται οι λόγοι για τους οποίους οι λύσεις αυτές δεν είναι πάντα επαρκείς. Στη συνέχεια, παρατίθενται νευροφυσιολογικά στοιχεία κυττάρων του ΑΟΣ, τα οποία σκιαγραφούν τον τρόπο με τον οποίο η εξέλιξη έχει αντιμετωπίσει το πρόβλημα αυτό. Ακολουθεί η περιγραφή της προτεινόμενης μεθόδου, η οποία υιοθετεί τα παραπάνω στοιχεία, καθώς και σειρά συγκριτικών πειραμάτων. Τέλος, παρουσιάζονται συμπεράσματα που αναλύουν τη συνεισφορά της προτεινόμενης μεθόδου.

3.1 Διατύπωση του προβλήματος

Τόσο στην ψηφιακή φωτογραφία όσο και στην αναλογική, η ποιότητα της τελικής εικόνας εξαρτάται κατά πολύ από το ποσοστό του προσπίπτοντος φωτός στη φωτοευαίσθητη επιφάνεια ή στο φιλμ. Πολύ μικρή ή πολύ μεγάλη ποσότητα φωτός οδηγεί σε μη ικανοποιητική αναπαραγωγή του ειδώλου. Αυτό μπορεί να αποτελέσει μεγάλο πρόβλημα σε περιπτώσεις μη ελεγχόμενων συνθηκών φωτισμού (π.χ. εξωτερικοί χώροι). Για το λόγο αυτό, έχουν αναπτυχθεί αλγόριθμοι Αυτόματης Έκθεσης (Automatic Exposure - AE), οι οποίοι ελέγχουν το χρόνο έκθεσης του ειδώλου στον αισθητήρα. Έτσι, βελτιστοποιείται η ποσότητα του προσπίπτοντος φωτός ανάλογα με τις εξωτερικές συνθήκες φωτισμού και κατ' επέκταση η ποιότητα της τελικής εικόνας. Τέτοιοι αλγόριθμοι έχουν ευρεία εφαρμογή σε ψηφιακές κάμερες και φωτογραφικές μηχανές, στις οποίες δίνουν δυνατότητα ικανοποιητικής λήψης σε μια μεγάλη γκάμα διαφορετικών συνθηκών φωτισμού. Εντούτοις, υπάρχουν περιπτώσεις στις οποίες οι αλγόριθμοι ΑΕ αδυνατούν να δώσουν ικανοποιητικό αποτέλεσμα. Τέτοιες είναι οι περιπτώσεις στις οποίες το δυναμικό εύρος της σκηνής ξεπερνάει το δυναμικό εύρος του αισθητήρα. Δυναμικό Εύρος (Dynamic Range) μιας σκηνής είναι ο λόγος του σημείου με τη μεγαλύτερη φωτεινότητα προς το σημείο με τη μικρότερη φωτεινότητα και δείχνει το εύρος των διαβαθμίσεων φωτεινότητας που υπάρχουν στη σκηνή.

Δυναμικό Εύρος σκηνής

3.1.1 Χαρακτηριστικά CCD αισθητήρων

Η Εικόνα 3.1 παρουσιάζει το χαρακτηριστικό διάγραμμα ενός τυπικού αισθητήρα CCD (Charge-Coupled Device), στο οποίο απεικονίζεται το δυναμικό του εύρος.



Εικόνα 3.1

Διάγραμμα έκθεσης-εξόδου για έναν τυπικό αισθητήρα CCD .

cd : candelas : κηρία (μονάδες μέτρησης φωτεινότητας).

Εικόνα από (Yamada, 2006)

Στο διάγραμμα της Εικόνας 3.1 φαίνεται ότι υπάρχει μια γραμμική περιοχή στην έξοδο του αισθητήρα, κατά την οποία η τάση εξόδου αυξάνεται γραμμικά σε σχέση με την έκθεση του ειδώλου. Η γραμμική αυτή περιοχή βρίσκεται ανάμεσα σε δύο όρια, τα οποία εξαρτώνται από τον ίδιο τον αισθητήρα. Το πρώτο όριο είναι ο Θόρυβος Σκότους (Dark Noise), ο οποίος είναι αποτέλεσμα του Ρεύματος Σκότους (Dark Current). Το Ρεύμα Σκότους είναι ένα ανεπιθύμητο φαινόμενο που παρατηρείται όταν ο CCD αισθητήρας δεν εκτίθεται σε ποσότητα φωτός. Οφείλεται στη συγκέντρωση φορτίου στο κύκλωμα του κάθε εικονοστοιχείου και εξαρτάται από τα χαρακτηριστικά του αισθητήρα, το χρόνο συγκέντρωσης του φορτίου και τη θερμοκρασία (Yamada, 2006). Το δεύτερο όριο είναι η Τάση Κορεσμού (Saturation Voltage), η οποία είναι η μέγιστη τάση που μπορεί να δώσει ο αισθητήρας στην έξοδό του. Πέρα από το σημείο αυτό, η έξοδος παραμένει σταθερή ακόμη και αν η έκθεση αυξάνεται. Τα δύο αυτά σημεία καθορίζουν το τμήμα Α'Β' που είναι το Δυναμικό Εύρος του αισθητήρα και δείχνει το εύρος των τιμών που μπορεί να έχει η τάση εξόδου του αισθητήρα. Το τμήμα ΑΒ είναι το Οπτικό Δυναμικό Εύρος (Optical

Dynamic Range) του αισθητήρα και ουσιαστικά καθορίζει το εύρος των τιμών έκθεσης που αντιστοιχούν στο δυναμικό του εύρους. Στόχος των αλγορίθμων ΑΕ είναι να υπολογίσουν το βέλτιστο χρόνο έκθεσης, έτσι ώστε οι τιμές που θα προκύψουν για τα διάφορα τμήματα της σκηνής να βρίσκονται εντός του Οπτικού Δυναμικού Εύρους.

3.1.2 Υποέκθεση & Υπερέκθεση

Η Εικόνα 3.2 δείχνει την επίδραση διαφορετικών χρόνων έκθεσης στην καταγραφή μιας σκηνής μεγάλου δυναμικού εύρους από έναν αισθητήρα CCD. Για πολύ μικρούς χρόνους έκθεσης (t1), το εύρος της σκηνής «συμπιέζεται», με αποτέλεσμα να μειώνεται σημαντικά η αντίθεση της εικόνας. Οι περιοχές υψηλής φωτεινότητας (Δ) βρίσκονται εντός του Οπτικού Δυναμικού Εύρους του αισθητήρα (αν και δε γίνεται εκμετάλλευση όλου του εύρους), ενώ οι περιοχές χαμηλής φωτεινότητας (Γ) βρίσκονται εκτός της περιοχής ΑΒ. Στην περίπτωση αυτή υπάρχει, υποέκθεση των περιοχών που βρίσκονται στο τμήμα ΓΑ και κάθε πληροφορία που βρίσκεται εκεί χάνεται μέσα στο Σκοτεινό Θόρυβο (Εικόνα 3.1). Για μεγαλύτερους χρόνους έκθεσης (t_2), υπάρχει μικρότερη συμπίεση στο εύρος της σκηνής, με αποτέλεσμα να υπάρχει εντονότερη αντίθεση σε σχέση με το χρόνο t1. Οι περιοχές υψηλής φωτεινότητας (Δ) καταλαμβάνουν όλο το Οπτικό Δυναμικό Εύρος του αισθητήρα, ενώ οι περιοχές χαμηλής φωτεινότητας (Γ) εξακολουθούν να βρίσκονται εκτός της περιοχής ΑΒ. Ως αποτέλεσμα, συνεχίζει να υπάρχει υποέκθεση των περιοχών που βρίσκονται στο τμήμα ΓΑ, μόνο που τώρα οι περιοχές αυτές είναι λιγότερες, λόγω της μικρότερης συμπίεσης του εύρους της σκηνής. Για ακόμη μεγαλύτερους χρόνους έκθεσης (t₃), δεν υπάρχει συμπίεση στο εύρος της σκηνής. Οι περιοχές υψηλής φωτεινότητας (Δ) βρίσκονται εκτός του Οπτικού Δυναμικού Εύρους του αισθητήρα, με αποτέλεσμα να υπάρχει υπερέκθεση των περιοχών που βρίσκονται στο τμήμα ΒΔ. Οποιαδήποτε πληροφορία υπάρχει στην περιοχή αυτή έχει χαθεί οριστικά λόγω της Τάσης Κορεσμού του αισθητήρα (Εικόνα 3.1). Οι περιοχές χαμηλής φωτεινότητας (Γ) εξακολουθούν να βρίσκονται εκτός της περιοχής ΑΒ, με αποτέλεσμα να υπάρχει υποέκθεση των περιοχών ΓΑ, μόνο που τώρα οι περιοχές αυτές είναι ακόμη λιγότερες σε σχέση με το χρόνο t₂, λόγω της μη συμπίεσης του εύρους της σκηνής. Τέλος, για πολύ μεγάλους χρόνους έκθεσης (t₄), δεν υπάρχει υποέκθεση, όμως αυξάνονται οι υπερεκτεθιμένες περιοχές (ΒΔ), με

αποτέλεσμα να υπάρχει μεγάλη απώλεια πληροφορίας. Σε τέτοιες περιπτώσεις σκηνών με μεγάλο δυναμικό εύρος, είναι προτιμότερο να αποφεύγεται η υπερέκθεση, αφού έτσι χάνονται τελείως οι οπτικές πληροφορίες της σκηνής. Αντίθετα, κατά την υποέκθεση, όπως θα δούμε αργότερα, υπάρχει δυνατότητα ανάκτησης μέρους της αρχικής πληροφορίας. Για το λόγο αυτό, οι αλγόριθμοι ΑΕ επιλέγουν την υποέκθεση από την υπερέκθεση σε περιπτώσεις σκηνών με μεγάλο δυναμικό εύρος.



Εικόνα 3.2

Αναπαράσταση του εύρους φωτεινοτήτων μιας σκηνής μεγάλου δυναμικού εύρους και τα αποτελέσματα στην ποιότητα της τελικής εικόνας για 4 διαφορετικούς χρόνους έκθεσης:

$$(t_1 < t_2 < t_3 < t_4)$$

Δεν υπάρχει χρόνος έκθεσης που να μπορεί να περιορίσει το δυναμικό εύρος της σκηνής εντός του οπτικού δυναμικού εύρους του αισθητήρα.

3.1.3 Σκηνές Μεγάλου Δυναμικού Εύρους

Στην πράξη, σκηνές μεγάλου δυναμικού εύρους εμφανίζονται πολύ συχνά στο εξωτερικό περιβάλλον. Αυτό συμβαίνει κυρίως, όταν στην ίδια σκηνή συνυπάρχει μια φωτεινή πηγή μαζί με μια σκιά ή όταν ένα αντικείμενο βρίσκεται μπροστά από πολύ φωτεινό φόντο. Μια τέτοια σκηνή παρουσιάζεται στην Εικόνα 3.3, έτσι όπως έχει καταγραφεί από μια κάμερα σε διαφορετικούς χρόνους έκθεσης. Σε τέτοιες περιπτώσεις το δυναμικό εύρος φωτεινοτήτων της σκηνής μπορεί να φτάσει έως και 100.000.000:1 (Devlin, 2002) τη στιγμή που το ΑΟΣ μπορεί να επεξεργαστεί διαβαθμίσεις φωτεινότητας από 1000:1 (Jobson, et al., 1997) έως 10.000:1 (Devlin, 2002), χωρίς να λαμβάνεται υπόψη η δυνατότητα χρονικής προσαρμογής, ενώ οι κάμερες Συνηθισμένου Δυναμικού Εύρους (ΣΔΕ) (Standard Dynamic Range) καταγράφουν αντιθέσεις από 256:1 έως 1024:1. Αυτή η μεγάλη διαφορά στη δυνατότητα σωστής απεικόνισης των σκηνών μπορεί να αποτελέσει σημαντικό πρόβλημα σε συστήματα όρασης μηχανών. Πιο συγκεκριμένα, συστήματα που θα πρέπει να λειτουργούν σε μη ελεγχόμενες συνθήκες φωτισμού, π.χ. αυτόματα συστήματα πλοήγησης, θα αντιμετωπίζουν πρόβλημα από την έλλειψη οπτικής πληροφορίας σε πιθανές υπερεκτεθειμένες ή υποεκτεθειμένες περιοχές της εικόνας. Σημαντικά επηρεάζονται επίσης και οι ψηφιακές φωτογραφικές μηχανές και οι κάμερες του εμπορίου, αφού το τελικό αισθητικό αποτέλεσμα θα απέχει σημαντικά από την πραγματική σκηνή. Είναι φανερό, λοιπόν, ότι απαιτούνται λύσεις στο πρόβλημα της σωστής απεικόνισης σκηνών μεγάλου δυναμικού εύρους, οι οποίες θα εξαλείφουν ή θα περιορίζουν τα φαινόμενα της υπερέκθεσης και της υποέκθεσης και θα συμβάλουν σε μια πιο ρεαλιστική απεικόνιση της αρχικής σκηνής.



Εικόνα 3.3

Σκηνή μεγάλου δυναμικού εύρους (αντικείμενο μπροστά από φωτεινό φόντο) σε διαφορετικούς χρόνους έκθεσης. Δεν υπάρχει χρόνος έκθεσης που να καταγράφει ταυτόχρονα και τις δύο περιοχές.

3.2 Ανασκόπηση υπάρχουσας βιβλιογραφίας

Στα πλαίσια του προβλήματος της ικανοποιητικής απεικόνισης σκηνών, που αναφέρθηκε στην προηγούμενη παράγραφο, έχει παρουσιαστεί μια μεγάλη γκάμα λύσεων.

3.2.1 Εξοπλισμός Μεγάλου Δυναμικού Εύρους

Η άμεση λύση στο πρόβλημα είναι η χρήση καμερών με Μεγάλο Δυναμικό Εύρος (ΜΔΕ) (High Dynamic Range – HDR). Στην περίπτωση αυτή, ο αισθητήρας της κάμερας διαθέτει αρκετό δυναμικό εύρος ώστε να καταγράφει περισσότερες διαβαθμίσεις έντασης από μια κάμερα ΣΔΕ, με αποτέλεσμα να εξαλείφονται τα φαινόμενα υποέκθεσης και υπερέκθεσης στις περισσότερες σκηνές. Εντούτοις, η λύση αυτή έχει μέχρι στιγμής πολύ περιορισμένη απήχηση εξαιτίας των ακόλουθων λόγων:

- Το κόστος των καμερών ΜΔΕ είναι ακόμη πολλαπλάσιες των καμερών ΣΔΕ.
 Όπως και σε πολλές άλλες πτυχές της τεχνολογίας, είναι σχεδόν σίγουρο ότι κάποια στιγμή στο μέλλον οι κάμερες ΜΔΕ θα επικρατήσουν στην αγορά. Στην παρούσα φάση όμως, το κόστος τους τις καθιστά απαγορευτικές σε εμπορικά προϊόντα και περιορίζει τη χρήση τους σε εφαρμογές όρασης μηχανών.
- 2. Το κόστος των συσκευών ΜΔΕ απεικόνισης (οθόνες) είναι απαγορευτικές σε σχέση με αυτές των ΣΔΕ. Η υπάρχουσα τεχνολογία στον τομέα αυτό δεν είναι τόσο ανεπτυγμένη όσο στον τομέα των αισθητήρων. Ως αποτέλεσμα, οι οθόνες ΜΔΕ είναι λιγότερο διαδεδομένες από τις κάμερες ΜΔΕ. Όπως και πριν, στο μέλλον είναι σχεδόν σίγουρο ότι θα επικρατήσουν οι οθόνες ΜΔΕ. Υπό τις παρούσες συνθήκες όμως, η αλλαγή αυτή φαίνεται να έπεται της επικράτησης των καμερών ΜΔΕ. Έτσι, μεγάλο μέρος της έρευνας έχει στραφεί προς το σχεδιασμό αλγορίθμων που χειρίζονται την απεικόνιση δεδομένων από κάμερες ΜΔΕ σε συσκευές ΣΔΕ (HDR Tone Mapping) (Reinhard, et al., 2005), (Meylan, 2006), (Qiu, et al., 2007).
- Τα περισσότερα ήδη υπάρχοντα συστήματα όρασης βασίζονται σε συσκευές ΣΔΕ. Η αναβάθμισή τους σε συστήματα ΜΔΕ, τις περισσότερες φορές απαιτεί εκτεταμένες τροποποιήσεις, εκτός από την αλλαγή της κάμερας ή της συσκευής απεικόνισης.

3.2.2 Πολλαπλή Έκθεση Σκηνών

Αλλη λύση στο πρόβλημα της σωστής απεικόνισης σκηνών, είναι η συγχώνευση εικόνων της ίδιας σκηνής που τραβήχτηκαν από μια κάμερα ΣΔΕ, σε διαφορετικούς χρόνους έκθεσης. Σύμφωνα με την τεχνική αυτή, λαμβάνεται μια αλληλουχία φωτογραφιών όμοια με αυτή της Εικόνας 3.3. Κάθε φωτογραφία περιέχει πληροφορίες για διαφορετικά τμήματα της εικόνας, ανάλογα με τους χρόνους έκθεσης και τη φωτεινότητα του κάθε τμήματος. Ειδικοί αλγόριθμοι, όπως οι (Debevec, et al., 1997), (Battiato, et al., 2003), και (Goshtasby, 2005) αναλαμβάνουν τη συγχώνευση των φωτογραφιών σε μια τελική εικόνα, που θα συνδυάζει μόνο τις σωστά εκτεθειμένες περιοχές των επιμέρους φωτογραφιών, και όχι τις υποεκτεθειμένες ή υπερεκτεθειμένες περιοχές τους. Η τεχνική αυτή δίνει τη δυνατότητα καταγραφής σκηνών ΜΔΕ χρησιμοποιώντας μόνο εξοπλισμό ΣΔΕ. Έτσι, παρακάμπτεται ένα από τα σημαντικότερα προβλήματα που αναφέρθηκαν πριν. Ωστόσο, εισάγεται ένα καινούριο: ο συνολικός απαραίτητος χρόνος έκθεσης για κάθε εικόναπλαίσιο, είναι το άθροισμα των χρόνων έκθεσης όλων των εικόνων της αλληλουχίας. Ο αριθμός των απαιτούμενων εικόνων της αλληλουχίας ποικίλει ανάλογα με το εύρος των διαβαθμίσεων φωτεινότητας που υπάρχουν στη σκηνή, συνήθως όμως απαιτούνται

Καταγραφή σκηνών ΜΔΕ από κάμερες ΣΔΕ τουλάχιστον 6 (Goshtasby, 2005). Από αυτές, εκείνες που αναδεικνύουν λεπτομέρειες μέσα σε σκοτεινές περιοχές, απαιτούν χρόνους έκθεσης πολλών δευτερολέπτων (π.χ. η πρώτη φωτογραφία της Εικόνας 3.3). Έτσι, το συνολικό άθροισμα χρόνου έκθεσης για την τελική εικόνα-πλαίσιο που θα αναπαριστά μια σκηνή ΜΔΕ, κυμαίνεται συνήθως στα 15 δευτερόλεπτα. Είναι φανερό λοιπόν, ότι η τεχνική αυτή, ακόμη και αν δε λάβουμε υπόψη το χρόνο επεξεργασίας της συγχώνευσης, είναι εξορισμού δύσκολο να χρησιμοποιηθεί για εφαρμογές βίντεο, όπου οι απαιτήσεις είναι 25 πλαίσια το δευτερόλεπτο. Για το λόγο αυτό, οι μέχρι τώρα εφαρμογές της περιορίζονταν μόνο σε στατικές σκηνές, όπου δεν υπάρχει καμία μετακίνηση του ειδώλου. Πρόσφατα, έχουν παρουσιαστεί λύσεις στο πρόβλημα αυτό με τη χρήση διαχωριστή ακτίνας (beam splitter) και πολλαπλών αισθητήρων ΣΔΕ, έτσι ώστε να υπάρχει ταυτόχρονα, πολλαπλή έκθεση ειδώλων (Mantiuk, et al., 2004). Η λύση όμως αυτή επαναφέρει το πρόβλημα του εξειδικευμένου εξοπλισμού, το οποίο όπως είδαμε περιορίζει τις εφαρμογές.

3.2.3 Βελτίωση Εικόνων από κάμερες ΣΔΕ

Μια διαφορετική προσέγγιση στο πρόβλημα της σωστής απεικόνισης σκηνών ΜΔΕ είναι η βελτίωση εικόνων που έχουν ληφθεί από κάμερες ΣΔΕ. Αυτός είναι και ο στόχος της προτεινόμενης μεθόδου. Στην περίπτωση αυτή, ο χρόνος έκθεσης θα πρέπει να είναι της τάξης t₁ και κατά προτίμηση t₂, σύμφωνα με το παράδειγμα της Εικόνας 3.2. Αυτό συμβαίνει γιατί όταν η έκθεση υπερβεί το Οπτικό Δυναμικό Εύρος του αισθητήρα, όλες οι πληροφορίες χάνονται οριστικά λόγω της Τάσης Κορεσμού. Στην περίπτωση του χρόνου έκθεσης t₂ υπάρχει η μικρότερη δυνατή απώλεια οπτικής πληροφορίας, γιατί γίνεται καλύτερη εκμετάλλευση του Δυναμικού Εύρους του αισθητήρα. Για το λόγο αυτό, οι αλγόριθμοι ΑΕ στις ψηφιακές κάμερες και φωτογραφικές μηχανές του εμπορίου, καθορίζουν τους χρόνους έκθεσης ώστε να είναι της τάξης t₂ (Sato, 2006). Για τέτοιους χρόνους έκθεσης, έχουν αναπτυχθεί αλγόριθμοι οι οποίοι βελτιώνουν την εικόνα, προσεγγίζοντας περισσότερο την αρχική σκηνή ΜΔΕ. Αυτή η κατηγορία αλγορίθμων μπορεί να εκτελέσει τις παρακάτω βελτιώσεις στην αρχική εικόνα:

- Να αναδείξει μέρος της πληροφορίας που βρίσκεται στις υποεκτεθειμένες περιοχές.
- 2. Να βελτιώσει την εμφάνιση περιοχών που έχουν προέλθει από λανθασμένη έκθεση, είτε λόγω λανθασμένης επιλογής του χρόνου έκθεσης από τον αλγόριθμο ΑΕ, είτε λόγω της μη γραμμικότητας του αισθητήρα κοντά στην περιοχή της Τάσης Κορεσμού.
- Να βελτιώσει την αντίθεση και την ευκρίνεια της εικόνας που προέρχεται από τη συμπίεση του Δυναμικού της Εύρους λόγω του περιορισμένου χρόνου έκθεσης.

3.2.3.1 Αλγόριθμοι Retinex

Κυριότεροι αλγόριθμοι αυτής της κατηγορίας είναι η οικογένεια των αλγορίθμων Retinex, οι οποίοι επιτυγχάνουν τα παραπάνω σημεία υιοθετώντας χαρακτηριστικά του ΑΟΣ. Ο πρώτος αλγόριθμος Retinex προτάθηκε από τον Edwin Land το 1964 (Land, 1964). Ο Land είναι ο εφευρέτης της στιγμιαίας φωτογραφίας Polaroid και δημιουργός μιας σειράς πειραμάτων που άλλαξαν την επιστημονική αντίληψη για τον τρόπο που λειτουργεί το ΑΟΣ στην αντίληψη του χρώματος και της φωτεινότητας. Το όνομα του αλγορίθμου προέρχεται από τη σύντμηση των λέξεων Retina και Cortex (Αμφιβληστροειδής και Εγκεφαλικός Φλοιός) και στόχος του είναι ο υπολογισμός της Φωτεινότητας (Lightness) σε μια σκηνή, έτσι όπως γίνεται αντιληπτή από το ΑΟΣ (βλέπε κεφάλαιο 1 – Διασαφήνιση Εννοιών).

Ο υπολογισμός της φωτεινότητας από το ΑΟΣ εξαρτάται σημαντικά από τις χωρικές σχέσεις των διαφόρων τμημάτων της σκηνής και δεν είναι μια απλή απεικόνιση της τιμής της φωτοβολίας. Στόχος λοιπόν του αλγόριθμου Retinex είναι η εξαγωγή της φωτεινότητας σε μια εικόνα από τις χωρικές σχέσεις των διάφορων τμημάτων της. Έτσι, προσεγγίζεται ο τρόπος με τον οποίο θα αντιλαμβανόταν ο ανθρώπινος παρατηρητής τη σκηνή και κατά συνέπεια επιτυγχάνεται μεγαλύτερη διαύγεια οπτικής πληροφορίας στις σκοτεινές περιοχές.

Ο αρχικός αλγόριθμος Retinex ενέπνευσε τη δημιουργία δεκάδων παρεμφερών αλγορίθμων. Όλοι αυτοί οι αλγόριθμοι χωρίζονται σε τρεις κύριες κατηγορίες, εκ των οποίων, η πιο διαδεδομένη είναι η Retinex Kέντρου-Περιφέρειας (Center-Surround Retinex) που προτάθηκε για πρώτη φορά το 1986 (Land, 1986). Κύριος σκοπός της ήταν να εισάγει την έννοια του Υποδεκτικού Πεδίου (Receptive Field) των κυττάρων στον υπολογισμό της lightness, ενσωματώνοντας έτσι νευροφυσιολογικά χαρακτηριστικά στον αλγόριθμο. Η μέθοδος Retinex Κέντρου-Περιφέρειας οδήγησε στις μεθόδους Retinex Μιας Κλίμακας (Single Scale Retinex – SSR) (Jobson, et al., 1997) και Retinex Πολλαπλής Κλίμακας (Multi Scale Retinex – MSR) (Jobson, et al., 1997). Από αυτές, η MSR είναι η πλέον διαδεδομένη. Χρησιμοποιείται από τη NASA στη βελτίωση φωτογραφιών από διαστημικές αποστολές (Ret07) και είναι η μόνη που έχει καταφέρει να εδραιωθεί στην εταιρεία TruView (Tru07).

3.2.3.2 Retinex Πολλαπλής Κλίμακας

Η μέθοδος MSR επεξεργάζεται ξεχωριστά κάθε χρωματική συνιστώσα σε διαφορετικές κλίμακες ως εξής:

$$R_{n_i}(x, y) = \log I_i(x, y) - \log[F_n(x, y) * I_i(x, y)]$$
(3.1)

όπου $R_{n_i}(x, y)$ είναι η έξοδος της μεθόδου για την κλίμακα n και τη χρωματική συνιστώσα $i, I_i(x, y)$ είναι η κατανομή της i χρωματικής συνιστώσας και '*' ο τελεστής της συνέλιξης. Η συνάρτηση $F_n(x, y)$ είναι η συνάρτηση της περιφέρειας (surround) για την κλίμακα n, σύμφωνα με την οποία καθορίζεται η επίδραση που θα έχουν τα γειτονικά εικονοστοιχεία στο κεντρικό, ανάλογα με την απόστασή τους από αυτό.

$$F_n(x,y) = K_n e^{-r^2/c_n^2}$$
(3.2)

όπου c_n είναι η σταθερά της Γκαουσιανής κατανομής, n είναι η κλίμακα, και K_n είναι μια σταθερά που επιλέγεται έτσι ώστε:

$$\iint F_n(x,y) \, dx \, dy = 1 \tag{3.3}$$

Αυτό γίνεται για να είναι ισοδύναμοι οι δύο όροι της εξίσωσης (3.1). Το τελικό αποτέλεσμα για πολλές κλίμακες δίνεται από την ακόλουθη εξίσωση:

Retinex Κέντρου-Περιφέρειας

$$R_{MSR_i} = \sum_{n=1}^{N} w_n R_{n_i} \tag{3.4}$$

όπου N είναι ο αριθμός των κλιμάκων, R_{n_i} είναι η έξοδος της i χρωματικής συνιστώσας για την κλίμακα n, και w_n είναι το βάρος με το οποίο συμμετέχει η κάθε κλίμακα στο τελικό αποτέλεσμα. Οι δημιουργοί της MSR κατέληξαν στη χρήση 3 διαφορετικών κλιμάκων, με μέγεθος συναρτήσεων περιφέρειας $c_{n1} = 5$, $c_{n2} = 20$ και $c_{n3} = 240$ εικονοστοιχεία, καθώς και ίδια βάρη ($w_1 = w_2 = w_3 = 0.333$).

Οι εξισώσεις (3.1)-(3.3) ουσιαστικά δείχνουν ότι η καινούρια τιμή για το κάθε Ερμηνεία του εικονοστοιχείο είναι η διαφορά του, με το μέσο όρο της γειτονιάς του, στο λογαριθμικό MSR πεδίο. Επειδή όμως η διαφορά λογαρίθμων είναι ισοδύναμη με τη διαίρεση των ορισμάτων τους, ουσιαστικά πρόκειται για το λόγο της τιμής του κεντρικού εικονοστοιχείου, με το μέσο όρο της γειτονιάς του. Αυτό γίνεται γιατί οι λόγοι φωτοβολίας μεταξύ διαφορετικών περιοχών μιας σκηνής τείνουν να μένουν αμετάβλητοι για διαφορετικά επίπεδα ομοιόμορφου φωτισμού, σε αντίθεση με τις τιμές έντασης που μεταβάλλονται (Jobson, et al., 1997). Η ξεχωριστή επεξεργασία κάθε χρωματικής συνιστώσας έχει τις βάσεις της στα πειράματα του Edwin Land, τα οποία έδειξαν ότι η αντίληψη του χρώματος στο ΑΟΣ είναι η σύγκριση των τριών αντιλήψεων lightness, που προκύπτουν από την ξεχωριστή επεξεργασία της κάθε χρωματικής συνιστώσας (Land, 1977). Το μέγεθος της περιφέρειας γύρω από το κεντρικό εικονοστοιχείο παίζει σημαντικό ρόλο στο τελικό αποτέλεσμα. Μικρές περιφέρειες τείνουν να εξάγουν πολλές λεπτομέρειες στις σκοτεινές περιοχές, υποβαθμίζοντας όμως την ποιότητα των τόνων της εικόνας. Αντίθετα, μεγάλες περιφέρειες διατηρούν τους τόνους της εικόνας, χωρίς να εξάγουν όμως αρκετές λεπτομέρειες. Η ταυτόχρονη χρήση διαφορετικών κλιμάκων συνδυάζει τα θετικά χαρακτηριστικά του κάθε μεγέθους και μειώνει τα αρνητικά τους. Η Εικόνα 3.4 παρουσιάζει το αποτέλεσμα του αλγορίθμου MSR για μια εικόνα με υποέκθεση.

Εικόνα 3.4

Εικόνα με μια υποεκτεθειμένη περιοχή και το αντίστοιχο αποτέλεσμα του αλγορίθμου MSR.



3.2.3.3 Ο αλγόριθμος ΑCE

Ο αλγόριθμος ACE (Automatic Color Equalization) ανήκει, όπως και ο Retinex, στην κατηγορία των αλγορίθμων που υπολογίζουν την εμφάνιση της σκηνής-εικόνας έτσι όπως

θα γινόταν αντιληπτή από το ΑΟΣ (Rizzi, et al., 2003), (Rizzi, et al., 2004). Όπως και στο Retinex, κάθε χρωματική συνιστώσα επεξεργάζεται ξεχωριστά. Έτσι, για κάθε εικονοστοιχείο *p* της χρωματικής συνιστώσας *i* υπάρχουν δύο στάδια επεξεργασίας. Το πρώτο περιγράφεται από την ακόλουθη εξίσωση:

$$R_{i}(p) = \sum_{j \in S, j \neq p} \frac{r[I_{i}(p) - I_{i}(j)]}{d[p, j]}$$
(3.5)

όπου *S* είναι είτε ένα υποσύνολο της εικόνας, είτε ολόκληρη η εικόνα, $d[\cdot]$ μια εξίσωση βάρους ανάλογα με την απόσταση του *p* από το εκάστοτε εικονοστοιχείο και $r[\cdot]$ μια εξίσωση που καθορίζει την επίδραση της διαφοράς τους. Για τη συνάρτηση $d[\cdot]$ χρησιμοποιείται συνήθως η ευκλείδεια απόσταση, για την $r[\cdot]$ η σιγμοειδής, ενώ για καλύτερα αποτελέσματα το *S* αντιπροσωπεύει ολόκληρη την εικόνα I_i . Η εξίσωση (3.5) μπορεί να έχει και αρνητικές τιμές. Για το λόγο αυτό, στο δεύτερο στάδιο επεξεργασίας τα αποτελέσματα R_i αντιστοιχίζονται στο διάστημα [0,255] με τον ακόλουθο τρόπο:

$$O_{i}(p) = \text{round}\left[127.5 + \frac{R_{i}(p) - R_{i}^{mean}}{R_{i}^{max} - R_{i}^{mean}} \times 127.5\right]$$
(3.6)

όπου $O_i(p)$ είναι η τελική τιμή του εικονοστοιχείου p για τη χρωματική συνιστώσα i, και R_i^{mean} και R_i^{max} η μέση και η μέγιστη τιμή του συνόλου των τιμών R_i , αντίστοιχα. Η συνάρτηση round[·] είναι η συνάρτηση της στρογγυλοποίησης.

Η εξίσωση (3.5) στην ουσία υλοποιεί το μηχανισμό Πλευρικής Αναστολής (Lateral Inhibition) που εμφανίζεται σε όλα σχεδόν τα στάδια του ΑΟΣ. Ο μηχανισμός αυτός είναι υπεύθυνος για τη διαφορική μορφή των σημάτων επεξεργασίας σε όλα σχεδόν τα επίπεδα του ΑΟΣ. Έτσι, για κάθε εικονοστοιχείο, ο αλγόριθμος ΑCE επεξεργάζεται το άθροισμα της επίδρασης όλων των εικονοστοιχείων της εικόνας ανάλογα με την απόστασή τους από αυτό. Η εξίσωση (3.6) υλοποιεί ταυτόχρονα δύο βασικούς κανόνες του Προβλήματος Αγκύρωσης (Anchoring Problem) (Ένθετο 3.1): Τον κανόνα του Γκρίζου Κόσμου (Gray World rule) και τον κανόνα του Λευκού Σημείου (White Patch rule). Έτσι, το φάσμα των τιμών του πρώτου σταδίου επεξεργασίας απεικονίζεται σε όλο το δυναμικό εύρος του μέσου απεικόνισης (0-255). Η Εικόνα 3.5 παρουσιάζει το αποτέλεσμα του αλγορίθμου ΑCE.



Ερμηνεία του ACE

Εικόνα 3.5

Εικόνα με μια υποεκτεθειμένη περιοχή και το αντίστοιχο αποτέλεσμα του αλγορίθμου ACE.


 i. Το πρόβλημα της Αγκύρωσης: Δεδομένου ενός εύρους φωτοβολίας μιας σκηνής-εικόνας, με ποιον τρόπο να απεικονίζονται οι τιμές αυτές στην κλίμακα αντιληπτών φωτεινοτήτων; (Gilchrist, et al., 1999).

ii. Κανόνες Αγκύρωσης.

Κανόνας του Λευκού Σημείου (White Patch rule): Η μέγιστη τιμής φωτοβολίας της σκηνής αντιστοιχίζεται στο αντιληπτό άσπρο.

Κανόνας του Γκρίζου κόσμου (Gray World rule): Η μέση τιμή της φωτοβολίας της σκηνής αντιστοιχίζεται στη μεσαία τιμή της κλίμακας αντιληπτής φωτεινότητας (γκρι). **Διπολικός Κανόνας (Bipolar rule):** Η μέγιστη και η ελάχιστη τιμή φωτοβολίας της σκηνής αντιστοιχίζονται στο αντιληπτό άσπρο και στο αντιληπτό μαύρο αντίστοιχα. Εκτός από τους παραπάνω 3 κανόνες υπάρχουν ακόμα πιο πολύπλοκοι, που συνδυάζουν τη φωτοβολία με το μέγεθος της κάθε περιοχής σε σχέση με το συνολικό οπτικό πεδίο (Gilchrist, et al., 1999).

3.2.3.4 Διόρθωση γ

Η διόρθωση γ είναι ένας μη γραμμικός μετασχηματισμός, που εφαρμόζεται στην έξοδο του αισθητήρα της κάμερας, και απεικονίζει τη γραμμική κωδικοποίηση της έντασης του

φωτός σε μια μη γραμμική κλίμακα. Ο μετασχηματισμός που εφαρμόζεται είναι της μορφής

$$f(x) = x^a \tag{3.7}$$

όπου *a* μια σταθερά, και είναι αντίστροφος από το μη γραμμικό μετασχηματισμό $g(x) = x^{1/a}$ που εισάγει η συσκευή απεικόνισης, έτσι ώστε $g(f(x)) = (x^a)^{1/a} = x$. Συνεπώς, οι δύο αντίστροφοι μετασχηματισμοί αναιρούνται, επιτυγχάνοντας καλύτερη απεικόνιση των εικόνων. Η σταθερά *a*, για ιδανικές συνθήκες έκθεσης, έχει συνήθως την τιμή 0.45 (Poynton, 1996). Σε περιπτώσεις όμως όπου υπάρχει υποέκθεση του ειδώλου, η χρήση τιμής μεγαλύτερης του 1 για τη σταθερά *a* μπορεί να διορθώσει αρκετά την εικόνα και να εμφανίσει οπτική πληροφορία που πριν δεν ήταν δυνατόν να γίνει αντιληπτή. Αντίστοιχα, σε περιπτώσεις όπου υπάρχει υποέκθεση του ειδώλου (όχι όμως τόση ώστε να χαθούν τελείως όλες οι πληροφορίες της σκηνής) η χρήση τιμής μικρότερης του 1 μπορεί να διορθώσει εν μέρει την εικόνα. Η Εικόνα 3.7 δείχνει το αποτέλεσμα της εφαρμογής διαφορετικών τιμών διόρθωσης γ σε δύο εικόνες, μια με υποεκτεθειμένη περιοχή και μια με υπερβολική έκθεση.



3.2.3.5 Λογαριθμική κωδικοποίηση

Ανάλογο αποτέλεσμα με τη διόρθωση γ έχει και η κωδικοποίηση της εικόνας σε λογαριθμική κλίμακα. Υποθέτοντας ότι η εικόνα *Ι* έχει θετικές τιμές, η λογαριθμική κωδικοποίησή της *I*_{log} δίνεται από την ακόλουθη εξίσωση (Bovik, 2004).

$$I_{log}(x,y) = \frac{\log[I(x,y)+1]}{\log 256} \times 255$$
(3.8)

Η μονάδα στην εξίσωση (3.8) εξασφαλίζει ότι το όρισμα του λογάριθμου θα είναι μεγαλύτερο ή ίσο του 1, έτσι ώστε οι τιμές που θα προκύψουν να μην είναι αρνητικές.

Το τελικό αποτέλεσμα της λογαριθμικής κωδικοποίησης είναι παρόμοιο με αυτό της διόρθωσης γ για τιμή $a \approx 4$.

3.2.3.6 Εξισορρόπηση ιστογράμματος (Histogram Equalization)

Η εξισορρόπηση ιστογράμματος είναι μια από τις πιο σημαντικές τεχνικές βελτίωσης της εικόνας. Ο βασικός στόχος αυτού του μετασχηματισμού είναι να δημιουργήσει όσο το δυνατόν πιο επίπεδο ιστόγραμμα σε μια εικόνα, γιατί αυτή η μορφή ιστογράμματος περιέχει τη μεγαλύτερη δυνατή πληροφορία (μέγιστη εντροπία) (Bovik, 2004). Έστω $h_I(k)$ το ιστόγραμμα μιας εικόνας *I*, τότε το κανονικοποιημένο ιστόγραμμα $H_I(k)$ θα δίνεται από την ακόλουθη εξίσωση.

$$H_I(k) = \frac{h_I(k)}{N}, \quad \forall k \in [0, K]$$
(3.9)

$$\sum_{k=0}^{K} H_I(k) = 1$$
(3.10)

όπου N ο αριθμός των εικονοστοιχείων της εικόνας I και K ο αριθμός των τμημάτων του ιστογράμματος. Το συσσωρευτικό κανονικοποιημένο ιστόγραμμα (cumulative normalized histogram) $CH_I(k)$ δίνεται από την εξίσωση (3.11).

$$CH_I(r) = \sum_{k=0}^r H_I(k), \quad \forall r \in [0, K]$$
 (3.11)

$$H_{I}(k) = CH_{I}(k) - CH_{I}(k-1)$$
(3.12)

Το συσσωρευτικό κανονικοποιημένο ιστόγραμμα είναι μια μονότονη αύξουσα συνάρτηση με $CH_I(K) = 1$, η οποία έχει την ιδιότητα $CH_I(r) = PR[I(x, y) \le r]$, όπου $PR[\cdot]$ η πιθανότητα του ορίσματος (Bovik, 2004). Η εικόνα με το εξισορροπημένο ιστόγραμμα HE δίνεται εφαρμόζοντας τη συνάρτηση του συσσωρευτικού κανονικοποιημένου ιστογράμματος στην αρχική εικόνα για κάθε εικονοστοιχείο ως εξής:

$$HE(x,y) = CH_I[I(x,y)]$$
(3.13)

Επειδή όμως η εικόνα *HE* δεν έχει τιμές εντός του διαστήματος [0,255] γίνεται γραμμική μετατόπιση του εύρους των τιμών στο διάστημα αυτό (Bovik, 2004).

$$I_{eq}(x,y) = \frac{HE(x,y) - \min[HE]}{\max[HE] - \min[HE]} \times 255$$
(3.14)

όπου I_{eq} η τελική εικόνα και min[HE], max[HE] το ελάχιστο και το μέγιστο της εικόνας HE. Σε περίπτωση που η αρχική εικόνα I είχε υποεκτεθειμένες ή υπερεκτεθιμένες περιοχές, τότε η τελική εικόνα I_{eq} αναδεικνύει τις πληροφορίες εκείνες που δεν ήταν ορατές στις περιοχές αυτές. Η Εικόνα 3.8 δείχνει τα αποτελέσματα της εξισορρόπησης ιστογράμματος για δύο τέτοιες περιπτώσεις.



Εικόνα 3.8

Αποτελέσματα της εξισορρόπησης ιστογράμματος για εικόνες με υποέκθεση και υπερέκθεση.

3.2.4 Αδυναμίες υπαρχόντων λύσεων

Οι υπάρχουσες λύσεις για τη βελτίωση εικόνων που έχουν υποστεί υποέκθεση ή μερική υπερέκθεση σε τμήματά τους παρουσιάζουν μερικά βασικά προβλήματα, που εμποδίζουν την εκτεταμένη χρήση τους.

Πιο συγκεκριμένα, οι καθολικές μέθοδοι (global methods), οι οποίες εφαρμόζουν τον ίδιο μετασχηματισμό σε ολόκληρη την εικόνα (διόρθωση γ, λογαριθμική κωδικοποίηση, εξισορρόπηση ιστογράμματος), διορθώνουν τις περιοχές που έχουν υποστεί λανθασμένη έκθεση, καταστρέφοντας όμως ταυτόχρονα τις περιοχές που είναι σωστά εκτεθειμένες. Έτσι, αναιρείται η όποια βελτίωση επιτυγχάνεται. Η Εικόνα 3.9 δείχνει μια τέτοια περίπτωση φωτογραφίας, η οποία έχει δύο διακριτές περιοχές: μια υποεκτεθειμένη περιοχή (γύρω από το παράθυρο) και μια σωστά εκτεθειμένη (στο εσωτερικό του παραθύρου). Και οι τρεις μέθοδοι, διορθώνουν την υποεκτεθειμένη περιοχή φωτίζοντάς την. Ταυτόχρονα όμως φωτίζουν και τη σωστά εκτεθειμένη περιοχή, με αποτέλεσμα να καταστρέφεται. Παράλληλα με τα αποτελέσματα κάθε μεθόδου στην Εικόνα 3.9 δίνεται και η διαφορά τους με την αρχική εικόνα (αφαίρεση της αρχικής εικόνας από τη διορθωμένη). Έτσι απεικονίζεται καλύτερα ο μετασχηματισμός που εισάγει η εκάστοτε μέθοδος. Είναι φανερό ότι και οι τρεις μέθοδοι αυξάνουν τις τιμές έντασης στο εσωτερικό του παραθύρου κατά 150 περίπου. Αυτό έχει σαν αποτέλεσμα, η σωστά εκτεθειμένη περιοχή της εικόνας, η οποία δε χρειαζόταν καμία μεταβολή, να υποστεί σημαντικές αλλαγές και κατά συνέπεια, οι πληροφορίες που περιλαμβάνει να μην είναι ευδιάκριτες. Για τον παραπάνω λόγο, οι καθολικές μέθοδοι δεν είναι δυνατό να αποτελέσουν εύρωστη λύση στο πρόβλημα της βελτίωσης εικόνων από κάμερες ΣΔΕ.

Καθολικές μέθοδοι



Εικόνα 3.9

Αποτελέσματα των τριών καθολικών μεθόδων για μια εικόνα με δύο διακριτές περιοχές: μια σωστά εκτεθειμένη και μια σε υποέκθεση. Το τρίγωνο σε κάθε εικόνα διαφοράς απεικονίζει την απόχρωση του μηδενός.

> Χωρικά μεταβλητές μέθοδοι

> Halo effects

Οι χωρικά μεταβλητές μέθοδοι (spatially varying methods), όπως η MSR και η ACE, σε αντίθεση με τις καθολικές μεθόδους, εφαρμόζουν διαφορετικό μετασχηματισμό σε κάθε σημείο της εικόνας, ανάλογα με την ευρύτερη γειτονιά της περιοχής. Έτσι, δεν επηρεάζονται το ίδιο οι υποεκτεθειμένες ή υπερεκτεθειμένες περιοχές σε σχέση με τις σωστά εκτεθειμένες. Εντούτοις, υπάρχουν μερικά βασικά προβλήματα τα οποία περιορίζουν την απόδοση των μεθόδων αυτών. Το πρώτο είναι η εμφάνιση φαινομένων άλω (γεν. της λέξης άλως) (halo effects) σε σημεία όπου υπάρχει απότομη μεταβολή των τιμών έντασης. Τα φαινόμενα αυτά έχουν τη μορφή «διάχυσης» φωτεινότητας μεταξύ των δύο τμημάτων, γύρω από τις ακμές. Η Εικόνα 3.10 απεικονίζει μια συνθετική σκηνή στην οποία υπάρχει ένα άσπρο τετράγωνο σε μαύρο φόντο, η οποία ευνοεί την εμφάνιση φαινομένων άλω. Είναι φανερό τόσο ο MSR όσο και ο ACE εμφανίζουν φαινόμενα άλω στο αποτέλεσμά τους. Από τους δύο, ο MSR φαίνεται να επηρεάζεται περισσότερο.



Τα φαινόμενα άλω επηρεάζουν σημαντικά την απόκριση των αλγορίθμων και σε πραγματικές σκηνές. Αυτό φαίνεται στην Εικόνα 3.11.



Εικόνα 3.11

Αποτελέσματα των αλγορίθμων MSR και ΑCΕ για μια εικόνα που ευνοεί την εμφάνιση φαινομένων άλω (στα αποτελέσματα σημειώνονται με το γράμμα 'Α'). Το τρίγωνο σε κάθε εικόνα διαφοράς απεικονίζει την απόχρωση του μηδενός.

Από τους δύο αλγορίθμους, ο MSR διατηρεί καλύτερα τη σωστά εκτεθειμένη περιοχή, όμως εμφανίζει εντονότερα φαινόμενα άλω. Ως αποτέλεσμα, οι σκοτεινές περιοχές γύρω από το παράθυρο, παραμένουν σκοτεινές και δεν αναδεικνύονται οι πληροφορίες τους. Αντίθετα ο ACE εμφανίζει πολύ λιγότερα φαινόμενα άλω, όμως επηρεάζει περισσότερο τη σωστά εκτεθειμένη σκηνή. Όπως φαίνεται από την αφαίρεση της αρχικής εικόνας από το αποτέλεσμά του αλγορίθμου ACE, οι περιοχές εντός του παραθύρου, οι οποίες έπρεπε να είχαν μείνει ανέπαφες, έχουν αυξήσει την έντασή τους κατά 85 μονάδες περίπου. Είναι φανερό λοιπόν ότι στις υπάρχουσες μεθόδους, η αποφυγή των φαινομένων άλω και η προστασία των σωστά εκτεθειμένων περιοχών είναι ιδιότητες συμπληρωματικές και δε συνυπάρχουν ταυτόχρονα.

Πολυπλοκότητα και χρόνοι εκτέλεσης

Εκτός από τα παραπάνω προβλήματα, βασική είναι και η πολυπλοκότητα των αλγορίθμων,
 η οποία επηρεάζει και το χρόνο εκτέλεσής τους. Ο αλγόριθμος MSR έχει πολυπλοκότητα
 O(2*N* log₂[2*N*]), όπου *N* ο αριθμός των εικονοστοιχείων της εικόνα, ενώ ο ACE έχει
 O(*N*²) (Rizzi, et al., 2003), (Rizzi, et al., 2004). Οι πολυπλοκότητες αυτές έχουν σαν αποτέλεσμα μεγάλες καθυστερήσεις στους χρόνους εκτέλεσης των αλγορίθμων. Ως αποτέλεσμα, εικόνες 1 εκατομμυρίου εικονοστοιχείων εκτελούνται σε 10 περίπου δευτερόλεπτα από ένα συμβατικό προσωπικό υπολογιστή με επεξεργαστή Intel Pentium 4 στα 3GHz. Ο χρόνος αυτός αυξάνεται κατά πολύ με το μέγεθος της εικόνας και φτάνει μερικά λεπτά για τη βελτίωση εικόνων άνω των 5 εκατομμυρίων εικονοστοιχείων.

Συνοψίζοντας, είναι φανερό ότι είναι αναγκαία η ύπαρξη ενός αλγορίθμου ο οποίος θα διορθώνει τις περιοχές εικόνων με λανθασμένη έκθεση, χωρίς να επηρεάζει τις σωστά εκτεθειμένες. Ο αλγόριθμος αυτός θα πρέπει να μπορεί να βελτιώσει τόσο την υποέκθεση, όσο και τη μερική υπερέκθεση, χωρίς όμως να εμφανίζει φαινόμενα άλω στις απότομες μεταβάσεις έντασης. Τέλος, η πολυπλοκότητα και ο χρόνος εκτέλεσης του αλγορίθμου θα πρέπει να είναι όσο το δυνατόν μικρότεροι, ώστε να μπορεί να εφαρμοστεί χωρίς δυσκολία σε μεγάλες εικόνες και να δίνει τη δυνατότητα για εκτέλεση σε πραγματικό χρόνο.

3.3 Η λύση του ΑΟΣ

Το πρόβλημα που εμφανίζεται στην όραση μηχανών με το περιορισμένο δυναμικό των καμερών και των συστημάτων επεξεργασίας, πρέπει να το αντιμετωπίσει και το ΑΟΣ. Πιο συγκεκριμένα, το ΑΟΣ είναι ένα σύστημα που αποτελείται από βιολογικούς νευρώνες, ο καθένας εκ των οποίων έχει ένα πεπερασμένο εύρος αποκρίσεων. Παρόλα αυτά, οι νευρώνες αυτοί πρέπει να κωδικοποιήσουν και να επεξεργαστούν ένα τεράστιο εύρος φωτοβολίας που μπορεί να υπάρξει στο εξωτερικό περιβάλλον. Είναι φανερό λοιπόν, ότι το ΑΟΣ έχει βρει λύση στο πρόβλημα αυτό, αφού ο ανθρώπινος παρατηρητής μπορεί να αντιληφθεί πολύ μεγάλο εύρος φωτοβολίας χρησιμοποιώντας τους πεπερασμένου εύρους νευρώνες του.

Η λύση αυτή φαίνεται να βρίσκεται στα δίκτυα κέντρου-περιφέρειας (center-surround networks) που βρίσκονται σε όλα σχεδόν τα επίπεδα του ΑΟΣ. Στα δίκτυα αυτά, επιμέρους νευρώνες χαμηλότερου επιπέδου συνεργάζονται με τέτοιο τρόπο ώστε να σχηματίσουν

υποδεκτικά πεδία (receptive fields) που παρουσιάζουν ανταγωνισμό μεταξύ του κέντρου και της περιφέρειάς τους. Οι πιο χαρακτηριστικοί νευρώνες, με τέτοιου είδους υποδεκτικά πεδία, είναι τα γαγγλιακά κύτταρα (ganglion cells) του αμφιβληστροειδούς. Σύμφωνα με το κοινώς αποδεκτό μαθηματικό μοντέλο του αμφιβληστροειδούς (standard retinal model) (Grossberg, et al., 1988), η απόκριση ενός γαγγλιακού κυττάρου, που βρίσκεται στη θέση (*i*, *j*) ενός πλέγματος από άλλα γαγγλιακά κύτταρα, είναι συνάρτηση του δυναμικού της μεμβράνης του *V_i* και περιγράφεται από την ακόλουθη εξίσωση.

$$\frac{dV_{ij}(t)}{dt} = g_{leak} (V_{rest} - V_{ij}) + C_{ij} (E_{ex} - V_{ij}) + S_{ij} (E_{inh} - V_{ij})$$
(3.15)

όπου g_{leak} είναι η αγωγιμότητα διαρροής (leak conductance) και V_{rest} το δυναμικό ηρεμίας που θα αποκτήσει το κύτταρο λόγω της αγωγιμότητας διαρροής. E_{ex} είναι το διεγερτικό δυναμικό αναστροφής (excitatory reversal potential), που ουσιαστικά είναι και η μεγαλύτερη τιμή δυναμικού V_{ij} που μπορεί να αποκτήσει το κύτταρο λόγω διεγερτικών ερεθισμάτων ($E_{ex} > V_{rest}$). E_{inh} είναι το ανασταλτικό δυναμικό αναστροφής (inhibitory reversal potential), που ουσιαστικά είναι και ο αποκτήσει το κύτταρο λόγω ανασταλτικών ερεθισμάτων ($E_{inh} < V_{rest}$). E_{inh} είναι το ανασταλτικό δυναμικό αναστροφής (inhibitory reversal potential), που ουσιαστικά είναι και η μικρότερη τιμή δυναμικού V_{ij} που μπορεί να αποκτήσει το κύτταρο λόγω ανασταλτικών ερεθισμάτων ($E_{inh} < V_{rest}$). C_{ij} και S_{ij} είναι οι είσοδοι στο κέντρο και στην περιφέρεια του υποδεκτικού πεδίου του κυττάρου και υπολογίζονται από τη συνέλιξη της κατανομής φωτοβολίας L που καταγράφουν οι φωτοϋποδοχείς (photoreceptors) με γκαουσιανά φίλτρα με τυπικές αποκλίσεις σ_c και σ_s αντίστοιχα.

$$C_{ij} = L \otimes G_{\sigma_C} = \sum_{p,q} L_{pq} G_{\sigma_C} \left(i - p, j - q \right)$$
(3.16)

$$S_{ij} = L \otimes G_{\sigma_S} = \sum_{p,q} L_{pq} G_{\sigma_S} \left(i - p, j - q \right)$$
(3.17)

όπου \otimes ο τελεστής της συνέλιξης και G_{σ} η ακόλουθη γκαουσιανή κατανομή.

$$G_{\sigma}(i,j) = \frac{1}{2\pi\sigma^2} exp\left(-\frac{i^2+j^2}{2\sigma^2}\right)$$
(3.18)

με

$$\iint G_{\sigma}(i,j) \, di \, dj = 1$$

Οι τυπικές αποκλίσεις είναι τέτοιες ώστε $\sigma_C < \sigma_S$. Όταν $t \to \infty$ και θεωρώντας ότι $V_{rest} = 0, E_{ex} = +1$ και $E_{inh} = -1$, η εξίσωση (3.15) παίρνει την ακόλουθη μορφή.

$$V_{ij,\infty} = \frac{C_{ij} - S_{ij}}{g_{leak} + C_{ij} + S_{ij}}$$
(3.19)

Η εξίσωση (3.19) είναι πολύ σημαντική γιατί αποτελεί τη βάση πάνω στην οποία στηρίζεται ο αυτόματος ελέγχος κέρδους (automatic gain control) του ΑΟΣ για διαφορετικές συνθήκες φωτισμού. Ο αριθμητής περιέχει τη διαφορά γκαουσιανών (difference of Gaussians – DoG), ενώ ο παρονομαστής το άθροισμα γκαουσιανών (sum of Gaussians – SoG) προσαυξημένο με την τιμή της αγωγιμότητας διαρροής *g*_{leak}. Ο λόγος της διαφοράς γκαουσιανών προς το άθροισμα γκαουσιανών αποτελεί στην ουσία μια μορφή κανονικοποίησης, αφού για οποιαδήποτε τιμή του αριθμητή, ο παρονομαστής θα είναι πάντα μεγαλύτερος, περιορίζοντας το αποτέλεσμα εντός των πεπερασμένων ορίων [0,1). Αυτή η μορφή βιολογικής κανονικοποίησης δίνει τη δυνατότητα στα γαγγλιακά κύτταρα του ΑΟΣ να αποκρίνονται σε ένα μεγάλο εύρος ερεθισμάτων και είναι γνωστότερη ως διακλαδισμένη αναστολή (shunting inhibition) ή διαιρετική αναστολή (divisive inhibition) (Bloomfield, 1974).

Η ύπαρξη της αγωγιμότητας διαρροής g_{leak} στον παρονομαστή παίζει επίσης πολύ σημαντικό ρόλο, γιατί εισάγει την έννοια της προσαρμογής, αφού ο όρος αυτός εξαρτάται από το γενικότερο επίπεδο φωτισμού της σκηνής. Για ένα συγκεκριμένο εύρος εντάσεων, το μέγεθος της g_{leak} σε σχέση με το μέγεθος του αθροίσματος κέντρου και περιφέρειας, καθορίζει τη χαρακτηριστική καμπύλη απόκρισης του κυττάρου. Αυτό φαίνεται στην Εικόνα 3.12.

Εικόνα 3.12

Γραφική παράσταση της εξίσωσης (3.19) για S_{ij} = 1, C_{ij} ∈ [0,255]και διαφορετικές τιμές του g_{leak}. Το διάστημα [0,255] επιλέχθηκε για λόγους απεικόνισης. Εντούτοις, η μορφή του διαγράμματος παραμένει ίδια και στο [0,+∞] για αντίστοιχα μεγαλύτερες τιμές του g_{leak}.

Shunting

inhibition



Είναι φανερό ότι τα γαγγλιακά κύτταρα μπορούν θεωρητικά να δεχτούν οποιοδήποτε μέγεθος ερεθίσματος εισόδου και να το απεικονίσουν στο διάστημα απόκρισής του [0,1), λόγω της διακλαδισμένης αναστολής. Επιπλέον, η απεικόνιση αυτή εξαρτάται από τον παράγοντα προσαρμογής *g*_{leak}. Μικρές συγκριτικά τιμές σε σχέση με το άθροισμα του κέντρου και της περιφέρειας δίνουν αποκρίσεις με μεγάλη μη γραμμικότητα. Όσο μεγαλώνει η τιμή του*g*_{leak}, συγκριτικά πάντα με το άθροισμα κέντρου και περιφέρειας, τόσο πιο γραμμική γίνεται η απόκριση του κυττάρου. Μια πιο εποπτική εικόνα για τη συμπεριφορά της εξίσωσης (3.19) φαίνεται στην Εικόνα 3.13.



Εικόνα 3.13

H τρισδιάστατη απεικόνιση της εξίσωσης (3.19) για $C_{ij}, S_{ij} ∈ [0,255]και$ $g_{leak} = 10.$

Όπως είναι φανερό από την παραπάνω εικόνα, η εξίσωση (3.19) είναι ένας εξαγωγέας ακμών, αφού όποτε το κέντρο έχει ίδια τιμή με την περιφέρεια ($C_{ij} = S_{ij}$), οι τιμές της είναι μηδέν. Εντούτοις, η συμπεριφορά της δεν είναι ίδια για όλες τις συνθήκες φωτισμού. Εδώ πρέπει να τονιστεί ότι η τιμή της περιφέρειας S_{ii} του υποδεκτικού πεδίου ενός γαγγλιακού κυττάρου στην ουσία δείχνει τη μέση τιμή έντασης στην ευρύτερη γειτονιά του, αφού η γκαουσιανή κατανομή της περιφέρειας εκτείνεται σε μεγαλύτερη απόσταση από αυτή του κέντρου. Όπως φαίνεται και από την τρισδιάστατη απεικόνιση, όταν οι συνθήκες φωτισμού είναι χαμηλές ($S_{ij} \to 0$) η μη γραμμικότητα a* ενισχύει την απόκριση για τιμές κέντρου C_{ij} μέχρι 100 περίπου, ενώ δεν επηρεάζει σημαντικά την απόκριση για τιμές κέντρου που είναι μεγαλύτερες από 100. Πρακτικά, αυτό σημαίνει ότι σε περιοχές του οπτικού πεδίου όπου δεν υπάρχει επαρκής φωτισμός, έστω και μικρές διαφορές μεταξύ κέντρου και περιφέρειας παράγουν έντονες αποκρίσεις στην έξοδο του κυττάρου. Αντίθετα, σε περιοχές όπου υπάρχει έντονος φωτισμός ($S_{ij} \rightarrow 255$) η σχεδόν γραμμικότητα a παράγει αποκρίσεις περίπου ανάλογες της διαφοράς μεταξύ κέντρου και περιφέρειας. Έτσι, η περιφέρεια κάθε κυττάρου διαμορφώνει την απόκρισή του ανάλογα με τις τοπικές συνθήκες φωτισμού. Αυτό έχει ως αποτέλεσμα το ΑΟΣ να επιτυγχάνει αύξηση της ευαισθησίας του σε περιοχές χαμηλού φωτισμού.

Ο παραπάνω μηχανισμός αποτελεί τη βάση για τον υπολογισμό της ανακλαστικότητας από το ΑΟΣ. Αυτό φαίνεται στο ακόλουθο παράδειγμα, το οποίο εφαρμόζει την εξίσωση (3.19) σε μια διάσταση. Η Εικόνα 3.14 και ο Πίνακας 3.1 απεικονίζουν μια κατανομή ανακλαστικότητας R(x), η οποία γενικότερα είναι ένα αμετάβλητο χαρακτηριστικό των επιφανειών (βλέπε κεφάλαιο 1), κάτω από δύο διαφορετικούς φωτισμούς: έναν ομοιόμορφο $E_1(x)$ και ένα μεταβαλλόμενο $E_2(x)$. Οι δύο αυτοί τύποι φωτισμού, όταν προσπίπτουν πάνω στην κατανομή R(x) παράγουν δύο διαφορετικές κατανομές φωτοβολίας L(x). Ένας κλασικός εξαγωγέας ακμών, όπως η διαφοράς γκαουσιανών G(x), η οποία ουσιαστικά είναι παρόμοια με τη λαπλασιανή της γκαουσιανής κατανομής (Marr, 1982), (Fiorentini, et al., 1990), επηρεάζεται από τον προσπίπτοντα φωτισμό και εξάγει διαφορετικά αποτελέσματα για τους φωτισμούς $E_1(x)$ και $E_2(x)$. Αντίθετα, η διακλαδισμένη αναστολή F(x), πραγματοποιεί κανονικοποίηση ως προς το τοπικό

επίπεδο φωτισμού, με αποτέλεσμα να εξάγει το ίδιο αποτέλεσμα και στις δύο περιπτώσεις. Το αποτέλεσμα αυτό σχετίζεται άμεσα με την κατανομή της ανακλαστικότητας R(x), η οποία είναι αμετάβλητη και ανεξάρτητη του φωτισμού. Στη συνέχεια, το ΑΟΣ χρησιμοποιεί την πληροφορία που εξήγαγε από τις ακμές της ανακλαστικότητας, και χρησιμοποιώντας μια μορφή διάχυσης που ονομάζεται γέμισμα (filling-in), αναδομεί το σήμα της ανακλαστικότητας (Grossberg, et al., 1988), (Grossberg, 2004). Η διαδικασία του γεμίσματος (που δε φαίνεται στην Εικόνα 3.14) ουσιαστικά διαχέει την πληροφορία που εξάχθηκε γύρω από τις διελεύσεις του μηδενός (zerocrossings), έτσι ώστε να προκύψουν ομοιόμορφες περιοχές. Οι διελεύσεις του μηδενός, αποτελούν τα όρια μέσα στα οποία πραγματοποιείται η διάχυση αυτή. Έτσι, το ΑΟΣ επιτυγχάνει να περιορίσει τις επιπτώσεις του φωτισμού πάνω στις επιφάνειες, υπολογίζοντας την ανακλαστικότητά τους R(x) από τη φωτοβολία τους L(x).



3.4 Προτεινόμενη μέθοδος

Στην ενότητα αυτή, παρουσιάζεται η προτεινόμενη μέθοδος για τη βελτίωση των αποχρώσεων σε εικόνες με προβλήματα έκθεσης από κάμερες ΣΔΕ. Η προτεινόμενη μέθοδος έχει σαν κύριο στόχο να συνδυάσει τα συμπεράσματα της ενότητας 3.2 με τα χαρακτηριστικά του ΑΟΣ που παρουσιάστηκαν στην ενότητα 3.3.

3.4.1 Υιοθέτηση στοιχείων του ΑΟΣ

Η εξίσωση (3.19), αν και είναι η βάση για την εξάλειψη της επίδρασης του προσπίπτοντος φωτισμού, έχει κάποια χαρακτηριστικά που δεν ευνοούν την άμεση εφαρμογή της για τη διόρθωση των αποχρώσεων σε εικόνες. Το κυριότερο είναι ότι ουσιαστικά πρόκειται για έναν εξαγωγέα ακμών. Αυτό σημαίνει ότι στις ομοιόμορφες περιοχές μιας εικόνας, στις οποίες δεν υπάρχει μεταβολή της έντασης, η έξοδός της θα είναι μηδέν. Το ΑΟΣ παρακάμπτει το πρόβλημα αυτό χρησιμοποιώντας το γέμισμα, κατά το οποίο, οι περιοχές αυτές «γεμίζουν» με την πληροφορία από τις περιβάλλουσες ακμές. Η υιοθέτηση όμως μιας τέτοιας αντιμετώπισης για εφαρμογές διόρθωσης αποχρώσεων εισάγει πολλά πρακτικά προβλήματα. Το σημαντικότερο είναι η καθυστέρηση, αφού ακόμη και στο ΑΟΣ η διαδικασία αυτή εισάγει σημαντική καθυστέρηση στη ροή της οπτικής αντίληψης (Paradiso, et al., 1991). Θα πρέπει να λοιπόν να αποφευχθεί η χρήση τεχνικών διάχυσης. Ένα άλλο χαρακτηριστικό της εξίσωσης (3.19) το οποίο δεν ευνοεί την άμεση εφαρμογή της, είναι ότι δε διατηρεί το ίδιο εύρος τιμών, για όλους τους πιθανούς φωτισμούς. Στην Εικόνα 3.13 φαίνεται ξεκάθαρα ότι για μεγάλες τιμές της περιφέρειας S_{ii}, το εύρος των τιμών μειώνεται στο μισό (b). Για τους σκοπούς λοιπόν της εφαρμογής μας πρέπει να σχεδιαστεί μια καινούρια συνάρτηση η οποία να έχει τα ακόλουθα χαρακτηριστικά:

- Θα πρέπει να μπορεί να διορθώσει τις αποχρώσεις τόσο στις υποεκτεθειμένες όσο και στις υπερεκτεθειμένες περιοχές της εικόνας.
- Θα πρέπει να χρησιμοποιεί την αρχιτεκτονική ανταγωνισμού κέντρουπεριφέρειας, ώστε να διατηρεί την τοπικότητα του μετασχηματισμού που θα εφαρμόζει.
- 3. Θα πρέπει να υιοθετήσει το μηχανισμό προσαρμογής της διακλαδισμένης αναστολής, και ιδιαίτερα τη μετάβαση από τη μη γραμμικότητα a* στη γραμμικότητα a (Εικόνα 3.13) ανάλογα με το τοπικό επίπεδο φωτισμού.

Απαραίτητα χαρακτηριστικά της καινούριας συνάρτησης

- Θα πρέπει να μην είναι εξαγωγέας ακμών, ώστε να αποφευχθεί η χρήση τεχνικών διάχυσης.
- Θα πρέπει να διατηρεί σταθερό εύρος τιμών εξόδου για όλες τις τιμές του τοπικού φωτισμού (b=0, Εικόνα 3.13).

Η εξίσωση (3.19) έχει τη γενικότερη μορφή

$$F(x) = \frac{B \cdot x}{A + x} \quad \forall x \in [0, \infty), A, B \in \mathbb{R}^+$$
(3.20)

με $x = C_{ij} - S_{ij}$, B = 1 και $A = g_{leak} + 2S_{ij}$. Η εξίσωση (3.20) απεικονίζει όλες τις τιμές εισόδου από το $[0,\infty)$ στο διάστημα [0, B] με μη γραμμικό τρόπο, ο οποίος εξαρτάται από

το λόγο A/B. Το γράφημά της, για το διάστημα $x \in [0, B]$, φαίνεται στην Εικόνα 3.15i. Στο διάστημα αυτό, η μέγιστη τιμή της F(x) δεν είναι σταθερή για διαφορετικές τιμές του λόγου A/B. Για το λόγο αυτό, μετασχηματίζουμε την εξίσωση (3.20) στην εξίσωση (3.21), η οποία διατηρεί σταθερό μέγιστο για οποιοδήποτε λόγο A/B στο διάστημα [0, B].

$$G(x) = \frac{(B+A) \cdot x}{A+x} \quad \forall x \in [0,B], A, B \in \mathbb{R}^+$$
(3.21)

Η γραφική παράσταση της G(x) φαίνεται στην Εικόνα 3.15ii και δείχνει ότι για διαφορετικούς λόγους A/B υπάρχουν καμπύλες με διαφορετική μη γραμμικότητα, οι οποίες εκτείνονται σε όλο το διάστημα [0, B]. Η μετατροπή αυτή ικανοποιεί την πέμπτη προϋπόθεση που τέθηκε για τα χαρακτηριστικά της καινούριας εξίσωσης που πρέπει να σχεδιαστεί. Αντίστροφη μορφή απεικόνισης παρουσιάζει η συνάρτηση H(x), η οποία είναι συμμετρική της συνάρτησης G(x) ως προς την αρχή των αξόνων. Η γραφική της παράσταση παρουσιάζεται στην Εικόνα 3.15ii.

$$H(x) = \frac{A \cdot x}{A + B - x} \quad \forall x \in [0, B], A, B \in \mathbb{R}^+$$
(3.22)



Οι εξισώσεις (3.21) και (3.22) είναι πολύ σημαντικές, γιατί μπορούν να χρησιμοποιηθούν σαν προσαρμοζόμενες συναρτήσεις απεικόνισης. Μόλις καθοριστεί το *B* ανάλογα με το εύρος των τιμών εισόδου, οι διαφορετικές τιμές του *A* δίνουν καμπύλες με μεταβλητό βαθμό μη γραμμικότητας, οι οποίες εκτείνονται έως και την τέλεια γραμμικότητα. Έτσι, η εκάστοτε τιμή του *A* καθορίζει τρόπο της απεικόνισης μεταξύ των τιμών εισόδου και εξόδου. Η ελεγχόμενη αυτή μετάβαση από καμπύλες με μεγάλο βαθμό μη γραμμικότητας σε γραμμικές καμπύλες, ικανοποιεί την τρίτη προϋπόθεση που τέθηκε για τα απαραίτητα χαρακτηριστικά που πρέπει να έχει η καινούρια εξίσωση. Στην περίπτωση της εφαρμογής για τη βελτίωση των εικόνων από κάμερες ΣΔΕ, οι τιμές έντασης κυμαίνονται στο διάστημα [0,255], οπότε και *B* = 255. Η εξίσωση (3.21) έχει την ιδιότητα να αυξάνει τις τιμές εισόδου που βρίσκονται κοντά στο 0. Για το λόγο αυτό μπορεί να χρησιμοποιηθεί για την αύξηση της έντασης των υποεκτεθειμένων περιοχών της εικόνας. Αντίθετα, η εξίσωση (3.22) μειώνει τις τιμές εισόδου που βρίσκονται κοντά στο όριο *B*. Έτσι, μπορεί να χρησιμοποιηθεί για τη μείωση των τιμών έντασης των υπερεκτεθειμένων περιοχών της εικόνας. Ο συνδυασμός των δύο εξισώσεων απεικόνισης ικανοποιεί την πρώτη προϋπόθεση για την ταυτόχρονη διόρθωση της υποέκθεσης και υπερέκθεσης σε μια εικόνα.

Αν οι εξισώσεις (3.21) και (3.22) εφαρμοστούν αυτούσιες για την ταυτόχρονη διόρθωση της απόχρωσης όλων των εικονοστοιχείων μιας εικόνας χρησιμοποιώντας την ίδια τιμή A, τότε δε θα διαφέρουν σημαντικά από τις καθολικές μεθόδους, αφού ουσιαστικά θα εφαρμόζουν τον ίδιο μετασχηματισμό σε όλα τα τμήματα της εικόνας. Για να ικανοποιηθεί και η δεύτερη προϋπόθεση των απαιτούμενων χαρακτηριστικών της νέας εξίσωσης, η τιμή A θα πρέπει να έχει τοπικό χαρακτήρα και να καθορίζεται ξεχωριστά για κάθε εικονοστοιχείο από της ευρύτερη περιφέρεια του. Θα πρέπει λοιπόν να είναι της μορφής $A = f(S_{ij})$ όπου S_{ij} η τιμή της περιφέρειας του εικονοστοιχείου με συντεταγμένες (i, j). Έτσι, ο μετασχηματισμός που θα εφαρμόζεται σε κάθε εικονοστοιχείο, θα εξαρτάται από την αρχική του τιμή και την τιμή της περιφέρειάς του.

Συνοψίζοντας όλα τα παραπάνω συμπεράσματα, η καινούρια συνάρτηση που πρέπει να σχεδιαστεί, θα είναι μια πολύπλοκη συνάρτηση απεικόνισης, η οποία θα δέχεται ως εισόδους την τιμή του εικονοστοιχείου και την τιμή της περιφέρειάς του και θα εξάγει τη διορθωμένη τιμή του εικονοστοιχείου. Μια τέτοια μορφή συνάρτησης ικανοποιεί και την τέταρτη προϋπόθεση, αφού αποφεύγεται η μορφή του εξαγωγέα ακμών, υπολογίζοντας άμεσα τη νέα τιμή κάθε εικονοστοιχείου και όχι έμμεσα, με τη διάχυση των τιμών γύρω από τις διελεύσεις του μηδενός. Η συνάρτηση απεικόνισης θα χρησιμοποιεί την εξίσωση (3.21) για τη διόρθωση των εικονοστοιχείων στις υποεκτεθειμένες περιοχές και την εξίσωση (3.22) για τη διόρθωση στις υπερεκτεθειμένες περιοχές.

Συμπεράσματα για την καινούρια συνάρτηση απεικόνισης

3.4.2 Αναλυτική περιγραφή της μεθόδου

Το συνολικό μπλοκ διάγραμμα της προτεινόμενης μεθόδου φαίνεται στην Εικόνα 3.16.



3.4.2.1 Μετατροπή χρωματικών χώρων

Η αρχική εικόνα, από το χρωματικό χώρο RGB που βρίσκεται, μετασχηματίζεται στο χώρο YCbCr για να εφαρμοστεί ο μετασχηματισμός διόρθωσης των αποχρώσεων, και στη συνέχεια επαναφέρεται στο χρωματικό χώρο RGB για να οδηγηθεί προς απεικόνιση. Η μετατροπή προς και από το χώρο YCbCr γίνεται βάσει των ακόλουθων εξισώσεων.

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 0.230R + 0.661G + 0.109B \\ -0.101R - 0.338G + 0.439B + 128 \\ 0.439R - 0.399G - 0.040B + 128 \end{bmatrix}$$
(3.23)

$$\begin{bmatrix} R\\ G\\ B \end{bmatrix} = \begin{bmatrix} 1.084Y + 1.793(Cr - 128)\\ 1.084Y - 0.534(Cr - 128) - 0.213(Cb - 128)\\ 1.084Y + 2.115(Cb - 128) \end{bmatrix}$$
(3.24)

Η μετατροπή αυτή επιλέχθηκε για να επιτευχθεί διαχωρισμός μεταξύ της χρωματικής και της αχρωματικής συνιστώσας της εικόνας. Αυτό γίνεται γιατί σκοπός της προτεινόμενης μεθόδου είναι η διόρθωση της υποέκθεσης και υπερέκθεσης και όχι η αλλαγή των χρωμάτων της εικόνας. Έτσι, από τις τρεις συνιστώσες του χώρου YCbCr, ο μετασχηματισμός εφαρμόζεται μόνο στη συνιστώσα *Y*, η οποία περιέχει την αχρωματική πληροφορία. Οι άλλες δύο συνιστώσες που περιέχουν τη χρωματική πληροφορία, *Cb* και *Cr*, παραμένουν ανέπαφες και συνενώνονται με τη διορθωμένη αχρωματική συνιστώσα κατά την επαναφορά στο χώρο RGB. Με τη διαδικασία αυτή τα χρώματα της εικόνας παραμένουν ανέπαφα.

Η επιλογή του χώρου YCbCr έναντι άλλων χρωματικών χώρων που διαχωρίζουν τη χρωματική και αχρωματική συνιστώσα, όπως η οικογένεια των χώρων HSV, HSI και HLS, έγινε για δύο κύριους λόγους. Πρώτον, τα οπτικά αποτελέσματα που έδωσε ο YCbCr έναντι των άλλων χώρων, κατά τη διαδικασία ανάπτυξης της μεθόδου ήταν βελτιωμένα από πλευράς εμφάνισης, αφού τα χρώματα στο τελικό αποτέλεσμα ήταν πιο «φυσικά». Δεύτερο, οι μετασχηματισμοί του χώρου YCbCr είναι γραμμικοί και απλούστεροι από τους μετασχηματισμούς των χώρων HSV, HSI και HLS, οι οποίοι εμπεριέχουν πιο πολύπλοκες πράξεις. Ίδιος είναι επίσης ο λόγος για τον οποίο αποκλείστηκε και ο χώρος L*a*b*, αφού απαιτεί ακόμη πιο δύσκολους μετασχηματισμούς.

3.4.2.2 Κλιμάκωση (0-255)

Μετά το διαχωρισμό χρωματικής και αχρωματικής πληροφορίας, η αχρωματική συνιστώσα Υ κλιμακώνεται γραμμικά, ώστε να καλύψει όλο το φάσμα του εύρους τιμών του καναλιού (0-255).

$$Y'_{ij} = \frac{Y_{ij} - \min[Y]}{\max[Y] - \min[Y]} \times 255$$
(3.25)

όπου min[Y] και max[Y] η ελάχιστη και η μέγιστη τιμή της συνιστώσας Y, Y_{ij} η εκάστοτε τιμή της και Y'_{ij} η μετασχηματισμένη τιμή. Ο μετασχηματισμός αυτός εφαρμόζεται γιατί μεγιστοποιεί την αντίθεση των εικόνων που δεν έχουν συνολικά διευρυμένο ιστόγραμμα, ενώ δεν επηρεάζει εκείνες τις εικόνες που δεν έχουν πρόβλημα αντίθεσης. Επιπλέον, είναι απαραίτητος για το επόμενο στάδιο του υπολογισμού των παραμέτρων της μεθόδου. Στο στάδιο αυτό υπολογίζεται το ποσοστό σκούρων, μεσαίων και φωτεινών αποχρώσεων της εικόνας βάσει δύο σταθερών κατωφλίων (85 και 170), που χωρίζουν την κλίμακα των 255 τιμών σε τρία ίσα τμήματα. Εξασφαλίζοντας ότι όλες οι εικόνες έχουν τιμές που

εκτείνονται σε ολόκληρο το διάστημα [0,255] ουσιαστικά προσαρμόζονται στην ίδια κλίμακα μέτρησης. Έτσι, τα σταθερά κατώφλια αποκτούν καθολικό χαρακτήρα και μπορούν να χρησιμοποιηθούν για τη μέτρηση των αποχρώσεων σε όλες ανεξαρτήτως τις εικόνες. Η Εικόνα 3.17 δείχνει την επίδραση της εξίσωσης (3.25) σε μια εικόνα με συνολικά χαμηλή αντίθεση.



Εικόνα 3.17

Η επίδραση του μετασχηματισμού (3.25) σε μια εικόνα με συνολικά χαμηλή αντίθεση. Το ιστόγραμμα της εικόνας κλιμακώνεται σε όλο το εύρος των τιμών, με αποτέλεσμα να βελτιώνεται η συνολική αντίθεση.

3.4.2.3 Απεικόνιση αποχρώσεων

Η βασική εξίσωση της μεθόδου, η οποία πραγματοποιεί την απεικόνιση των αποχρώσεων, και η οποία σχεδιάστηκε βάσει των συμπερασμάτων που εξήχθησαν στην ενότητα 3.4.1 είναι η ακόλουθη.

$$Y_{out,ij}(Y',S) = \begin{cases} \frac{[B + A(S_{ij})] \cdot Y'_{ij}}{A(S_{ij}) + Y'_{ij}} & \forall S_{ij} < \frac{B}{2} \\ \frac{A(S_{ij}) \cdot Y'_{ij}}{A(S_{ij}) + B - Y'_{ij}} & \forall S_{ij} \ge \frac{B}{2} \end{cases}$$
(3.26)

όπου

$$A(S) = \begin{cases} [M_{dark} + q(S)] \cdot d(S) & \forall S < \frac{B}{2} \\ [M_{bright} + q(B - S)] \cdot d(B - S) & \forall S \ge \frac{B}{2} \end{cases}$$
(3.27)

$$q(x) = \frac{x^2}{lobe}$$
(3.28)

$$d(x) = \frac{\frac{B}{2}}{\frac{B}{2} - x} \quad \forall x \in \left[0, \frac{B}{2}\right)$$
(3.29)

Η εξίσωση (3.26) είναι ο συνδυασμός των εξισώσεων (3.21) και (3.22), στις οποίες ο παράγοντας μη γραμμικότητας A είναι συνάρτηση της περιφέρειας S_{ii} του εκάστοτε εικονοστοιχείου, και η μεταβλητή απεικόνισης x, η τιμή του, Y'_{ij} . Η τρισδιάστατη απεικόνιση της εξίσωσης (3.26) φαίνεται στην Εικόνα 3.18.



ί. Απεικόνιση της εξίσωσης (3.26) σε τρεις διαστάσεις, στην οποία φαίνεται η επίδραση των σταθερών M_{dark} και

> *ii. Η ίδια απεικόνιση* από διαφορετική οπτική γωνία, στην οποία φαίνεται η επίδραση της σταθεράς lobe

Το εύρος Β των τιμών της εικόνας, που στην περίπτωσή μας είναι [0-255], χωρίζεται σε Ερμηνεία της δύο τμήματα. Στο χαμηλότερο, [0, B/2], ενεργοποιείται η συνάρτηση (3.21) η οποία έχει εξίσωσης (3.26) την ιδιότητα να απεικονίζει τις χαμηλές τιμές έντασης των εικονοστοιχείων Y'_{ij} σε υψηλότερες τιμές. Αυτό πραγματοποιείται μέσω της μη γραμμικότητας a*, η οποία

σταδιακά καταλήγει στη γραμμικότητα a, καθώς η τιμή της περιφέρειας S_{ij} του εικονοστοιχείου πλησιάζει στο μέσο της κλίμακας (B/2). Στο άλλο τμήμα, [B/2, B], ενεργοποιείται η συνάρτηση (3.22), η οποία απεικονίζει τις υψηλές τιμές έντασης των εικονοστοιχείων σε χαμηλότερες τιμές. Η μετατροπή αυτή πραγματοποιείται μέσω της μη γραμμικότητας b*, η οποία είναι αντίστροφη της a*, και η οποία εκφυλίζεται στη γραμμικότητα a, καθώς η τιμή της περιφέρειας, του εκάστοτε εικονοστοιχείου, πλησιάζει προς το μέσο της κλίμακας.

Ο παράγοντας μη γραμμικότητας A, ο οποίος καθορίζει τη μορφή των a* και b*, περιγράφεται από την εξίσωση (3.27). Ο παράγοντας αυτός είναι συνάρτηση των δύο μορφών προσαρμογής που διαθέτει το ΑΟΣ: της τοπικής, στην οποία προσαρμόζεται στην τοπική φωτοβολία των περιοχών του οπτικού πεδίου (π.χ. μεμονωμένες σκιές) και της γενικευμένης προσαρμογής, στην οποία προσαρμόζεται στη γενικότερη φωτοβολία ολόκληρου του οπτικού πεδίου (π.χ. όραση στο ημίφως). Η τοπική προσαρμογή καθορίζει τη μορφή των μη γραμμικοτήτων a* και b* ξεχωριστά για κάθε εικονοστοιχείο, ανάλογα με την τιμή έντασης της περιφέρειάς του S_{ij} . Η γενικευμένη προσαρμογή καθορίζει τη γενική μορφή των μη γραμμικοτήτων a* και b*, και πιο συγκεκριμένα, το μέγιστο βαθμό μη γραμμικότητας στις χαμηλές και στις υψηλές τιμές της περιφέρειας S_{ij} . Η ρύθμιση αυτή γίνεται μέσω των σταθερών M_{dark} και M_{bright} . Όσο μικρότερες είναι οι δύο σταθερές αυτές καθορίζονται από τη, ποσοστιαία κατανομή των αποχρώσεων στην εικόνα. Η περιγραφή του υπολογισμού τους θα γίνει στην ενότητα υπολογισμού των παραμέτρων. Η επίδρασή τους στη μορφή της συνάρτησης απεικόνισης φαίνεται στην Εικόνα 3.18i.

Η εξίσωση (3.28) καθορίζει τον τρόπο με τον οποίο γίνεται η μετάβαση από τις μη γραμμικότητες a* και b* στη γραμμικότητα a. Όπως και η εξίσωση (3.27), έτσι και η εξίσωση αυτή εξαρτάται από την τοπική και γενικευμένη προσαρμογή. Ως αποτέλεσμα, η τιμή της περιφέρειας *S_{ij}* του εκάστοτε εικονοστοιχείου εκφράζει την επίδραση της τοπικής προσαρμογής, ενώ η σταθερά *lobe* εκφράζει την επίδραση της γενικευμένης προσαρμογής. Όπως και πριν, η σταθερά *lobe* εξαρτάται από τα στατιστικά στοιχεία της εικόνας και ο υπολογισμός της θα δοθεί στην ενότητα του υπολογισμού των παραμέτρων. Η επίδρασή της στη συνάρτηση απεικόνισης φαίνεται στην Εικόνα 3.18ii. Όσο μεγαλύτερη είναι η τιμή της σταθεράς *lobe*, τόσο πιο γραμμικός είναι ο ρυθμός μετατροπής των μη γραμμικοτήτων a* και b* στη γραμμικότητα a.

Η εξίσωση (3.29) είναι ένας διορθωτικός παράγοντας, ο οποίος είναι απαραίτητος για την ομαλή συνέχεια της συνάρτησης απεικόνισης στο σημείο ($S_{ij} = B/2$), στο οποίο γίνεται η αλλαγή από την εξίσωση (3.21) στην εξίσωση (3.22). Ο ρόλος της είναι τέτοιος, ώστε, καθώς οι τιμές της περιφέρειας S_{ij} πλησιάζουν στο σημείο ($S_{ij} = B/2$), ο διορθωτικός παράγοντας d αυξάνεται μη γραμμικά. Έτσι, όταν πολλαπλασιαστεί με τον παράγοντα A, εξασφαλίζει ότι και οι δύο εξισώσεις (3.21) και (3.22) θα καταλήξουν στην τέλεια γραμμικότητα a, επιτυγχάνοντας ομαλή συνέχεια στο σημείο μετάβασης ($S_{ij} = B/2$). Στην περίπτωση που απουσίαζε ο διορθωτικός παράγοντας d, δε θα υπήρχε ομαλή μετάβαση από την εξίσωση (3.21) στην εξίσωση (3.22), με αποτέλεσμα να δημιουργηθεί ένα «σκαλοπάτι» στο σημείο ($S_{ij} = B/2$) της τρισδιάστατης απεικόνισης της συνάρτησης

Ερμηνεία της εξίσωσης (3.27)

Ερμηνεία της εξίσωσης (3.28)

Ερμηνεία της εξίσωσης (3.29) (3.26). Αποτέλεσμα αυτού θα ήταν να εμφανίζονται ανεπιθύμητες αλλοιώσεις στην απεικόνιση των αποχρώσεων που βρίσκονται γύρω από την τιμή *B*/2.

Η χρήση του ορίσματος (255 – S_{ij}) στις εξισώσεις (3.28) και (3.29) από την εξίσωση (3.27) αποσκοπεί στη δημιουργία «κατοπτρικής συμπεριφοράς» γύρω από το σημείο ($S_{ij} = B/$ 2). Έτσι, τα σημεία (B/2) – 1 και (B/2) + 1 προκαλούν αντίστοιχο αποτέλεσμα στις συναρτήσεις (3.21) και (3.22) αντίστοιχα, υπό την προϋπόθεση ότι $M_{dark} = M_{bright}$. Παρομοίως, το σημείο (B/2) – 2 με το (B/2) + 2, το (B/2) – 3 με το (B/2) + 3 κ.ο.κ.

3.4.2.4 Υπολογισμός της περιφέρειας S_{ij}

Πολύ σημαντικό ρόλο στην ποιότητα του τελικού αποτελέσματος παίζει το μέγεθος της περιφέρειας S_{ij} του κάθε εικονοστοιχείου. Η Εικόνα 3.19 παρουσιάζει την εφαρμογή της μεθόδου, σε μια εικόνα με υποεκτεθειμένη περιοχή για δύο διαφορετικά μεγέθη περιφέρειας.



Η επίπτωση του μεγέθους της περιφέρειας στο τελικό αποτέλεσμα. 'A'= φαινόμενα άλω.



Όπως φαίνεται και από την Εικόνα 3.19, οι περιφέρειες μικρού μεγέθους έχουν σαν αποτέλεσμα τη χαμηλή τοπική αντίθεση στις διορθωμένες περιοχές. Όσο αυξάνεται το μέγεθος της περιφέρειας τόσο αυξάνεται και η τοπική αντίθεση. Παράλληλα με την αύξηση του μεγέθους της περιφέρειας όμως, ανακύπτει και ένα σημαντικό πρόβλημα. Τα φαινόμενα άλω, τα οποία υποβαθμίζουν την αισθητική ποιότητα του τελικού αποτελέσματος και αποκρύπτουν την οπτική πληροφορία στις περιοχές στις οποίες εμφανίζονται. Η Εικόνα 3.20 συνοψίζει τα παραπάνω.



Από τα παραπάνω, είναι φανερό ότι το βέλτιστο αποτέλεσμα επιβάλει τη χρήση μεγάλου μεγέθους περιφέρειας, για την επίτευξη ικανοποιητικής αντίθεσης, σε συνδυασμό με ένα μηχανισμό απαλοιφής των φαινομένων άλω.

Όπως έγινε φανερό από τις Εικόνες 3.10, 3.11 και 3.19, η χρήση μεγάλου μεγέθους περιφέρειας με αμετάβλητο σχήμα, οδηγεί αναπόφευκτα στην εμφάνιση φαινομένων άλω, για όλες τις μεθόδους χωρικής επεξεργασίας αποχρώσεων (MSRCR, ACE, προτεινόμενη μέθοδος). Λύση στο πρόβλημα αυτό προτάθηκε από την (Meylan, et al., 2006), η οποία παρουσίασε ένα φίλτρο για τον υπολογισμό της περιφέρειας που προσαρμόζεται στις ακμές της εικόνας. Η λύση αυτή, εντούτοις, διέπεται από κάποια βασικά προβλήματα. Πρώτον, αναφέρεται στην απεικόνιση εικόνων ΜΔΕ, ενώ η προτεινόμενη μέθοδος στοχεύει στη βελτίωση εικόνων ΣΔΕ. Δεύτερον, παρουσιάζει υψηλή πολυπλοκότητα $O(N + N^2)$, όπου N ο αριθμός των εικονοστοιχείων της εικόνας, η οποία έχει ως αποτέλεσμα μεγάλους χρόνους εκτέλεσης.

Η προτεινόμενη μέθοδος χρησιμοποιεί μια διαφορετική προσέγγιση που βασίζεται στα κύτταρα κλίσης της περιοχής V1 στο ΑΟΣ. Τα κύτταρα αυτά, προσεγγίζονται από τις 60 μάσκες που ήδη έχουν παρουσιαστεί στο κεφάλαιο 2 και πιο συγκεκριμένα στον Πίνακα 2.3. Για λόγους ευκολίας στην ανάγνωση, παρουσιάζονται επίσης στην Εικόνα 3.21.



Οι μάσκες της Εικόνας 3.21 χρησιμοποιούνται για τη δημιουργία του Χάρτη Κλίσεων (Orientation Map – OM), ο οποίος αναφέρθηκε στο μπλοκ διάγραμμα της Εικόνας 3.16. Ο στόχος της κατασκευής του Χάρτη Κλίσεων είναι να υπολογιστεί σε κάθε σημείο της συνιστώσας Y' η τοπική κλίση της κατανομής έντασης, καθώς και η τοπική μέση τιμή της. Αρχικά, η συνιστώσα Y'αναλύεται σε ένα πλέγμα από μη επικαλυπτόμενες περιοχές 10×10 εικονοστοιχείων, όμοιες με το μέγεθος των μασκών. Σε κάθε περιοχή, υπολογίζεται η απόκριση και των 60 μασκών. Η μάσκα με τη μεγαλύτερη απόκριση είναι και αυτή που περιγράφει καλύτερα την τοπική κλίση της έντασης. Έστω (i, j) οι συντεταγμένες του πάνω αριστερού εικονοστοιχείου, μιας από τις μη επικαλυπτόμενες περιοχές του πλέγματος, και (u, v) οι συντεταγμένες της περιοχής αυτής στο Χάρτη Κλίσεων (Εικόνα 3.22). Εφόσον οι μάσκες έχουν σταθερό μέγεθος 10×10 εικονοστοιχείων και εφαρμόζονται σε μη επικαλυπτόμενες περιοχές του συντεταγμένες του πάνω για του χάρτης Κλίσεων είναι 10 φορές μικρότερος στο συνολικό μέγεθος) από τη συνιστώσα Y'.

απόκριση out_{uv}^K της μάσκας K για την περιοχή (u, v) του Χάρτη Κλίσεων, δίνεται από την εξίσωση (3.30).



$$out_{uv}^{K} = \left| M_{uv}^{A_{K}} - M_{uv}^{B_{K}} \right|, \quad u = \frac{i}{10}, v = \frac{j}{10}, \ u, v \in \mathbb{Z}, \quad K \in [1, 60]$$
(3.30)

όπου $M_{uv}^{A_K}$ και $M_{uv}^{B_K}$ η διέγερση του τμήματος A και του τμήματος B της μάσκας Kστην περιοχή (u, v), αντίστοιχα, οι οποίες δίνονται από την ακόλουθη εξίσωση.

$$M_{uv}^{A_K} = \frac{1}{N^{A_K}} \sum_{y=i}^{i+9} \sum_{x=j}^{j+9} Y'_{yx}, \forall Y'_{yx} \in A_K \qquad M_{uv}^{B_K} = \frac{1}{N^{B_K}} \sum_{y=i}^{i+9} \sum_{x=j}^{j+9} Y'_{yx}, \forall Y'_{yx} \in B_K$$
(3.31)

όπου N^{A_K} και N^{B_K} είναι ο αριθμός των εικονοστοιχείων που απαρτίζουν το τμήμα A και το τμήμα B της μάσκας K. Ουσιαστικά, η εξίσωση (3.31) δίνει το μέσο όρο της έντασης των εικονοστοιχείων της συνιστώσας Y'που βρίσκονται στα τμήματα A και B της μάσκας K για την περιοχή (u, v). Η εξίσωση (3.30) δίνει τη διαφορά του μέσου όρου έντασης μεταξύ των δύο τμημάτων κάθε μάσκας. Έτσι, όσο μεγαλύτερη είναι η διαφορά αυτή τόσο περισσότερο η κλίση της μάσκας ταιριάζει με την κλίση της έντασης στην εκάστοτε περιοχή. Για το λόγο αυτό, σε κάθε θέση (u, v) του Χάρτη Κλίσεων OM επιλέγεται η μάσκα εκείνη K', η οποία έχει τη μέγιστη απόκριση $out_{uv}^{K'}$ μεταξύ των 60 συνολικά μασκών για την περιοχή αυτή.

$$OM_{uv}^{A} = M_{uv}^{A_{K'}} \qquad OM_{uv}^{B} = M_{uv}^{B_{K'}}$$
(3.32)

$$K': \max[out_{uv}^{K}]_{K=1}^{60} = out_{uv}^{K'}$$
(3.33)

Έτσι, ο χάρτης κλίσεων OM περιέχει σε κάθε θέση του (u, v) τρία είδη πληροφορίας: την τοπική κλίση της έντασης, η ο οποία δίνεται από τον αριθμό K' της επιλεγμένης μάσκας, καθώς και το μέσο όρο της έντασης των δύο τμημάτων της μάσκας.

Για κάθε εικονοστοιχείο Y'_{ij} της συνιστώσας Y', υπολογίζονται δύο διαφορετικές τιμές περιφέρειας: η $S_{1,ij}$, η οποία αποτελεί την κλασική περιφέρεια, παρόμοια με τους αλγορίθμους MSR και ACE, καθώς και η $S_{2,ij}$ που είναι βασισμένη στο Χάρτη Κλίσεων. Η

τελική τιμή της περιφέρειας S_{ij} για το εκάστοτε εικονοστοιχείο Y'_{ij} υπολογίζεται με παρεμβολή μεταξύ των δύο αυτών διαφορετικών τιμών περιφέρειας και θα περιγραφεί αναλυτικότερα παρακάτω. Η κλασική περιφέρεια $S_{1,ij}$ του εικονοστοιχείου Y'_{ij} , ορίζεται ως «ο μέσος όρος έντασης των εικονοστοιχείων που βρίσκονται σε ένα τετράγωνο πλευράς 51×51 , στο κέντρο του οποίου βρίσκεται το εικονοστοιχείο Y'_{ij} ». Το μέγεθος της περιφέρειας αυτής επιλέχθηκε έτσι ώστε να αποδίδει καλή τοπική αντίθεση στο τελικό αποτέλεσμα (βλέπε Εικόνα 3.19). Μικρότερο μέγεθος περιφέρειας θα οδηγούσε σε μικρότερη αντίθεση στην τελική εικόνα, ενώ για μεγαλύτερο μέγεθος το υπολογιστικό κόστος θα αύξανε πολύ χωρίς αυτό να ακολουθείται από σημαντική βελτίωση στην αντίθεση του τελικού αποτελέσματος. Ο υπολογισμός της περιφέρειας $S_{1,ij}$ δίνεται από την παρακάτω εξίσωση.

$$S_{1,ij} = \frac{1}{51^2} \sum_{y=i-25}^{i+25} \sum_{x=j-25}^{j+25} Y'_{yx}$$
(3.34)

Όπως φάνηκε από την Εικόνα 3.19, το μεγάλο μέγεθος της περιφέρειας $S_{1,ij}$, εκτός από την καλή αντίθεση, οδηγεί και στην εμφάνιση φαινομένων άλω. Για το λόγο αυτό, υπολογίζεται η περιφέρεια $S_{2,ij}$. Η περιφέρεια $S_{2,ij}$ ανήκει σε μια περιοχή H του Χάρτη Κλίσεων OM, η οποία έχει μέγεθος 5×5 μάσκες (ουσιαστικά καλύπτοντας μια περιοχή 50×50 εικονοστοιχείων), και της οποίας η κεντρική μάσκα είναι αυτή στην οποία βρίσκεται το εικονοστοιχείο Y'_{ij} (Εικόνα 3.23). Η περιοχή H τημηματοποιείται σε δύο ξεχωριστές περιοχές E_1 και E_2 χρησιμοποιώντας το τοπικό κατώφλι th_H . Αυτές οι περιοχές καθορίζουν την τιμή της περιφέρειας $S_{2,ij}$ ανάλογα με το πού ανήκει το εικονοστοιχείο Y'_{ij} . Αυτό φαίνεται στις ακόλουθες εξισώσεις.

$$th_H = \frac{OM_H^{max} + OM_H^{min}}{2} \tag{3.35}$$

$$OM_H^{max} = \max[OM_z^G], \qquad OM_H^{min} = \min[OM_z^G], \qquad \forall \ G \in \{A, B\}, z \in H$$
(3.36)

$$ME_1 = \frac{1}{N_{E1}} \sum OM_z^G \ \forall OM_z^G < th_H \qquad ME_2 = \frac{1}{N_{E2}} \sum OM_z^G \ \forall OM_z^G \ge th_H \qquad (3.37)$$

$$S_{2,ij} = \begin{cases} ME_1 \ \epsilon \dot{\alpha} \nu \ Y'_{ij} \in E_1 \\ ME_2 \ \epsilon \dot{\alpha} \nu \ Y'_{ij} \in E_2 \end{cases}$$
(3.38)

όπου N_{E1} και N_{E2} είναι ο συνολικός αριθμός τμημάτων μασκών (A ή B) που σχηματίζουν τις περιοχές E_1 και E_2 , αντίστοιχα. Η εξίσωση (3.36) καθορίζει τη μέγιστη και την ελάχιστη τιμή των τμημάτων των μασκών που ανήκουν στην περιοχή H. Αυτά τα δύο άκρα καθορίζουν το τοπικό κατώφλι th_H στην εξίσωση (3.35), το οποίο είναι απαραίτητο για την τμηματοποίηση της περιοχής H στις περιοχές E_1 και E_2 . Η εξίσωση (3.37) δείχνει ότι τα ME_1 και ME_2 είναι ουσιαστικά ο μέσος όρος έντασης των περιοχών E_1 και E_2 , αντίστοιχα. Η τελική τιμή της προσαρμοσμένης περιφέρειας $S_{2,ij}$ καθορίζεται από τη θέση του εικονοστοιχείου Y'_{ij} ανάλογα με το σε ποια περιοχή (E_1 ή E_2) ανήκει. Κατά συνέπεια, η προσαρμοσμένη περιφέρεια επιτρέπεται να λάβει τιμές μόνο από μια εκ των δύο περιοχών E_1 και E_2 και ποτέ και από τις δύο ταυτόχρονα. Έτσι, δε συμβάλει στη δημιουργία φαινομένων άλω.



Η τελική τιμή της περιφέρειας S_{ij} του εικονοστοιχείου Y'_{ij} , καθορίζεται από τη διαφορά dif_H μεταξύ της μέγιστης τιμής OM_H^{max} και της ελάχιστης τιμής OM_H^{min} των τμημάτων των μασκών, για την περιοχή H.

$$dif_{H} = OM_{H}^{max} - OM_{H}^{min}, \quad dif_{H} \in [0,255]$$
 (3.39)

$$S_{ij} = \frac{S_{1,ij}(255 - dif_H) + S_{2,ij} \cdot dif_H}{255}$$
(3.40)

Η διαφορά dif_H αποτελεί το ρυθμιστικό παράγοντα, ο οποίος καθορίζει με παρεμβολή, την τελική τιμή της περιφέρειας S_{ij} . Σε περίπτωση που υπάρχει οξεία μετάβαση έντασης (sharp intensity transition) εντός της περιοχής H, τότε η διαφορά dif_H αποκτά μεγάλη τιμή, με αποτέλεσμα στην τελική τιμή της περιφέρειας S_{ij} να συμμετέχει η προσαρμοσμένη περιφέρεια $S_{2,ij}$ με μεγαλύτερο βάρος σε σχέση με την κλασική περιφέρεια $S_{1,ij}$. Έτσι, ελαχιστοποιούνται τα φαινόμενα άλω γύρω από τις ακμές. Αντίθετα, σε περίπτωση που δεν υπάρχει έντονη μετάβαση έντασης εντός της περιοχής H, η διαφορά dif_H αποκτά μικρή τιμή, με αποτέλεσμα στην τελική τιμή της τερική τιμή της περιφέρειας S_{ij}

να συμμετέχει η κλασική περιφέρεια $S_{1,ij}$ με μεγαλύτερο βάρος σε σχέση με την προσαρμοσμένη περιφέρεια $S_{2,ij}$. Έτσι, επιτυγχάνεται καλή τοπική αντίθεση στις περιοχές αυτές.

3.4.2.5 Υπολογισμός παραμέτρων

Το μπλοκ υπολογισμού παραμέτρων υπολογίζει τις τιμές των παραμέτρων M_{dark} , M_{bright} και lobe, οι οποίες εξαρτώνται από τα στατιστικά στοιχεία της εκάστοτε εικόνας. Οι πρώτες δύο παράμετροι καθορίζουν τη μορφή των μη-γραμμικοτήτων a* και b* για τιμές περιφέρειας $S_{ij} = 0$ και $S_{ij} = 255$ αντίστοιχα και μπορούν να λάβουν τιμές στο διάστημα $(0,\infty)$. Δε μπορούν να έχουν μηδενική τιμή, γιατί στην περίπτωση αυτή θα μηδένιζαν τον παρονομαστή της εξίσωσης (3.26) για συγκεκριμένες τιμές των S_{ij} και Y'_{ij} . Όταν οι παράμετροι M_{dark} και M_{bright} λαμβάνουν τιμές πολύ κοντά στο μηδέν, έχουν σαν αποτέλεσμα τη δημιουργία πολύ απότομων μη-γραμμικοτήτων a* και b* (Εικόνα 3.18). Κατά συνέπεια, οι υποεκτεθειμένες περιοχές φωτίζονται σημαντικά, ενώ οι υπερεκτεθειμένες περιοχές σκουραίνουν επίσης αρκετά. Αντίθετα, μεγάλες τιμές για τις παραμέτρους M_{dark} και M_{briaht} μετατρέπουν τις μη-γραμμικότητες a* και b* σε γραμμικότητες, με αποτέλεσμα να μη συντελείται βελτίωση των υποεκτεθειμένων και υπερεκτεθειμένων περιοχών. Πρακτικά, τιμές μεγαλύτερες του 300 οδηγούν σε οπτικά μη αντιληπτό αποτέλεσμα στην εικόνα εξόδου. Επιπλέον, τιμές μικρότερες του 30 έχουν σαν αποτέλεσμα να ενισχύουν πέρα από το απαιτούμενο τις υποεκτεθειμένες και υπερεκτεθειμένες περιοχές. Παράλληλα, ενισχύουν σημαντικά το θόρυβο που αναπόφευκτα υπάρχει στις υποεκτεθειμένες περιοχές λόγω του χαμηλού λόγου σήματος προς θόρυβο στις περιοχές χαμηλού φωτισμού. Η Εικόνα 3.24 παρουσιάζει την τομή της τρισδιάστατης παράστασης στην Εικόνα 3.18
ί για το επίπεδο όπου το κέντρο Y'_{ii} είναι ίσο με την περιφέρεια S_{ij} , για διάφορους συνδυασμούς τιμών των παραμέτρων M_{dark} και M_{briaht}, διατηρώντας σταθερή την παράμετρο lobe. Αυτή η τομή δείχνει την απόκριση της μεθόδου για ομοιόμορφες περιοχές των εικόνων, στις οποίες δεν υπάρχουν διακυμάνσεις έντασης και αναπόφευκτα η τιμή του κεντρικού εικονοστοιχείου Υ_{ii} είναι ίση με την τιμή της περιφέρει
άς του $S_{ij}.$ Από τα γραφήματα αυτά φαίνεται καθαρά το ικανοποιητικό εύρος των τιμών των παραμέτρων M_{dark} και M_{briaht} . Για τιμές μικρότερες του 30, εκτός από το πρόβλημα του ενισχυμένου θορύβου στις υποεκτεθειμένες περιοχές, παρουσιάζεται και μια ανεπιθύμητη αντιστροφή των αποχρώσεων: οι σκοτεινές αποχρώσεις στο διάστημα [0,50] απεικονίζονται σε μεγαλύτερη τιμή εξόδου Y_{out} από ότι οι μεσαίες αποχρώσεις του διαστήματος [50,128], οι οποίες κανονικά θα έπρεπε να βρίσκονται ψηλότερα από τις σκοτεινές. Αντίστοιχα, οι πολύ φωτεινές αποχρώσεις στο διάστημα [200,255], οι οποίες θα έπρεπε κανονικά να έχουν τη μεγαλύτερη τιμή, απεικονίζονται χαμηλότερα από ότι οι μεσαίες αποχρώσεις του διαστήματος [128,200]. Αυτό το φαινόμενο εξαφανίζεται για τιμές μεγαλύτερες του 30. Επίσης, για τιμές των παραμέτρων M_{dark} και M_{bright} μεγαλύτερες του 300 δεν υφίσταται σχεδόν καμία μεταβολή στη συνάρτηση απεικόνισης. Κατά συνέπεια, μπορούμε με ασφάλεια να επιλέξουμε το εύρος των τιμών των M_{dark} και M_{briaht} στο διάστημα [30,300]. Οι τελικές τιμές τους καθορίζονται από ένα στατιστικό μέτρο ανάλογο με το κλειδί της εικόνας (key of the image), το οποίο δείχνει κατά πόσο μια εικόνα φαίνεται να έχει σκοτεινές, μεσαίες ή φωτεινές αποχρώσεις.

Καθορισμός των παραμέτρων M_{dark} και M_{bright}



Εικόνα 3.24

Τομές της γραφικής παράστασης της Εικόνας 3.18i στο επίπεδο Y_i'=S_i για διαφορετικούς συνδυασμούς των παραμέτρων M_{dark}, M_{bright} και lobe:

i. lobe=σταθερό

ii. $M_{dark} = M_{bright} = \sigma \tau \alpha \theta \varepsilon \rho \dot{\alpha}$

Η εκτίμηση αυτή των συνολικών αποχρώσεων της εικόνας γίνεται με τον υπολογισμό ενός ποσοστιαίου ιστογράμματος τριών θέσεων (bin_{low} [0,84], bin_{middle} [85,169], bin_{high} [170,255]) το οποίο χωρίζει τις 255 διαβαθμίσεις αποχρώσεων σε τρία αδρά σύνολα: σκοτεινές, μεσαίες και φωτεινές αποχρώσεις. Οι ακόλουθες εξισώσεις ρυθμίζουν γραμμικά τις τιμές των M_{dark} και M_{bright} μεταξύ των τιμών 30, όταν 100% των εικονοστοιχείων έχουν μεσαίες αποχρώσεις.

$$bin_{low} = \frac{100}{py \cdot px} \sum_{i=1}^{py} \sum_{j=1}^{px} u(85 - Y'_{ij})$$
(3.41)

$$M_{dark} = \frac{270}{100} (100 - bin_{low}) + 30 \tag{3.42}$$

$$bin_{high} = \frac{100}{py \cdot px} \sum_{i=1}^{py} \sum_{j=1}^{px} u(Y'_{ij} - 170)$$
(3.43)

$$M_{bright} = \frac{270}{100} \left(100 - bin_{high} \right) + 30 \tag{3.44}$$

όπου u(x) είναι η βηματική συνάρτηση και py, px το μήκος και το πλάτος της εικόνας.

Η παράμετρος lobe μπορεί να λάβει τιμές στο διάστημα (0,∞). Η Εικόνα 3.24ii δείχνει την τομή της τρισδιάστατης παράστασης στην Εικόνα 3.18i, για το επίπεδο, όπου το κέντρο Y'_{ii} είναι ίσο με την περιφέρεια S_{ij} , για διαφορετικές τιμές της παραμέτρου lobe, διατηρώντας σταθερές τις τιμές των M_{dark} και M_{bright} . Η τομή αυτή δείχνει καθαρά τη συμπεριφορά της συνάρτησης απεικόνισης για διαφορετικές τιμές της παραμέτρου lobe. Καθώς οι τιμές της αυξάνονται, οι μεσαίες αποχρώσεις απεικονίζονται όλο και σε πιο μικρό εύρος τιμών εξόδου, μειώνοντας έτσι συνεχώς την αντίθεση της εικόνας εξόδου. Ειδικότερα, όταν lobe =100, οι μεσαίες αποχρώσεις του διαστήματος [50,200], απεικονίζονται σε μια μόνο τιμή εξόδου, στο μέσο της κλίμακας των αποχρώσεων (128). Ως αποτέλεσμα, η τοπική αντίθεση εξαλείφεται. Αυτό το φαινόμενο μείωσης της έντασης γίνεται οπτικά αντιληπτό για τιμές μεγαλύτερες του 30. Για τον παραπάνω λόγο, το εύρος τιμών της παραμέτρου lobe μπορεί με ασφάλεια να θεωρηθεί στο διάστημα [1,30]. Χαμηλές τιμές της παραμέτρου lobe (γύρω στο 1) διατηρούν την αντίθεση στις μεσαίες αποχρώσεις, έχοντας ικανοποιητική κλίση απεικόνισης μεταξύ εισόδου Y'_{ij} και εξόδου Y_{out} . Αντίθετα, μεγάλες τιμές (γύρω στο 30), βελτιώνουν τις σκοτεινές και φωτεινές αποχρώσεις λόγω των μη γραμμικοτήτων στα άκρα της κλίμακας (κοντά στο 0 και στο 255), μειώνοντας ωστόσο την κλίση απεικόνισης μεταξύ εισόδου και εξόδου στις μεσαίες αποχρώσεις. Η τελική τιμή της παραμέτρου lobe καθορίζεται γραμμικά από την τιμή της μεσαίας θέσης του ιστογράμματος τριών θέσεων της εικόνας, η οποία ανταποκρίνεται στο ποσοστό των εικονοστοιχείων που έχουν μεσαίες αποχρώσεις.

$$bin_{middle} = 100 - bin_{low} - bin_{high} \tag{3.45}$$

$$lobe = \frac{29}{100} (100 - bin_{middle}) + 1$$
(3.46)

Είναι σημαντικό να τονιστεί ότι όλα τα αποτελέσματα της προτεινόμενης μεθόδου, που παραθέτονται στη συνέχεια αυτού του κεφαλαίου, έχουν εξαχθεί χρησιμοποιώντας τις αυτόματα ρυθμιζόμενες παραμέτρους των εξισώσεων (3.41)-(3.46), χωρίς επιπλέον ρυθμίσεις.

3.5 Συγκριτικά αποτελέσματα

Η προτεινόμενη μέθοδος συγκρίνεται με δύο άλλες καθιερωμένες μεθόδους, οι οποίες μπορούν να βελτιώσουν τη λανθασμένη έκθεση σε εικόνες ΣΔΕ. Σε αυτές περιλαμβάνεται η MSR (Jobson, et al., 1997), η οποία είναι μια από τις πιο διαδεδομένες μεθόδους της οικογένειας των αλγορίθμων Retinex, και ο αλγόριθμος ACE (Rizzi, et al., 2003), (Rizzi, et al., 2004), ο οποίος υπολογίζει τον τρόπο αντίληψης μιας σκηνής από τον ανθρώπινο

Καθορισμός

παραμέτρου

της

lobe

115

παρατηρητή. Ο αλγορίθμος MSR που χρησιμοποιήθηκε στα πειράματα σύγκρισης είναι αυτός που εμπεριέχεται στο λογισμικό πακέτο επεξεργασίας εικόνας PhotoFlair από την εταιρεία TruView (Tru07), και ο οποίος είναι άμεση υλοποίηση της μεθόδου που περιγράφεται στην εργασία (Jobson, et al., 1997). Επιπλέον, η συγκεκριμένη υλοποίηση είναι βελτιστοποιημένη από πλευράς προγραμματιστικής απόδοσης, ώστε να ελαχιστοποιηθεί ο χρόνος εκτέλεσής της και έτσι να ανταποκρίνεται στις απαιτήσεις ενός εμπορικού προϊόντος. Αυτό ενισχύει την εγκυρότητα στις συγκρίσεις των χρόνων εκτέλεσης που θα παρουσιαστούν στην ενότητα αυτή, αφού όλοι οι αλγόριθμοι που χρησιμοποιούνται βρίσκονται στη βελτιστοποιημένη τους μορφή. Τέλος, το λογισμικό PhotoFlair έχει ήδη χρησιμοποιηθεί για την αξιολόγηση του αλγορίθμου MSR στις εργασίες (Meylan, et al., 2006) και (Meylan, et al., 2007). Οι παράμετροι της μεθόδου MSR που χρησιμοποιήθηκαν για την εξαγωγή όλων των αποτελεσμάτων, είναι αυτοί που δίνονται ως βέλτιστοι από τους συγγραφείς της εργασίας (Jobson, et al., 1997): 3 χωρικές κλίμακες με διαμέτρους 5, 20 και 240 εικονοστοιχεία, καθώς και όμοια βάρη συμμετοχής τους στο τελικό αποτέλεσμα. Η υλοποίηση του αλγορίθμου ΑCE που χρησιμοποιήθηκε στη σύγκριση, δόθηκε από τους ίδιους τους συγγραφείς του αλγορίθμου (Rizzi, et al., 2003), (Rizzi, et al., 2004). Αυτό εξασφαλίζει την αντικειμενικότητα των αποτελεσμάτων που χρησιμοποιήθηκαν στη σύγκριση. Μια απλούστερη υλοποίηση του αλγορίθμου ΑCE, που όμως μπορεί να βρεθεί στη διεύθυνση (ΑCE07). Οι παράμετροι του αλγορίθμου ΑCE που χρησιμοποιήθηκαν, είναι αυτοί που παρέχονται από τους συγγραφείς: παράγοντας υποδειγματοληψίας (sub-sampling factor)=8, κλίση εξόδου (slope)=5, απόσταση άλφα (distance alpha)=0.01 και αναπαραγωγή δυναμικού εύρους με τη χρήση των μεθόδων αγκύρωσης WP+GW (λευκού σημείου και γκρίζου κόσμου ταυτόχρονα). Η προτεινόμενη μέθοδος υλοποιήθηκε σε γλώσσα προγραμματισμού C. Όλες οι εικόνες που χρησιμοποιήθηκαν στις συγκρίσεις, είναι τυπικές φωτογραφίες που ελήφθησαν από εμπορικά μοντέλα ψηφιακών φωτογραφικών μηχανών, διαφόρων εταιρειών, σε καθημερινές συνθήκες. Οι περισσότερες από αυτές έχουν υποεκτεθειμένες ή υπερεκτεθειμένες περιοχές, έτσι ώστε να συγκριθεί η ικανότητα βελτίωσης αποχρώσεων των αλγορίθμων που συγκρίνονται. Εκτός από τη βελτίωση της αντίθεσης στις υποεκτεθειμένες και υπερεκτεθειμένες περιοχές, οι αλγόριθμοι MSR και ACE πραγματοποιούν επίσης και διόρθωση χρώματος. Για το λόγο αυτό, όλες οι εικόνες τροποποιήθηκαν σε κλίμακα του γκρι πριν την εφαρμογή των αλγορίθμων, έτσι ώστε να αποτραπεί οποιαδήποτε συμβολή των διεργασιών διόρθωσης χρώματος στα τελικά αποτελέσματα.

3.5.1 Αποτελέσματα σε εικόνες με υποέκθεση και υπερέκθεση

Η Εικόνα 3.25 παρουσιάζει τα αποτελέσματα των αλγορίθμων σε συνδυασμό με τη διαφορά τους από την αρχική εικόνα. Η συγκεκριμένη εικόνα εισόδου τραβήχτηκε με τη χρήση φλας σε κοντινή απόσταση από τον άνθρωπο σε σκοτεινό περιβάλλον. Σαν αποτέλεσμα, το φόντο είναι υποεκτεθειμένο, το πρόσωπο του ανθρώπου σωστά εκτεθειμένο, ενώ το πουκάμισό του υπερεκτεθειμένο. Σε κάθε διαφορική εικόνα, καταγράφεται ο μέσος όρος έντασης δύο 10×10 περιοχών, οι οποίες βρίσκονται στο λαιμό και στο πουκάμισο του ανθρώπου. Οι περιοχές αυτές έχουν επιλεχθεί επειδή ο λαιμός είναι μια περιοχή σωστά εκτεθειμένη, στην οποία δεν είναι αναγκαία καμιά βελτίωση, ενώ

το πουκάμισο είναι μια περιοχή που χρειάζεται χαμηλότερες τιμές έντασης. Έτσι, φαίνεται η συμπεριφορά των αλγορίθμων στον τρόπο επεξεργασίας αυτών των περιοχών. Εκτός από τις μεθόδους MSR και ACE, περιλαμβάνονται και τα αποτελέσματα της Εξισορρόπησης Ιστογράμματος, καθώς και τριών διαφορετικών τιμών Διόρθωσης γ. Είναι φανερό ότι η Διόρθωση γ δε μπορεί να βελτιώσει επιτυχώς ταυτόχρονα την υποεκτεθειμένη και την υπερεκτεθειμένη περιοχή λόγω της καθολικής μη-γραμμικότητας που εφαρμόζει στην εικόνα. Πιο συγκεκριμένα, η Διόρθωση γ=1.5 δε βελτιώνει αισθητά το υποεκτεθειμένο φόντο, το οποίο χρειάζεται πιο έντονη βελτίωση, ενώ αυξάνει την ένταση της σωστά εκτεθειμένης περιοχής κατά +25 μονάδες και της υπερεκτεθειμένης περιοχής κατά +3 μονάδες. Καθώς η τιμή της Διόρθωσης γ αυξάνει, η σωστά εκτεθειμένη και η υπερεκτεθειμένη περιοχή επηρεάζονται όλο και περισσότερο. Έτσι, η Διόρθωση γ=2.5, αυξάνει την ένταση κατά +49 μονάδες για τη σωστά εκτεθειμένη περιοχή, ενώ κατά +7 μονάδες για την υπερεκτεθειμένη περιοχή. Αυτό μειώνει την ποιότητα του τελικού αποτελέσματος. Η Εξισορρόπηση Ιστογράμματος βελτιώνει σημαντικά το υποεκτεθειμένο φόντο, εμφανίζοντας πληροφορίες που πριν δεν ήταν αντιληπτές. Εντούτοις, αυτή η βελτίωση γίνεται και πάλι σε βάρος της σωστά εκτεθειμένης και υπερεκτεθειμένης περιοχής, οι οποίες έχουν υποβαθμιστεί σημαντικά. Οι περιοχές αυτές έχουν δεχτεί σχεδόν την ίδια λανθασμένη αύξηση έντασης με τη Διόρθωση γ=2.5. Αντίθετα με τις καθολικές μεθόδους, η MSR επιτυγχάνει να βελτιώσει σημαντικά το υπερεκτεθειμένο φόντο, επιτυγχάνοντας καλή τοπική αντίθεση. Επιπλέον, δεν επηρεάζει τη σωστά εκτεθειμένη περιοχή, όπως γινόταν με τις προηγούμενες μεθόδους. Εντούτοις, εμφανίζει σημαντικά φαινόμενα άλω στα σημεία απότομης μετάβασης της έντασης. Τα φαινόμενα αυτά αποκρύπτουν οπτικές πληροφορίες που βρίσκονται στα σημεία εκείνα, και υποβαθμίζουν την αισθητική εικόνα του τελικού αποτελέσματος. Σε αντίθεση με όλες τις προηγούμενες μεθόδους, η μέθοδος ΑCE είναι η μόνη που ελαττώνει την ένταση της υπερεκτεθειμένης περιοχής (-11) και βελτιώνει το υποεκτεθειμένο φόντο, επιτυγχάνοντας καλή τοπική αντίθεση. Παρόλα αυτά, έχει αυξήσει την ένταση της σωστά εκτεθειμένης περιοχής (+19), η οποία δε χρειάζεται καμία αλλαγή. Η προτεινόμενη μέθοδος είναι η μόνη που επεξεργάζεται σωστά και τις τρεις περιοχές. Το φόντο έχει βελτιωθεί σημαντικά, έχοντας καλή τοπική αντίθεση, ενώ η σωστά εκτεθειμένη περιοχή δεν έχει αλλάξει. Επιπλέον, η προτεινόμενη μέθοδος είναι η μόνη η οποία χαμηλώνει σημαντικά την ένταση της υπερεκτεθειμένης περιοχής (-45), αυξάνοντας την τοπική αντίθεση και αποκαλύπτοντας λεπτομέρειες που πριν δεν ήταν διακριτές.

Η Εικόνα 3.26 δείχνει τα αποτελέσματα των αλγορίθμων για μια εικόνα με υπερεκτεθειμένη και υποεκτεθειμένη περιοχή. Επιπλέον, παρουσιάζεται και η γραφική παράσταση για μια μόνο γραμμή (scan-line) του κάθε αποτελέσματος, η οποία διέρχεται ταυτόχρονα και από τις δύο περιοχές. Έτσι, μπορεί να γίνει φανερή η ακριβής επίδραση του κάθε αλγορίθμου στην αρχική εικόνα. Τα αποτελέσματα δείχνουν ότι η προτεινόμενη μέθοδος βελτιώνει με επιτυχία την υποεκτεθειμένη περιοχή, εμφανίζοντας πληροφορίες που πριν δεν ήταν διακριτές. Επιπλέον, δεν εμφανίζει φαινόμενα άλω στο σημείο μετάβασης από την υπερεκτεθειμένη στην υποεκτεθειμένη περιοχή. Το πιο σημαντικό χαρακτηριστικό της όμως, που τη διαφοροποιεί περισσότερο από τις άλλες μεθόδους, είναι ότι βελτιώνει τοπικά την αντίθεση στην υπερεκτεθειμένη περιοχή (σημειώνεται με

Διόρθωση γ

Εξισορρόπηση Ιστογράμματος

MSR

ACE

Προτεινόμενη μέθοδος

Αποτελέσματα για μια γραμμή (scan-line) κύκλο στην Εικόνα 3.26). Αντίθετα, οι άλλες δύο μέθοδοι βελτιώνουν την υποεκτεθειμένη περιοχή, όχι όμως και την υπερεκτεθειμένη. Η MSR εμφανίζει επιπλέον φαινόμενα άλω στη μετάβαση των δύο περιοχών της εικόνας, υποβαθμίζοντας το τελικό αποτέλεσμα.

Αρχική Εικόνα -255 28 διαφορά = διορθωμένη - αρχική Διόρθωση γ=1.5 Διόρθωση γ=2.0 Διόρθωση γ=2.5 0 0 0 +3 +5 +7 +25 +35 +49 Εξισορρόπηση Ιστογράμματος MSR ACE

Εικόνα 3.25

Τα αποτελέσματα όλων των μεθόδων για μια εικόνα με τρεις διακριτές περιοχές: υποεκτεθειμένη, σωστά εκτεθειμένη και υπερεκτεθειμένη. Παράλληλα με τα αποτελέσματα παραθέτονται και οι διαφορές τους από την αρχική εικόνα, ώστε να φαίνεται η επίδραση κάθε αλγορίθμου. Τα φαινόμενα άλω υποδεικνύονται με το γράμμα 'Α'.





Η Εικόνα 3.27 δείχνει τα αποτελέσματα των αλγορίθμων για τρεις εικόνες με σημαντικά προβλήματα έκθεσης σε κάποιες από τις περιοχές τους. Παράλληλα με τα αποτελέσματα δίνονται και τα ιστογράμματά τους, ώστε να φανεί καλύτερα η επίδραση που έχει ο κάθε αλγόριθμος στην εικόνα. Για μια ακόμη φορά φαίνεται ξεκάθαρα η ιδιότητα του MSR να εμφανίζει φαινόμενα άλω. Εκτός από αυτό, τα αποτελέσματά του χαρακτηρίζονται από καλή τοπική αντίθεση και καλή διατήρηση των σωστά εκτεθειμένων περιοχών. Ο αλγόριθμος ΑCE δεν εμφανίζει φαινόμενα άλω στα αποτελέσματά του, τα οποία χαρακτηρίζονται από μικρότερη τοπική αντίθεση σε σχέση με αυτά του MSR. Το βασικότερο πρόβλημα του ΑCE είναι ότι σε ορισμένες περιπτώσεις, όπου υπάρχει πολύ ισχυρή υποέκθεση, οι σωστά εκτεθειμένες περιοχές επηρεάζονται. Αυτό είναι πιο φανερό στις Εικόνες 3.27i και 3.27iii, στις οποίες έχει χαθεί σημαντικό ποσοστό πληροφορίας από τις περιοχές αυτές. Η προτεινόμενη μέθοδος είναι αυτή που δίνει την καλύτερη ευκρίνεια στις υποεκτεθειμένες περιοχές, αφού δεν εξάγει φαινόμενα άλω, ενώ παράλληλα επιτυγχάνει πολύ καλή τοπική αντίθεση. Αυτή η διαφορά είναι φανερή σε όλες τις εικόνες και ιδιαίτερα στην Εικόνες 3.27i και 3.27ii. Επιπλέον, οι σωστά εκτεθειμένες περιοχές δεν επηρεάζονται, διατηρώντας όλες τις λεπτομέρειές τους. Ένα άλλο βασικό στοιχείο της προτεινόμενης μεθόδου είναι ότι δεν επηρεάζει τις περιοχές με πολύ έντονη υποέκθεση, όπου ουσιαστικά δεν υπάρχει οπτική πληροφορία. Αντίθετα, οι άλλες δύο μέθοδοι, στην προσπάθειά τους να εξάγουν πληροφορία από τις περιοχές αυτές ενισχύουν το θόρυβο του αισθητήρα μειώνοντας έτσι τη φυσικότητα του τελικού αποτελέσματος (Εικόνα 3.27iii). Αυτό το φαινόμενο είναι πιο έντονο στην Εικόνα 3.28, όπου η υποέκθεση γύρω από την σωστά εκτεθειμένη περιοχή είναι τόσο έντονη, που κάθε πληροφορία έχει χαθεί. Η μεγέθυνση που παρατίθεται επιβεβαιώνει ότι οι αλγόριθμοι MSR και ACE ενισχύουν το θόρυβο του αισθητήρα στην περιοχή αυτή, ενώ η προτεινόμενη μέθοδος ελάχιστα. Αυτό επαληθεύεται και από τα ιστογράμματα των αποτελεσμάτων. Έτσι, τα ιστογράμματα των MSR και ACE έχουν μεταβληθεί αρκετά λόγω της ενίσχυσης του θορύβου στην υποεκτεθειμένη περιοχή. Αντίθετα, το ιστόγραμμα της προτεινόμενη μεθόδου δεν έχει αλλάξει σημαντικά. Επιπλέον, στην ίδια εικόνα δίνεται και η μεγέθυνση της υπερεκτεθειμένης περιοχής, στην οποία φαίνεται ξεκάθαρα ότι μόνο η προτεινόμενη μέθοδος επιτυγχάνει κάποια βελτίωση.

Περίπτωση πολύ ισχυρής υποέκθεσης





MSR









3.5.2 Αποτελέσματα για άλλες εικόνες

Μέχρι τώρα παρουσιάστηκαν αποτελέσματα για εικόνες με υποεκτεθειμένες και υπερεκτεθειμένες περιοχές. Εντούτοις όμως, δεν είναι απαραίτητο ότι κάθε σκηνή που φωτογραφίζεται έχει δυναμικό εύρος μεγαλύτερο από την κάμερα. Υπάρχουν σκηνές με μικρότερο δυναμικό εύρος από αυτό της κάμερας, οι οποίες έχουν καθολικά χαμηλή αντίθεση και χρειάζονται επεξεργασία για να αναδείξουν τις οπτικές πληροφορίες που έχουν. Επίσης, υπάρχουν εικόνες που έχουν εκ των προτέρων καλή ποιότητα και δε χρειάζονται κάποια επεξεργασία. Είναι σημαντικό λοιπόν ένας αλγόριθμος βελτίωσης αποχρώσεων να μην επηρεάζει τέτοιου είδους εικόνες. Η Εικόνα 3.29 παρουσιάζει τα αποτελέσματα των αλγορίθμων για αυτή την κατηγορία εικόνων. Συγκεκριμένα, η εικόνα 1 έχει μικρότερο δυναμικό εύρος από αυτό της κάμερας, γιατί λήφθηκε σε συνθήκες χαμηλής ορατότητας. Κατά συνέπεια, το ιστόγραμμά της δεν καταλαμβάνει ολόκληρη την κλίμακα εντάσεων και για το λόγο αυτό, έχει καθολικά χαμηλή αντίθεση. Ο αλγόριθμος MSR φαίνεται ότι δε μπορεί να βελτιώσει τέτοιες εικόνες, αφού ουσιαστικά το ιστόγραμμα του αποτελέσματος συνεχίζει να έχει το ίδιο εύρος. Την καλύτερη καθολική αντίθεση την παρουσιάζει ο αλγόριθμος ACE, ο οποίος εκτείνει το αρχικό ιστόγραμμα περισσότερο από όλους τους άλλους αλγορίθμους. Εντούτοις, αυτό γίνεται αποκόπτοντας λεπτομέρειες από τις ακραίες τιμές του ιστογράμματος. Έτσι, οι πληροφορίες που βρίσκονται στον αφρό των κυμάτων και σε ορισμένα μέρη του κάστρου χάνονται. Η προτεινόμενη μέθοδος, έχει μικρότερη καθολική αντίθεση στο τελικό αποτέλεσμά της, όμως διατηρεί όλες τις οπτικές πληροφορίες της σκηνής χωρίς απώλειες. Έτσι, οι λεπτομέρειες στον αφρό των κυμάτων φαίνονται στο τελικό αποτέλεσμα. Παρόμοιο είναι το αποτέλεσμα και για την εικόνα 2, η οποία είναι μια ακτινογραφία. Ο αλγόριθμος ΑCE έχει πάλι μεγαλύτερη καθολική αντίθεση, όμως η προτεινόμενη μέθοδος εμφανίζει περισσότερη πληροφορία στις φωτεινές περιοχές. Έτσι, υπάρχει περισσότερη διαύγεια στην περιοχή της κλείδωσης και του περιγράμματος της σάρκας. Τέλος, η εικόνα 3 δείχνει μια φωτογραφία που δε χρειάζεται κάποια βελτίωση στις αποχρώσεις της, αφού το ιστόγραμμά της εκτείνεται σε ολόκληρη την κλίμακα των εντάσεων και δεν έχει κάποια περιοχή χαμηλής ευκρίνειας. Στην εικόνα αυτή, η προτεινόμενη μέθοδος είναι η μόνη που διατηρεί καλύτερα το αρχικό ιστόγραμμα. Ακολουθεί ο αλγόριθμος MSR, και τέλος ο ACE, ο οποίος αλλάζει σε μεγαλύτερο βαθμό τη μορφή του ιστογράμματος. Αποδεικνύεται λοιπόν, ότι η προτεινόμενη μέθοδος μπορεί να διατηρήσει σχεδόν ανέπαφες τις εικόνες που δε χρειάζονται βελτίωση των αποχρώσεών τους.


Εικόνα 3.29

Αποτελέσματα των αλγορίθμων για εικόνες χωρίς υποέκθεση ή υπερέκθεση.

3.5.3 Ψυχοφυσιολογική σύγκριση

Δυστυχώς δεν υπάρχει ένα αντικειμενικό αριθμητικό μέτρο, το οποίο να εκφράζει τη συνολική ποιότητα μιας εικόνας, έτσι όπως γίνεται αντιληπτή από τον ανθρώπινο παρατηρητή. Για το λόγο αυτό, τα αποτελέσματα των αλγορίθμων χρησιμοποιήθηκαν σε μιας ψυχοφυσιολογική μελέτη, έτσι ώστε να αποκαλυφθούν τα υποκειμενικά ποιοτικά χαρακτηριστικά των αλγορίθμων. Ψυχοφυσιολογικά πειράματα χρησιμοποιούνται εκτεταμένα για την αξιολόγηση αλγορίθμων, οι οποίοι συμπιέζουν το δυναμικό εύρος ΜΔΕ εικόνων, ώστε να απεικονιστούν σε συσκευές ΣΔΕ (Drago, et al., 2003), (Kuang, et al., 2004), (Ledda, et al., 2005), (Yoshida, et al., 2005), (Čadík, et al., 2006). Η μεθοδολογία αυτής της σύγκρισης μπορεί να χρησιμοποιηθεί επίσης και στην περίπτωσή μας.

Βαθμολόγηση και κατάταξη Δύο είναι οι κύριες προσεγγίσεις που ακολουθούνται σε αυτά τα πειράματα: η βαθμολόγηση (rating) και η κατάταξη (ranking). Στη βαθμολόγηση, οι παρατηρητές συγκρίνουν τα αποτελέσματα των αλγορίθμων με την ιδανική εικόνα (ground truth image) και τα βαθμολογούν βάσει κάποιας συγκεκριμένης ιδιότητας. Ως ιδανική εικόνα μπορεί να χρησιμοποιηθεί είτε η ίδια η σκηνή, είτε μια εικόνα ΜΔΕ προβαλλόμενη σε μια οθόνη ΜΔΕ. Στην κατάταξη, δεν υπάρχει ιδανική εικόνα και οι παρατηρητές κατατάσσουν τα αποτελέσματα των αλγορίθμων βάσει κάποιων συγκεκριμένων ιδιοτήτων. Στην προκειμένη περίπτωση, στόχος μας είναι η σύγκριση αλγορίθμων που εφαρμόζονται σε εικόνες ΣΔΕ. Κατά συνέπεια, δεν είναι δυνατή η ύπαρξη ιδανικής εικόνας, αφού οι εικόνες ΣΔΕ έχουν εξ' ορισμού υποβαθμιστεί από το μικρότερο δυναμικό εύρος της κάμερας. Για το λόγο αυτό, η προσέγγιση της κατάταξης είναι πιο κατάλληλη.

Οι ιδιότητες βάσει των οποίων οι παρατηρητές κλήθηκαν να κατατάξουν τα αποτελέσματα των αλγορίθμων είναι πέντε: συνολική φωτεινότητα, συνολική αντίθεση, αναπαραγωγή λεπτομερειών στις σκοτεινές περιοχές, αναπαραγωγή λεπτομερειών στις φωτεινές περιοχές και φυσικότητα. Οι ίδιες ακριβώς ιδιότητες, εκτός του χρώματος, το οποίο δε μας ενδιαφέρει στην προκειμένη περίπτωση, έχουν χρησιμοποιηθεί και στις εργασίες (Yoshida, et al., 2005) και (Čadík, και συν., 2006). Είκοσι παρατηρητές, δέκα άντρες και δέκα γυναίκες, διαφόρων ηλικιών, συμμετείχαν στο πείραμα. Παρατήρησαν τα αποτελέσματα των αλγορίθμων για είκοσι ΣΔΕ εικόνες σε αποχρώσεις του γκρι, οι οποίες είχαν υποεκτεθειμένες και υπερεκτεθειμένες περιοχές, όμοιες με αυτές που παρουσιάστηκαν στις Εικόνες 3.25 έως 3.28. Κατά συνέπεια, τα συμπεράσματα του πειράματος αυτού αναφέρονται μόνο για εικόνες με προβλήματα έκθεσης και όχι συνολικά για όλα τα είδη των εικόνων. Η παρατήρηση έγινε σε μια διαβαθμισμένη CRT οθόνη, σε σκοτεινό δωμάτιο και ελεγχόμενες συνθήκες περιβάλλοντος φωτισμού. Οι παρατηρητές κλήθηκαν να κατατάξουν τους τρεις αλγορίθμους (MSR, ACE και προτεινόμενη μέθοδο) με έναν ακέραιο αριθμό μεταξύ 1-3: 1 για τον καλύτερο και 3 για το χειρότερο.

Η Εικόνα 3.30 συνοψίζει τα αποτελέσματα του ψυχοφυσιολογικού ελέγχου, μαζί με το τυπικό σφάλμα τους. Όπως φαίνεται, η προτεινόμενη μέθοδος επιτυγχάνει την καλύτερη μέση κατάταξη για τις ιδιότητες της συνολικής φωτεινότητας, της αναπαραγωγής λεπτομερειών στις σκοτεινές και φωτεινές περιοχές, καθώς και της φυσικότητας. Αυτό έρχεται σε πλήρη συμφωνία με τα αποτελέσματα των προηγούμενων ενοτήτων, στις οποίες φάνηκε ότι η προτεινόμενη μέθοδος αναδεικνύει καλύτερα τις λεπτομέρειες στις υποεκτεθειμένες και υπερεκτεθειμένες περιοχές της εικόνας. Επιπλέον, η απουσία φαινομένων άλω και η προστασία των σωστά εκτεθειμένων περιοχών οδήγησε την προτεινόμενη μέθοδο να καταταχθεί από τους παρατηρητές πρώτη στην κατηγορία της φυσικότητας. Αυτό σημαίνει ότι τα αποτελέσματα της προτεινόμενης μεθόδου έχουν μεγαλύτερη πιθανότητα να θεωρηθούν πιο «φυσικά» από έναν ανθρώπινο παρατηρητή, σε σχέση με τα αποτελέσματα των άλλων μεθόδων. Η προτεινόμενη μέθοδος ξεπεράστηκε μόνο από τη MSR στην κατηγορία της συνολικής αντίθεσης, στην οποία κατατάχθηκε δεύτερη. Σύμφωνα με τους παρατηρητές, η MSR παρουσιάζει τη μεγαλύτερη συνολική αντίθεση, καθώς και ικανοποιητική αναπαραγωγή λεπτομερειών στις σκοτεινές περιοχές, αφού κατατάχθηκε δεύτερη στην κατηγορία αυτή. Στις ιδιότητες της αναπαραγωγής λεπτομερειών σε φωτεινές περιοχές, συνολική φωτεινότητα και φυσικότητα, η MSR κατατάσσεται τρίτη. Ο κύριος λόγος γι αυτό, είναι τα φαινόμενα άλω που παρουσιάζει, καθώς και η αδυναμία για βελτίωση των υπερεκτεθειμένων περιοχών. Ο αλγόριθμος ACE κατατάσσεται δεύτερος στις κατηγορίες της συνολικής φωτεινότητας, αναπαραγωγής λεπτομερειών στις φωτεινές περιοχές και στη φυσικότητα. Η κατάταξη αυτή οφείλεται κατά πάσα πιθανότητα στην απουσία φαινομένων άλω και στη δυνατότητα διόρθωσης υπερεκτεθειμένων περιοχών. Τέλος, κατατάσσεται τρίτος στην αναπαραγωγή λεπτομερειών στις σκοτεινές περιοχές, αφού όπως φάνηκε και στις προηγούμενες εικόνες, ο ACE δε βελτιώνει σε ικανοποιητικό βαθμό τις υποεκτεθειμένες περιοχές.

Η διαδικασία του πειράματος

Ανάλυση αποτελεσμάτων



3.5.4 Πολυπλοκότητα και χρόνος εκτέλεσης

Σημαντικό ρόλο στη συνολική εικόνα του κάθε αλγορίθμου παίζει ο χρόνος εκτέλεσής του. Η προτεινόμενη μέθοδος έχει πολυπλοκότητα της τάξεως O(N), όπου N ο αριθμός των εικονοστοιχείων της εικόνας. Αυτό σημαίνει ότι ο χρόνος εκτέλεσης της προτεινόμενης μεθόδου αυξάνεται γραμμικά σε σχέση με το μέγεθος της εικόνας. Αντίθετα, ο αυθεντικός αλγόριθμος ACE έχει αναφερθεί ότι έχει πολυπλοκότητα $O(N^2)$ (Rizzi, et al., 2003), (Rizzi, et al., 2004). Εντούτοις, χρησιμοποιώντας την τεχνική του Τοπικού Γραμμικού Πίνακα Απεικόνισης (Local Linear LUT – LLL) (Gatta, et al., 2006), η οποία εφαρμόζει τον αλγόριθμο σε μια μικρότερη εκδοχή της αρχική εικόνας, ο χρόνος εκτέλεσής του βελτιώνεται σημαντικά. Ο αλγόριθμος MSR έχει πολυπλοκότητα $O(N \log_2 N)$. Αυτή η διαφορά στις πολυπλοκότητες αντανακλάται και στους μέσους χρόνους εκτέλεσης των τριών αλγορίθμων, οι οποίοι παρατίθενται στον Πίνακα 3.2. Προτεινόμενη μέθοδος

Αλγόριθμος	Μέγεθος Εικόνας						
	640×480	1MP	2MP	3MP			
ΑCE+LLL (υποδειγματοληψία=8)	4 sec	40 sec	152 sec	339 sec			
ΑCE+LLL (υποδειγματοληψία=16)	1 sec	7 sec	19 sec	36 sec			
MSR	2 sec	5 sec	12 sec	20 sec			

1 sec

2.1 sec

3.3 sec

0.3 sec

Πίνακας 3.2

Χρόνοι εκτέλεσης των τριών αλγορίθμων για διάφορα μεγέθη εικόνας από έναν Pentium4 στα 3 GHz σε περιβάλλον Windows XP.

Οι χρόνοι του Πίνακα 3.2 αναφέρονται σε εκτέλεση από έναν επεξεργαστή Pentium 4 στα 3 GHz σε λειτουργικό σύστημα Windows XP. Ο χρόνος εκτέλεσης της προτεινόμενης μεθόδου είναι ο μικρότερος από όλους. Αυτό είναι άμεση συνέπεια της πολυπλοκότητας O(N) που παρουσιάζει. Ο ACE, που είναι ο αλγόριθμος με τη μεγαλύτερη πολυπλοκότητα, έχει και το μεγαλύτερο χρόνο εκτέλεσης, ακόμα και όταν χρησιμοποιηθεί η τεχνική επιτάχυνσης LLL. Ο παράγοντας υποδειγματοληψίας d της τεχνικής LLL παίζει πολύ σημαντικό ρόλο στον τελικό χρόνο εκτέλεσης. Όταν d = 8 που σημαίνει ότι κάθε πλευρά της εικόνας διαιρείται κατά 8 και κατά συνέπεια, ο αλγόριθμος εφαρμόζεται σε μια εικόνα 64 φορές μικρότερη, η επιτάχυνση που επιτυγχάνεται δεν είναι αρκετή για εικόνες που πλησιάζουν τα 2 εκατομμύρια εικονοστοιχεία. Σε τέτοιες περιπτώσεις, μεγαλύτεροι παράγοντες υποδειγματοληψίας μπορούν να δώσουν μεγαλύτερες επιταχύνσεις. Εντούτοις, όπως είναι αναμενόμενο, η επιτάχυνση στο χρόνο εκτέλεσης του αλγορίθμου γίνεται σε βάρος των υψηλών συχνοτήτων της εικόνας, οι οποίες χάνονται λόγω της υποδειγματοληψίας. Ο αλγόριθμος MSR, όντας ο δεύτερος σε πολυπλοκότητα, έχει το δεύτερο μεγαλύτερο χρόνο εκτέλεσης. Είναι σημαντικό να αναφερθεί όμως ότι οι αλγόριθμοι ACE και MSR εφαρμόζονται και στις τρεις χρωματικές συνιστώσες της εικόνας, ενώ αντίθετα η προτεινόμενη μέθοδος μόνο στη συνιστώσα της φωτοβολίας. Παρόλα αυτά, ακόμα και αν οι χρόνοι εκτέλεσης των αλγορίθμων ACE και MSR διαιρεθούν με έναν παράγοντα 3, προσεγγίζοντας έτσι το χρόνο εκτέλεσης για μια μόνο χρωματική συνιστώσα, οι χρόνοι που προκύπτουν εξακολουθούν να είναι μεγαλύτεροι αυτών της προτεινόμενης μεθόδου.

3.6 Συμπεράσματα

Στο κεφάλαιο αυτό παρουσιάστηκε μια νέα μέθοδος για τη βελτίωση των αποχρώσεων σε εικόνες ΣΔΕ, οι οποίες έχουν υποέκθεση, υπερέκθεση ή και τα δύο ταυτόχρονα, σε ορισμένα από τα τμήματά τους. Βασικός στόχος της προτεινόμενης μεθόδου είναι να

βελτιώσει τα τμήματα αυτά των εικόνων χωρίς να επηρεάσει τις σωστά εκτεθειμένες περιοχές. Η βελτίωση πραγματοποιείται με την αύξηση της τοπικής αντίθεσης σε περιοχές με υποέκθεση και υπερέκθεση. Στα πλαίσια της προτεινόμενης μεθόδου, παρουσιάστηκε μια σειρά από πρωτότυπα στοιχεία:

- Μια νέα συνάρτηση απεικόνισης αποχρώσεων, (εξίσωση (3.26)), βασισμένη στη διακλαδισμένη αναστολή των δικτύων ανταγωνισμού κέντρου-περιφέρειας του ΑΟΣ. Η συνάρτηση αυτή υιοθετεί στοιχεία, όπως ρυθμιζόμενες μη γραμμικότητες, χωρίς να εμπίπτει στα πολύπλοκα νευροβιολογικά μοντέλα που μοντελοποιούν την αντίληψη της φωτεινότητας.
- 2. Ένας νέος τρόπος υπολογισμού της περιφέρειας κάθε εικονοστοιχείου, ο οποίος βασίζεται στο συνδυασμό του κλασικού μέσου όρου και μιας περιφέρειας προσαρμοσμένης στην κατανομή εντάσεων της εικόνας. Η προσαρμοσμένη αυτή περιφέρεια βασίζεται στα κύτταρα κλίσης της περιοχής V1 του ΑΟΣ και είναι αρκετά γρήγορη, αφού δεν ενεργεί σε μεμονωμένα εικονοστοιχεία, αλλά σε 10×10 περιοχές της εικόνας.
- 3. Δύο νέες συναρτήσεις απεικόνισης (εξισώσεις (3.21) και (3.22)), οι οποίες μπορούν να χρησιμοποιηθούν για την ελεγχόμενη αλλαγή της σχέσης μεταξύ εισόδου και εξόδου. Οι συναρτήσεις αυτές δεν περιορίζονται μόνο στις εντάσεις των εικονοστοιχείων, αλλά και σε οποιοδήποτε άλλο μονοδιάστατο σήμα. Επίσης, σε αντίθεση με την εκθετική και λογαριθμική απεικόνιση, οι συναρτήσεις αυτές έχουν συμμετρική καμπύλη απεικόνισης ως προς τον άξονα y=x.

Η προτεινόμενη μέθοδος συγκρίθηκε με δύο άλλες καθιερωμένες μεθόδους στο χώρο των βιολογικά εμπνευσμένων μεθόδων βελτίωσης εικόνας. Η σύγκριση έδειξε ότι η προτεινόμενη μέθοδος παρουσιάζει τα παρακάτω πλεονεκτήματα έναντι των άλλων δύο:

- Μπορεί να βελτιώνει ταυτόχρονα τις αποχρώσεις τόσο στις υποεκτεθειμένες όσο και στις υπερεκτεθειμένες περιοχές των εικόνων.
- Η βελτίωση αυτή επιτυγχάνεται χωρίς την επιβάρυνση των σωστά εκτεθειμένων περιοχών.
- 3. Δεν εξάγει φαινόμενα άλω στις απότομες μεταβάσεις έντασης.
- Επηρεάζει λιγότερο τις ισχυρά υποεκτεθειμένες περιοχές, στις οποίες κάθε οπτική πληροφορία έχει χαθεί μέσα στο σκοτεινό θόρυβο.
- 5. Η αυτόματη επιλογή παραμέτρων εξασφαλίζει την ευρωστία της προτεινόμενης μεθόδου καθώς και τα ικανοποιητικά αποτελέσματά της σε ένα μεγάλο σύνολο εικόνων, χωρίς να απαιτείται κάποια επιπλέον ρύθμιση.
- Επηρεάζει ελάχιστα εικόνες που δεν απαιτούν βελτίωση και διατηρεί την αρχική μορφή του ιστογράμματος για τις περιπτώσεις αυτές.
- Σε εικόνες με συμπιεσμένο ιστόγραμμα, εκτείνει τις τιμές τις εικόνας σε ολόκληρο το εύρος εντάσεων, χωρίς να αποκόπτει οποιαδήποτε οπτική πληροφορία.
- Δίνει αποτελέσματα που χαρακτηρίζονται από τον ανθρώπινο παρατηρητή ως «πιο φυσικά».
- Τα αποτελέσματά της έχουν μεγαλύτερη ευκρίνεια οπτικών λεπτομερειών στις υποεκτεθειμένες και υπερεκτεθειμένες περιοχές.

 Είναι ταχύτερη στην εκτέλεση από τους άλλους αλγορίθμους, λόγω της μικρότερης πολυπλοκότητάς της.

Λαμβάνοντας υπόψη τα παραπάνω, η προτεινόμενη μέθοδος μπορεί να χρησιμοποιηθεί με επιτυχία σε εφαρμογές βελτιστοποίησης εικόνας που εξαρτώνται σημαντικά από το χρόνο εκτέλεσης. Τέτοιες εφαρμογές είναι αυτόματα συστήματα όρασης που απαιτείται να λειτουργούν σε περιβάλλοντα μη ελεγχόμενου φωτισμού, στα οποία η εναλλαγή του φωτισμού μπορεί να επηρεάσει σημαντικά τους μετέπειτα αλγορίθμους επεξεργασίας εικόνας. Παραδείγματα τέτοιων εφαρμογών είναι:

- Συστήματα αναγνώρισης προσώπων, στα οποία οι σκιάσεις το προσπίπτοντος φωτισμού μπορούν να επηρεάσουν σημαντικά την επιτυχία της αναγνώρισης.
- Αυτόματα συστήματα πλοήγησης.
- Ρομποτική όραση

Παράλληλα με τις προαναφερθείσες εφαρμογές, η προτεινόμενη μέθοδος μπορεί να χρησιμοποιηθεί και ως λογισμικό βελτίωσης εικόνων, αφού τείνει να εξάγει φυσικά αποτελέσματα.

Μελλοντικές κατευθύνσεις εξέλιξης της προτεινόμενης μεθόδου περιλαμβάνουν κυρίως την εισαγωγή αυτόματου παράγοντα διόρθωσης χρώματος για την επίτευξη χρωματικής σταθερότητας σε εικόνες που έχουν ληφθεί κάτω από έγχρωμο φωτισμό.



- 164 Συμπεράσματα
- 152 Συγκρίσεις με υπάρχουσες μεθόδους
- 143 Προτεινόμενη μέθοδος
- 139 Νευροφυσιολογικά στοιχεία του ΑΟΣ
- 135 Ανασκόπηση υπάρχουσας βιβλιογραφίας
- 134 Διατύπωση του προβλήματος

4

Δυαδικοποίηση Εγγράφων

Το κεφάλαιο αυτό αναφέρεται στο πρόβλημα της δυαδικοποίησης εγγράφων υποβαθμισμένης ποιότητας, τα οποία προέκυψαν ως αποτέλεσμα της κακής συντήρησης, της εκτεταμένης ή κακής φωτοτύπησης, του ελλιπούς ή ανομοιόμορφου φωτισμού κατά την ψηφιοποίηση και γενικότερα εγγράφων τα οποία παρουσιάζουν έντονη εναλλαγή μεταξύ φωτεινών και σκοτεινών περιοχών. Γίνεται μια σύντομη περιγραφή των σημαντικότερων μεθόδων που έχουν αναπτυχθεί μέχρι τώρα για την επίλυση αυτού του προβλήματος και παρουσιάζονται οι λόγοι για τους οποίους οι λύσεις αυτές δεν είναι πάντα επαρκείς. Στη συνέχεια παραθέτονται νευροφυσιολογικά στοιχεία κυττάρων του ΑΟΣ, τα οποία σκιαγραφούν τον τρόπο με τον οποίο το ΑΟΣ παράγει τις αντιλήψεις του «φωτεινού» και «σκοτεινού». Ακολουθεί η περιγραφή της προτεινόμενης μεθόδου, η οποία υιοθετεί τα παραπάνω στοιχεία, καθώς και σειρά συγκριτικών πειραμάτων. Τέλος, παρουσιάζονται συμπεράσματα που αναλύουν τη συνεισφορά της προτεινόμενης μεθόδου.

4.1 Διατύπωση του προβλήματος

Η αυτόματη επεξεργασία κειμένου είναι μια αλληλουχία πολλών διακριτών βημάτων, η οποία περιλαμβάνει την τμηματοποίηση των εγγράφων σε κείμενο και εικόνες, τη δυαδικοποίηση των περιοχών κειμένου, την οπτική αναγνώριση χαρακτήρων (Optical Character Recognition – OCR) και σε επόμενα στάδια, την εξαγωγή συμπερασμάτων για το περιεχόμενο του κειμένου. Η δυαδικοποίηση του κειμένου, όντας στα πρώτα στάδια αυτής της αλληλουχίας, είναι από τα πιο βασικά βήματα, αφού οποιοδήποτε σφάλμα στο επίπεδο αυτό θα διαδοθεί και στα επόμενα επίπεδα επεξεργασίας.

Η ποιότητα των εικόνων που απεικονίζουν έγγραφα, μπορεί να έχει μεγάλες διακυμάνσεις και κατά συνέπεια, να επηρεάσει σημαντικά την απόδοση των αλγορίθμων δυαδικοποίησης. Εικόνες, που έχουν προέλθει από σάρωση κειμένων εκτυπωμένων από σημερινούς εκτυπωτές, έχουν πολύ καλή ποιότητα και οδηγούν συνήθως σε πολύ ικανοποιητικά αποτελέσματα δυαδικοποίησης από τις υπάρχουσες μεθόδους. Εντούτοις, η περίπτωση αυτή δεν είναι πάντα ο κανόνας. Στο άλλο άκρο βρίσκονται τα ιστορικά κείμενα, τα οποία λόγω φθοράς και κακής συντήρησης, καθιστούν πολύ δύσκολη την ανάλυσή τους. Το ίδιο χαμηλή ποιότητα μπορεί να έχουν και σημερινά κείμενα, τα οποία όμως έχουν υποστεί εκτεταμένη φωτοτύπηση, κάτι που συνήθως οδηγεί στην παραμόρφωση των χαρακτήρων και στην εμφάνιση σκιάσεων. Τέλος, μια καινούρια κατηγορία εικόνων που απεικονίζουν έγγραφα, είναι εικόνες που έχουν ληφθεί με ψηφιακές φωτογραφικές μηχανές κάτω από μη ιδανικές συνθήκες φωτισμού. Στις περιπτώσεις αυτές, οι εικόνες κειμένου που προκύπτουν μπορεί να έχουν εκτεταμένα προβλήματα σκιάσεων ή γενικότερα χαμηλής αντίθεσης, αφού ο προσπίπτων φωτισμός δε μπορεί να είναι το ίδιο ελεγχόμενος όπως σε ένα σαρωτή. Οι τρεις αυτές περιπτώσεις εγγράφων παρουσιάζονται στην Εικόνα 4.1.

0 ΕΦΗΜΕΡΙΣ ΤΗΣ ΚΥΒΕΡΝΗΣΕΩΣ ΤΗΣ ΕΛΛΗΝΙΚΗΣ ΔΗΜΟΚΡΑΤΙΑΣ Έγγραφο μετά από Έγγραφο από ψηφιακή κάμερα Ιστορικό Έγγραφο εκτεταμένη φωτοτύπηση και ανομοιόμορφο φωτισμό

Εικόνα 4.1

Τρεις περιπτώσεις εγγράφων που παρουσιάζουν δυσκολίες στη δυαδικοποίηση.

Είναι λοιπόν αναγκαίο να υπάρξει ένας αλγόριθμος, ο οποίος θα δίνει ικανοποιητικά αποτελέσματα και στις τρεις κατηγορίες των παραπάνω εγγράφων και γενικότερα, θα μπορεί να ανταπεξέλθει σε έγγραφα με μεγάλη υποβάθμιση ποιότητας. Πηγή έμπνευσης

στην κατεύθυνση αυτή μπορεί να αποτελέσει το ΑΟΣ, το οποίο αναμφισβήτητα διαθέτει είναι το καλύτερο σύστημα δυαδικοποίησης, αφού ο ανθρώπινος παρατηρητής μπορεί να διακρίνει τους χαρακτήρες από το φόντο κάτω από πολύ δύσκολες συνθήκες.

4.2 Ανασκόπηση υπάρχουσας βιβλιογραφίας

Οι υπάρχουσες μέθοδοι δυαδικοποίησης εγγράφων χωρίζονται σε δυο μεγάλες κατηγορίες: τις καθολικές (global) και τις τοπικές (local).

4.2.1 Καθολικές μέθοδοι δυαδικοποίησης

Οι καθολικές μέθοδοι αποτελούν τις πρώτες χρονολογικά λύσεις που προτάθηκαν στο πρόβλημα της δυαδικοποίησης εγγράφων. Στόχος τους είναι να ορίσουν ένα καθολικό κατώφλι, αναλύοντας τα στατιστικά χαρακτηριστικά της εικόνας. Το κατώφλι αυτό χωρίζει τις τιμές έντασης της εικόνας σε δύο κλάσεις: το φόντο και τους χαρακτήρες. Η κατηγορία αυτή αλγορίθμων παρουσιάζει ικανοποιητικά αποτελέσματα σε εικόνες με ομοιόμορφο φόντο. Όπως θα δειχθεί όμως και παρακάτω, όταν η εικόνα έχει ανομοιόμορφο φόντο, τότε η ποιότητα των αποτελεσμάτων τους μειώνεται σημαντικά.

Μέθοδος Otsu Η μέθοδος (Otsu, 1979) είναι ίσως η γνωστότερη μέθοδος δυαδικοποίησης, τόσο στον τομέα των εγγράφων όσο και στην επεξεργασία εικόνας γενικότερα. Παρόλο που προτάθηκε πριν πολύ καιρό, εξακολουθεί να χρησιμοποιείται συχνά σε πολλές εφαρμογές. Κύρια δύναμή της είναι η απλότητα και τα καλά αποτελέσματα, υπό συγκεκριμένες προϋποθέσεις, που εξάγει. Στον τομέα της δυαδικοποίησης εγγράφων έχει καθιερωθεί σαν «κλασική» μέθοδος, ενώ δεν υπάρχει εργασία στο συγκεκριμένο τομέα που να μη συγκρίνει τα αποτελέσματά της με αυτή. Η μέθοδος (Otsu, 1979) εφαρμόζεται στο ιστόγραμμα της εικόνας και βασίζεται στην απλή ιδέα ότι τα στοιχεία των δύο κλάσεων που θα προκύψουν, θα πρέπει να είναι όσο το δυνατόν πιο όμοια μεταξύ τους. Κατά συνέπεια, η διακύμανση των στοιχείων κάθε κλάσης θα πρέπει να είναι η ελάχιστη. Έτσι, στόχος της μεθόδου (Otsu, 1979) είναι η εύρεση εκείνου του κατωφλίου που ελαχιστοποιεί τη σταθμισμένη διακύμανση των δύο κλάσεων (weighted within-class variance). Αυτό περιγράφεται στις ακόλουθες εξισώσεις:

$$th = K$$
, $K \in [1, N]$: $\sigma_w^2(K) = \max[\sigma_w^2(k)]_{k=1}^N$ (4.1)

$$\sigma_w^2(k) = q_1(k)\sigma_1^2(k) + q_2(k)\sigma_2^2(k)$$
(4.2)

$$q_1(k) = \sum_{i=1}^k H(i) \qquad q_2(k) = \sum_{i=k+1}^N H(i) \qquad (4.3)$$

$$\mu_1(k) = \sum_{i=1}^k \frac{iH(i)}{q_1(k)} \qquad \qquad \mu_2(k) = \sum_{i=k+1}^N \frac{iH(i)}{q_2(k)} \tag{4.4}$$

$$\sigma_1^2(k) = \sum_{i=1}^k [i - \mu_1(k)]^2 \frac{H(i)}{q_1(k)} \qquad \qquad \sigma_2^2(k) = \sum_{i=k+1}^N [i - \mu_2(k)]^2 \frac{H(i)}{q_2(k)} \qquad (4.5)$$

όπου th το τελικό κατώφλι και $\sigma_w^2(k)$ η σταθμισμένη διακύμανση των δύο κλάσεων, η οποία ελαχιστοποιείται για μια τιμή Κ της κλίμακας εντάσεων της εικόνας στο διάστημα [1,N]. Στις περισσότερες εικόνες το N παίρνει την τιμή 256. $q_1(k)$ και $q_2(k)$ είναι οι πιθανότητες των δύο κλάσεων που χρησιμοποιούνται για τον υπολογισμό της σταθμισμένης διακύμανσης και $\mu_1(k)$ και $\mu_2(k)$ ο μέσος όρος των δύο κλάσεων, αντίστοιχα, για μια τιμή έντασης k στο διάστημα [1,N]. Τέλος, H(i) είναι η συνάρτηση του ιστογράμματος της εικόνας.

Αν και μέχρι σήμερα έχουν παρουσιαστεί πιο βελτιωμένες καθολικές μέθοδοι δυαδικοποίησης, όλες διέπονται από το πρόβλημα του ενός καθολικού κατωφλίου. Η των Εικόνα 4.2 δείχνει σε μεγέθυνση την τρίτη φωτογραφία της Εικόνας 4.1, στην οποία όπως φαίνεται, είναι αδύνατο να υπάρξει ένα κατώφλι που να τμηματοποιεί σωστά και τις δύο περιοχές.

Η αδυναμία καθολικών μεθόδων

Σκοτεινή περιοχή	Φωτεινή περιοχή
ο κατί ^φ όντο 48 του Ν. 3316/2005 διαδικασίας του ά	ακολουθεί την ανοική με την η ^{χαρακτήρες} 5, με ανάλογη εφαρική ρθρου 7 του ίδιου νάμι
δίους πόρους (25	%), με το επι∂έξιμο και: τα την αμοιβή, χωρί¢™

Εικόνα 4.2

Περίπτωση εικόνας που είναι αδύνατο να τμηματοποιηθεί από ένα και μόνο κατώφλι. Οι χαρακτήρες στη φωτεινή περιοχή είναι πιο ανοιχτόχρωμοι (91) από το φόντο της σκοτεινής περιοχής (48).

Όπως είναι φανερό από την Εικόνα 4.2 για εικόνες με μεταβαλλόμενο φόντο, λόγω σκίασης ή ελλιπούς φωτισμού, είναι αδύνατο να υπάρξει ένα και μόνο κατώφλι που να τμηματοποιεί τις δύο περιοχές του εγγράφου. Αν για παράδειγμα επιλέξουμε το κατώφλι βάσει της φωτεινής περιοχής, τότε στη συγκεκριμένη περίπτωση θα πρέπει να είναι μια τιμή έντασης μεγαλύτερη του 91, αφού αυτή την τιμή έχουν οι χαρακτήρες της περιοχής. Με την τιμή αυτή όμως, ολόκληρη η σκοτεινή περιοχή θα τμηματοποιηθεί ως χαρακτήρες. Αναλόγως, αν καθορίσουμε το κατώφλι βάσει της σκοτεινής περιοχής, τότε αυτό θα πρέπει να έχει μια τιμή μεγαλύτερη του 8 και μικρότερη του 48. Στην περίπτωση αυτή όμως, ολόκληρη η φωτεινή περιοχή θα τμηματοποιηθεί ως φόντο. Είναι φανερό λοιπόν, ότι όλες οι μέθοδοι που υπολογίζουν ένα μόνο κατώφλι δε μπορούν να δώσουν ικανοποιητικά αποτελέσματα.

4.2.2 Τοπικές μέθοδοι δυαδικοποίησης

Οι τοπικές μέθοδοι δυαδικοποίησης καθορίζουν διαφορετικό κατώφλι για κάθε εικονοστοιχείο βάσει των τοπικών στατιστικών χαρακτηριστικών της εικόνας. Οι κυριότερες μέθοδοι της κατηγορίας αυτής είναι οι ακόλουθες.

4.2.2.1 Μέθοδος Niblack

Η μέθοδος (Niblack, 1986) είναι μια από τις πρώτες μεθόδους της κατηγορίας αυτής. Αποτελεί για την κατηγορία των τοπικών μεθόδων, ότι και η μέθοδος (Otsu, 1979) για τις καθολικές μεθόδους δυαδικοποίησης, αφού έχει καθιερωθεί πλέον ως κλασική στην κατηγορία και χρησιμοποιείται σαν μέτρο σύγκρισης σε όλες σχεδόν τις προτεινόμενες μεθόδους. Η μέθοδος (Niblack, 1986) υπολογίζει διαφορετικό κατώφλι για κάθε εικονοστοιχείο, μετατοπίζοντας ένα παραλληλόγραμμο παράθυρο επάνω στην εικόνα. Το κατώφλι *T* για το κεντρικό εικονοστοιχείο του παραθύρου υπολογίζεται χρησιμοποιώντας το μέσο όρο *m* και τη διακύμανση *s* των τιμών έντασης του παραθύρου ως εξής:

$$T = m + ks \tag{4.6}$$

όπου k είναι μια σταθερά η οποία έχει την τιμή -0.2. Η σταθερά k χρησιμοποιείται για να καθορίσει το ποσοστό συμμετοχής της διακύμανσης του παραθύρου στο τελικό κατώφλι για το εκάστοτε εικονοστοιχείο. Η μέθοδος αυτή μπορεί να διακρίνει τους χαρακτήρες από το φόντο, σε περιοχές που βρίσκονται κοντά στους χαρακτήρες. Τα αποτελέσματα δεν επηρεάζονται πολύ από το μέγεθος του παραθύρου, εφόσον το παράθυρο καλύπτει το λιγότερο 1-2 χαρακτήρες. Εντούτοις, ο θόρυβος που υπάρχει στο φόντο παραμένει κυρίαρχος και στο τελικό αποτέλεσμα. Κατά συνέπεια, εάν οι χαρακτήρες είναι διάσπαρτοι στην εικόνα, αρκετός θόρυβος θα υπάρξει στο αποτέλεσμα.

4.2.2.2 Μέθοδος Sauvola

Οι Sauvola και Pietikainen (Sauvola, et al., 2000) πρότειναν μια παραλλαγή της μεθόδου (Niblack, 1986), η οποία λύνει το πρόβλημα του θορύβου, προσθέτοντας την ακόλουθη υπόθεση για τις τιμές έντασης του κειμένου και του φόντου: οι χαρακτήρες έχουν τιμές έντασης πιο κοντά στο 0 και το φόντο έχει τιμές έντασης πιο κοντά στο 255. Αυτό οδηγεί στην ακόλουθη σχέση:

$$T = m + \left[1 - k\left(1 - \frac{s}{R}\right)\right] \tag{4.7}$$

όπου οι μεταβλητές *m* και *s* είναι ίδιες με αυτές της μεθόδου (Niblack, 1986), *R* είναι το δυναμικό εύρος της τυπικής απόκλισης που είναι σταθερό στο 128, ενώ η παράμετρος k παίρνει θετικές τιμές (συνήθως 0.5). Η προσέγγιση αυτή δίνει βελτιωμένα αποτελέσματα για εικόνες εγγράφων σε σχέση με τη μέθοδο (Niblack, 1986).

4.2.2.3 Μέθοδος προσαρμοσμένης λογικής (Adaptive Logical Method – ALM)

Οι Yang και Yan πρότειναν μια μέθοδο προσαρμοσμένης λογικής (Yang, et al., 2000), η οποία χρησιμοποιεί το ιστόγραμμα συνδεδεμένου μήκους (run-length histogram) για να εξάγει συμπεράσματα για τα χαρακτηριστικά του πάχους της γραμματοσειράς (character stroke) σε κάθε περιοχή της εικόνας. Αυτά τα συμπεράσματα χρησιμοποιούνται για τη ρύθμιση του μεγέθους της γειτονιάς στην οποία γίνεται δυαδικοποίηση και τον υπολογισμό του κατωφλίου για τη συγκεκριμένη περιοχή.

4.2.2.4 Μέθοδος Gatos

Μια πρόσφατη μέθοδος η οποία χρησιμοποιεί πολλά στάδια επεξεργασίας και δίνει πολύ ικανοποιητικά αποτελέσματα παρουσιάστηκε στην εργασία (Gatos, et al., 2006). Στο

πρώτο στάδιο επεξεργασίας της μεθόδου χρησιμοποιείται ένα φίλτρο Weiner, το οποίο θολώνει την εικόνα χωρίς όμως να επηρεάζει σημαντικά τις ακμές. Έτσι, μειώνεται ο θόρυβος της εικόνας χωρίς όμως να αλλοιώνονται οι χαρακτήρες του κειμένου. Στη συνέχεια, χρησιμοποιείται η μέθοδος (Sauvola, et al., 2000) για μια πρώτη τμηματοποίηση του εγγράφου. Η τμηματοποίηση αυτή, η οποία δεν είναι η τελική, οδηγεί σε μια προσεγγιστική εκτίμηση των εικονοστοιχείων που αποτελούν τους χαρακτήρες. Έχοντας αυτή την εκτίμηση, υπολογίζεται στη συνέχεια μια εκτίμηση του φόντου της εικόνας. Τα εικονοστοιχεία που αποτελούν χαρακτήρες, και για τα οποία δεν υπάρχει πληροφορία σχετική με το φόντο, υπολογίζονται με γραμμική παρεμβολή από τα γειτονικά εικονοστοιχεία φόντου. Έχοντας λοιπόν μια εκτίμηση για τους χαρακτήρες και το φόντο, το επόμενο στάδιο υπολογίζει το τοπικό κατώφλι για κάθε εικονοστοιχείο, λαμβάνοντας υπόψη ότι σε σημεία όπου το φόντο είναι σκούρο, το κατώφλι πρέπει να έχει πιο χαμηλές τιμές. Τέλος, γίνεται μια ανάλυση της συνδετικότητας του κάθε εικονοστοιχείου με τα γειτονικά του, ώστε να ελαχιστοποιηθούν τα εικονοστοιχεία του φόντου που λανθασμένα έχουν τμηματοποιηθεί σαν χαρακτήρες, και αντίστοιχα τα εικονοστοιχεία χαρακτήρων που έχουν τμηματοποιηθεί σαν φόντο. Τα αποτελέσματα της μεθόδου αυτής, όντας και τα πιο πρόσφατα, υπερτερούν έναντι των προηγούμενων μεθόδων.

4.2.3 Αδυναμίες των υπαρχόντων μεθόδων

Αν και ομολογουμένως η απόδοση των αλγορίθμων δυαδικοποίησης εγγράφων έχει βελτιωθεί σημαντικά σε σχέση με τις αρχικές προσπάθειες στο πεδίο αυτό, ακόμα υπάρχουν περιθώρια βελτίωσης, ειδικότερα για εικόνες με σημαντική υποβάθμιση ποιότητας λόγω θορύβου και έντονων σκιάσεων. Αυτό φαίνεται στην Εικόνα 4.3, η οποία δείχνει μια δύσκολη περίπτωση εγγράφου.



Όπως αναφέρθηκε, οι καθολικές μέθοδοι δε μπορούν να δώσουν σωστά αποτελέσματα σε μη ομοιόμορφο φόντο. Η μέθοδος (Niblack, 1986) επηρεάζεται ξεκάθαρα από το θόρυβο της εικόνας. Η μέθοδος (Sauvola, et al., 2000) δίνει βελτιωμένα αποτελέσματα σε σχέση με αυτή του (Niblack, 1986), όμως ο θόρυβος εξακολουθεί να είναι εμφανής, ειδικότερα στις σκοτεινές περιοχές της εικόνας. Η μέθοδος ALM επηρεάζεται λιγότερο από το θόρυβο, όμως οι χαρακτήρες που εξάγει δεν είναι υψηλής ποιότητας. Τέλος, η μέθοδος του Γάτου δίνει χαρακτήρες καλής ποιότητας, χωρίς θόρυβο, όμως δεν τμηματοποιεί σωστά τις περιοχές στις οποίες υπάρχει πολύ έντονη σκίαση. Είναι φανερό λοιπόν, ότι υπάρχει δυνατότητα για βελτίωση των αποτελεσμάτων που εξάγουν οι ήδη υπάρχουσες μέθοδοι, με στόχο την καλύτερη δυαδικοποίηση εγγράφων με μειωμένη ποιότητα.

4.3 Ανίχνευση κειμένου από το ΑΟΣ

Αναμφισβήτητα, το ΑΟΣ υπερέχει μακράν όλων των μεθόδων δυαδικοποίησης που έχουν παρουσιαστεί μέχρι σήμερα. Ο άνθρωπος μπορεί να ξεχωρίσει το κείμενο από το φόντο, όσο ανομοιόμορφο και να είναι αυτό, σε συνθήκες ισχυρής σκίασης, θορύβου και λεκέδων. Η βασική αιτία της ευρωστίας που παρουσιάζει το ΑΟΣ στον τομέα αυτό, είναι τα διαφορετικά νευρικά κυκλώματα αντίληψης της αίσθησης του σκοτεινού και του φωτεινού που διαθέτει.

Τα μηχανικά συστήματα όρασης και οι αλγόριθμοι επεξεργασίας εικόνας δε διαθέτουν κάποιον ποιοτικό διαχωρισμό μεταξύ των αισθήσεων του φωτεινού και του σκοτεινού. Όλες οι επεξεργασίες και οι υπολογισμοί γίνονται απευθείας στις τιμές έντασης που συλλαμβάνουν οι αισθητήρες της κάμερας. Έτσι, μια συγκεκριμένη τιμή, συνήθως στο διάστημα [0,255], χαρακτηρίζει την ένταση του εκάστοτε εικονοστοιχείου. Τιμές κοντά στο 0 θεωρούνται ότι περιγράφουν σκοτεινά ερεθίσματα, ενώ τιμές κοντά στο 255 φωτεινά.

Σε αντίθεση με τα τεχνητά συστήματα όρασης, το ΑΟΣ δεν εφαρμόζει τους κύριους αλγορίθμους επεξεργασίας του πάνω στις τιμές έντασης που συλλαμβάνουν οι φωτοϋποδοχείς του. Οι τιμές έντασης κωδικοποιούνται από τα πρώτα κιόλας στάδια σε σήματα διαφορικής μορφής, στα οποία ενεργούν όλα τα μετέπειτα στάδια επεξεργασίας. Επιπλέον, νευροφυσιολογικές μελέτες έχουν αποδείξει ότι οι αντιλήψεις της φωτεινότητας και της σκοτεινότητας διαφέρουν ποιοτικά και δεν είναι δύο διαφορετικές τιμές σε μια μοναδική κλίμακα αντιληπτής έντασης (Fiorentini, et al., 1990), (Fiorentini, 2004). Πιο συγκεκριμένα, έχει βρεθεί ότι η αντίληψη της φωτεινότητας και της σκοτεινότητας υλοποιείται από δύο κυτταρικούς πληθυσμούς με ανταγωνιστικές αποκρίσεις: τα γαγγλιακά κύτταρα φωτεινού και σκοτεινού κέντρου (ON-center and OFF-center ganglion cells) αντίστοιχα (Fiorentini, et al., 1990), (Fiorentini, 2004), (Nelson, et al., 2004). Τα γαγγλιακά κύτταρα είναι η μόνη έξοδος από τον αμφιβληστροειδή και οι νευράξονές τους σχηματίζουν το οπτικό νεύρο. Τα γαγγλιακά κύτταρα φωτεινού και σκοτεινού κέντρου δειγματοληπτούν ανεξάρτητα το μωσαϊκό των φωτοϋποδοχέων και οι πληθυσμοί τους σχηματίζουν δύο ανεξάρτητα μωσαϊκά. Η Εικόνα 4.4 δείχνει το διαφορικό τρόπο δειγματοληψίας των φωτοϋποδοχέων και το σχηματισμό του υποδεκτικού πεδίου με ανταγωνισμό κέντρου-περιφέρειας των γαγγλιακών κυττάρων. Η απόκριση των γαγγλιακών κυττάρων είναι ουσιαστικά η διαφορά μεταξύ της δειγματοληψίας ενός κεντρικού συνόλου φωτοϋποδοχέων και ενός περιφερειακού συνόλου που περιβάλει το κεντρικό. Έτσι, δεν καταγράφεται η απόλυτη τιμή φωτεινής έντασης που προσλαμβάνουν οι φωτοϋποδοχείς, αλλά η διαφορά του μέσου όρου των τιμών έντασης του κέντρου από την ένταση της περιφέρειας.

Αντίληψη φωτεινού σκοτεινού

Απόκριση των γαγγλιακών κυττάρων



Εικόνα 4.4

Η διαφορική δειγματοληψία του μωσαϊκού των φωτοϋποδοχέων από τα γαγγλιακά κύτταρα φωτεινού και σκοτεινού κέντρου, και ο σχηματισμός του υποδεκτικού τους πεδίου.

Το υποδεκτικό πεδίο των γαγγλιακών κυττάρων μοντελοποιείται συνήθως από τη διαφορά Γκαουσιανών (Difference of Gaussians – DoG) (Rodieck, 1965), (Fiorentini, et al., 1990), (Nelson, et al., 2004), ή από τη Λαπλασιανή Γκαουσιανών (Laplacian of Gaussians) (Marr, 1982). Είναι επίσης γνωστή και ως Μεξικάνικο Καπέλο (Mexican Hat) λόγω του χαρακτηριστικού σχήματός της. Από τα δύο μοντέλα, αυτό που έχει επικρατήσει περισσότερο είναι η διαφορά Γκαουσιανών (Fiorentini, 2004). Η Εικόνα 4.5 δείχνει τη διαφορά Γκαουσιανών για κύτταρα φωτεινού και σκοτεινού κέντρου, ενώ οι εξισώσεις (4.8) και (4.9) εκφράζουν τη μαθηματική μορφή της.

Μοντελοποίηση των γαγγλιακών κυττάρων



Εικόνα 4.5

Το υποδεκτικό πεδίο των γαγγλιακών κυττάρων που περιγράφεται από τη διαφορά Γκαουσιανών.

$$DoG_{ON}(x,y) = G_1(x,y) - G_2(x,y) = k_1 e^{-\left(\frac{x^2 + y^2}{\sigma_1^2}\right)} - k_2 e^{-\left(\frac{x^2 + y^2}{\sigma_2^2}\right)}$$
(4.8)

$$DoG_{OFF}(x,y) = G_2(x,y) - G_1(x,y) = k_2 e^{-\left(\frac{x^2 + y^2}{\sigma_2^2}\right)} - k_1 e^{-\left(\frac{x^2 + y^2}{\sigma_1^2}\right)}$$
(4.9)

όπου G_1 και G_2 είναι η Γκαουσιανή του κέντρου και της περιφέρειας, και αντίστοιχα, σ_1 και σ_2 η τυπική τους απόκλιση. Ισχύει επίσης ότι $\sigma_2 > \sigma_1$, αφού εξ' ορισμού η κατανομή της περιφέρειας έχει μεγαλύτερη ακτίνα. Οι σταθερές k_1 και k_2 επιλέγονται έτσι ώστε να μηδενίζουν το άθροισμα των θετικών και αρνητικών τιμών των κατανομών κέντρου και περιφέρειας, όπως φαίνεται στην ακόλουθη εξίσωση.

$$\iint DoG(x,y) \ dxdy = 0 \tag{4.10}$$

Με τον τρόπο αυτό εξαλείφεται η DC συνιστώσα όταν η διαφορά Γκαουσιανών συνελίσσεται με μια εικόνα *I*.

$$out(i,j) = \iint I(i,j) \cdot DoG(i-y,j-x) \, dx \, dy \tag{4.11}$$

Αυτό σημαίνει ότι τα γαγγλιακά κύτταρα δεν παράγουν απόκριση, όταν το κέντρο και η περιφέρειά τους δέχονται το ίδιο ερέθισμα ($G_1 = G_2$). Αντίθετα, αντιδρούν μόνο όταν υπάρχει μεταβολή της έντασης που καταγράφουν οι φωτοϋποδοχείς μέσα στο υποδεκτικό τους πεδίο. Η συμπεριφορά αυτή ισοδυναμεί ουσιαστικά με την εξαγωγή ακμών, αφού σε ομοιόμορφες περιοχές η απόκριση είναι μηδέν, ενώ στις περιοχές που υπάρχει αλλαγή της έντασης του φωτός, υπάρχει απόκριση.

Γαγγλιακά κύτταρα σκοτεινού κέντρου Τα γαγγλιακά κύτταρα σκοτεινού κέντρου αυξάνουν την απόκρισή τους όταν η περιφέρεια του υποδεκτικού τους πεδίου (G_2) δεχθεί ένα φωτεινό ερέθισμα, ενώ μειώνουν την απόκρισή τους όταν στο κέντρο του υποδεκτικού τους πεδίου (G_1) αυξηθεί η ένταση του φωτός. Ως αποτέλεσμα, τα γαγγλιακά κύτταρα σκοτεινού κέντρου αποκρίνονται στην τοπική μείωση της έντασης του φωτός (light decrements) και είναι ιδανικά για την ανίχνευση σκούρων σημείων σε φωτεινό φόντο (Dacey, 1993), (Croner, et al., 1995), (Fiorentini, 2004). Αντίθετα, παρουσιάζουν τη μικρότερη απόκρισή τους για φωτεινά σημεία σε σκούρο φόντο (Εικόνα 4.6i).

Γαγγλιακά κύτταρα φωτεινού κέντρου

Τα γαγγλιακά κύτταρα φωτεινού κέντρου αυξάνουν την απόκρισή τους όταν το κέντρο του υποδεκτικού τους πεδίου (*G*₁) δεχθεί ένα φωτεινό ερέθισμα, ενώ μειώνουν την απόκρισή τους όταν στην περιφέρεια του υποδεκτικού τους πεδίου (*G*₂) αυξηθεί η ένταση του φωτός. Ως αποτέλεσμα, τα γαγγλιακά κύτταρα φωτεινού κέντρου αποκρίνονται στην τοπική αύξηση της έντασης του φωτός (light increments) και είναι ιδανικά για την ανίχνευση φωτεινών σημείων σε σκούρο φόντο (Dacey, 1993), (Croner, et al., 1995), (Fiorentini, 2004). Αντίθετα, παρουσιάζουν τη μικρότερη απόκρισή τους για σκοτεινά σημεία σε φωτεινό φόντο (Εικόνα 4.6ii).

Η Εικόνα 4.7 παρουσιάζει την απόκριση των γαγγλιακών κυττάρων για τη μετάβαση από μια περιοχή με χαμηλή ένταση φωτός σε μια περιοχή με μεγάλη ένταση φωτός.



Γαγγλιακά κύτταρα σκοτεινού κέντρου

Γαγγλιακά κύτταρα φωτεινού κέντρου

Συνήθως το κείμενο είναι σκούρο και βρίσκεται πάνω σε ανοιχτόχρωμο φόντο. Αυτό σημαίνει ότι το κείμενο, σαν οπτικό σήμα, διεγείρει τα γαγγλιακά κύτταρα σκοτεινού κέντρου, με αποτέλεσμα το ΑΟΣ να βασίζεται σε αυτά για την ανίχνευση των χαρακτήρων. Στην αντίθετη περίπτωση όπου έχουμε ανοιχτόχρωμο κείμενο σε σκούρο φόντο, τότε τα γαγγλιακά κύτταρα φωτεινού κέντρου είναι αυτά που θα ανιχνεύουν τους χαρακτήρες. Αυτό απεικονίζεται στην Εικόνα 4.8.

Το κείμενο σαν οπτικό σήμα

Έχει βρεθεί, επίσης, ότι και τα δύο είδη γαγγλιακών κυττάρων χωρίζονται σε δύο υποκατηγορίες, βάσει του μεγέθους του υποδεκτικού τους πεδίου: στα M (από το magni, δηλαδή μεγάλα) και στα P (από το parvi, δηλαδή μικρά). Όπως φανερώνει και το όνομά τους, τα κύτταρα M διαθέτουν μεγαλύτερο υποδεκτικό πεδίο, σε σχέση με τα P, σε κάθε σημείο του αμφιβληστροειδούς (Croner, et al., 1995). Αυτό σημαίνει ότι το ΑΟΣ χρησιμοποιεί το λιγότερο δυο διαφορετικά μεγέθη γαγγλιακών κυττάρων για την ανίχνευση των οπτικών σημάτων. Αυτή τη σχέση των δύο μεγεθών υιοθετεί και η προτεινόμενη μέθοδος στα τεχνητά γαγγλιακά κύτταρα που χρησιμοποιεί για τη δυαδικοποίηση των εγγράφων.

Μεγέθη γαγγλιακών κυττάρων



4.4 Περιγραφή της προτεινόμενης μεθόδου

Όπως αναφέρθηκε στην ενότητα 4.2, δύο είναι οι κατηγορίες αλγορίθμων για τη δυαδικοποίηση εγγράφων: οι καθολικές και οι τοπικές μέθοδοι. Η προτεινόμενη μέθοδος ανήκει και στις δύο. Βασικό χαρακτηριστικό της είναι ότι χρησιμοποιώντας παρόμοια γαγγλιακά κύτταρα σκοτεινού κέντρου με το ΑΟΣ, μετασχηματίζει την αρχική εικόνα κειμένου σε μια νέα μορφή, η οποία μπορεί να τμηματοποιηθεί από καθολικές μεθόδους. Η καθολική μέθοδος που δυαδικοποιεί τη μετασχηματισμένη μορφή της αρχικής εικόνας είναι η μέθοδος (Otsu, 1979). Το μπλοκ διάγραμμα της μεθόδου παρουσιάζεται στην Εικόνα 4.9.

Εικόνα 4.9

Το μπλοκ διάγραμμα της προτεινόμενης μεθόδου.



4.4.1 Προεπεξεργασία

Το πρώτο στάδιο της προτεινόμενης μεθόδου είναι η προεπεξεργασία, στην οποία η τιμή *O_{ij}* του κάθε εικονοστοιχείου της αρχικής εικόνας *O*, που είναι στις αποχρώσεις του γκρι, εκτείνεται γραμμικά σε ολόκληρο το εύρος της κλίμακας [0,255], όπως δείχνει η εξίσωση (4.12).

$$O_{ij}' = \frac{O_{ij} - O_{min}}{O_{max} - O_{min}} \times 255$$
(4.12)

όπου O_{min} είναι η ελάχιστη τιμή της αρχικής εικόνας O, O_{max} είναι η μέγιστη τιμή της και O_{ij}^{\prime} η τιμή εξόδου του εκάστοτε εικονοστοιχείου για το στάδιο αυτό. Ο γραμμικός αυτός μετασχηματισμός εξασφαλίζει ότι η αρχική εικόνα εκτείνεται σε όλο το εύρος της κλίμακας των γκρι αποχρώσεων, χωρίς να χαθεί κάποιο τμήμα της οπτικής πληροφορίας του εγγράφου. Σε πολλές περιπτώσεις, το στάδιο αυτό δεν είναι απαραίτητο, αφού στα περισσότερα έγγραφα οι τιμές των χαρακτήρων καταλαμβάνουν τις τιμές κοντά στο 0 και οι τιμές του φόντου τις τιμές κοντά στο 255. Εντούτοις, μπορεί να είναι ιδιαίτερα χρήσιμο σε ορισμένα ιστορικά ή κατεστραμμένα έγγραφα, στα οποία η συνολική αντίθεση είναι μικρή. Επιπλέον, δεν επηρεάζει εικόνες με ήδη καλή αντίθεση, εισάγει ελάχιστο υπολογιστικό κόστος και βελτιώνει τη στιβαρότητα της προτεινόμενης μεθόδου. Για το στάδιο αυτό της προεπεξεργασίας, εξετάσθηκε επίσης η χρήση της τεχνικής εξισορρόπησης ιστογράμματος. Διαπιστώθηκε όμως, ότι πολλές φορές παρουσίαζε μη φυσικές εικόνες, καθώς ενίσχυε λεπτομέρειες όπως η υφή του χαρτιού, οι οποίες δυσκόλευαν τη δυαδικοποίηση και κατέστρεφε ορισμένες χρήσιμες οπτικές πληροφορίες. Η Εικόνα 4.10 δείχνει μια σύγκριση της έκτασης ιστογράμματος, που χρησιμοποιείται στο στάδιο της προεπεξεργασίας και της εξισορρόπησης ιστογράμματος για δύο διαφορετικά έγγραφα: ένα με χαμηλή αντίθεση (δεν καταλαμβάνει όλες τις θέσεις της κλίμακας [0,255]) και ένα με καλή αντίθεση (οι τιμές έντασης εκτείνονται σε όλη την κλίμακα [0,255]).



Φαίνεται ξεκάθαρα ότι η έκταση ιστογράμματος αποτελεί πιο αξιόπιστη λύση για τη συγκεκριμένη εφαρμογή, αφού βελτιώνει τη συνολική αντίθεση της εικόνας όταν χρειάζεται, ενώ δεν επηρεάζει τις εικόνες που δε χρειάζονται βελτίωση. Αντίθετα η εξισορρόπηση ιστογράμματος επηρεάζει σημαντικά όλες τις εικόνες, αναδεικνύοντας λεπτομέρειες που δυσκολεύουν τη διαδικασία της δυαδικοποίησης.

4.4.2 Τεχνητά γαγγλιακά κύτταρα σκοτεινού κέντρου

Μετά την προεπεξεργασία, η προτεινόμενη μέθοδος χρησιμοποιεί μονάδες επεξεργασίας παρόμοιες με τα γαγγλιακά κύτταρα σκοτεινού κέντρου του ΑΟΣ, τα οποία, όπως είδαμε στην ενότητα 4.3, ανιχνεύουν σκούρο κείμενο σε ανοιχτόχρωμο φόντο. Το στάδιο αυτό είναι η καρδιά της προτεινόμενης μεθόδου, αφού μετασχηματίζει όλες τις εικόνες κειμένου σε μια μορφή ικανή να τμηματοποιηθεί από τη μέθοδο (Otsu, 1979), εξαλείφοντας σκιάσεις, σκοτεινά σημεία, λεκέδες και γενικότερα το ανομοιόμορφο φόντο.

4.4.2.1 Υποδεκτικά πεδία

Όπως και το ΑΟΣ χρησιμοποιεί δύο βασικά μεγέθη γαγγλιακών κυττάρων (τα Μ και Ρ), έτσι και η προτεινόμενη μέθοδος χρησιμοποιεί δυο μεγέθη τεχνητών γαγγλιακών κυττάρων για την επεξεργασία κάθε εικονοστοιχείου της εικόνας *Ο*', όπως φαίνεται στην Εικόνα 4.11.



Τα Μ κύτταρα αποκρίνονται γενικότερα σε ερεθίσματα μεγάλου μεγέθους, ενώ τα Ρ κύτταρα σε ερεθίσματα μικρού μεγέθους. Τα γαγγλιακά αυτά κύτταρα σκοτεινού κέντρου έχουν τετράγωνα ομόκεντρα υποδεκτικά πεδία, τα οποία περιγράφονται από τις ακόλουθες εξισώσεις.

$$Surround_{ij}^{K} = \frac{1}{(2S^{K} + 1)^{2}} \sum_{y=i-S^{K}}^{i+S^{K}} \sum_{x=j-S^{K}}^{j+S^{K}} O_{yx}^{\prime}$$
(4.13)

$$Center_{ij}^{K} = \frac{1}{(2C^{K}+1)^{2}} \sum_{y=i-C^{K}}^{i+C^{K}} \sum_{x=j-C^{K}}^{j+C^{K}} O'_{yx}$$
(4.14)

$$K = \{M, P\} \qquad M: Magni, \qquad P: Parvi$$

$$S^{K} > C^{K} \forall K, \quad S^{M} > S^{P} \qquad (4.15)$$

όπου (i, j) οι συντεταγμένες του προς επεξεργασία εικονοστοιχείου της εικόνας O', K το μέγεθος του γαγγλιακού κυττάρου, ενώ S^K και C^K τα μεγέθη της περιφέρειας και του κέντρου για κύτταρο K μεγέθους, αντίστοιχα. Όπως είναι φανερό από τις εξισώσεις (4.13) και (4.14), οι τιμές *Surround^K_{ij}* και *Center^K_{ij}* είναι ο μέσος όρος έντασης της εικόνας O' στη θέση (i, j) για τις τετράγωνες περιοχές της περιφέρειας και του κέντρου, στο μέγεθος K. Αυτό είναι διαφορετικό από τις κυκλικές Γκαουσιανές κατανομές των βιολογικών γαγγλιακών κυττάρων. Εντούτοις, βρέθηκε ότι δεν υπάρχουν ορατές διαφορές στο τελικό αποτέλεσμα σε σχέση με τη χρήση κυκλικών κατανομών. Για το λόγο αυτό, οι τετράγωνες περιοχές προτιμήθηκαν έναντι των κυκλικών, επειδή απαιτούν μικρότερο χρόνο εκτέλεσης. Η εξίσωση (4.15) περιγράφει τις χωρικές σχέσεις που πρέπει να ισχύουν για τα μεγέθη των κυττάρων της προτεινόμενης μεθόδου. Στη συνέχεια, η περιφέρεια και το κέντρου και περιφέρειας. Οι τελικές αυτές τιμές θα χρησιμοποιηθούν για την εξαγωγή της τελικής απόκρισης στη θέση (i, j).

$$Surround_{ij} = \frac{WS^{M} \cdot Surround_{ij}^{M} + WS^{P} \cdot Surround_{ij}^{P}}{WS^{M} + WS^{P}}$$
(4.16)

$$Center_{ij} = \frac{WC^{M} \cdot Center_{ij}^{M} + WS^{P} \cdot Center_{ij}^{P}}{WC^{M} + WC^{P}}$$
(4.17)

όπου WS^M , WS^P τα βάρη των περιφερειών και WC^M , WS^P τα βάρη των κέντρων για τα δύο μεγέθη κυττάρων. *Surround*_{ij} και *Center*_{ij} είναι οι τελικές τιμές της περιφέρειας και του κέντρου. Όλες οι τιμές των βαρών βρίσκονται στο διάστημα [0, ∞).

4.4.2.2 Συνάρτηση απόκρισης

Για το σχεδιασμό της συνάρτησης απόκρισης του κάθε γαγγλιακού κυττάρου, χρησιμοποιείται σαν βάση η ακόλουθη εξίσωση, η οποία έχει χρησιμοποιηθεί και στο κεφάλαιο 3.

$$G(x) = \frac{(B+A) \cdot x}{A+x} \quad \forall x \in [0,B], A, B \in \mathbb{R}^+$$
(4.18)

Η ανάλυση των ιδιοτήτων της περιγράφεται εκτενέστατα στην ενότητα 3.4.1 και για το λόγο αυτό, θα αναφέρουμε εδώ μόνο τις βασικές της ιδιότητες, χωρίς να επεκταθούμε περισσότερο. Η συνάρτηση (4.18) είναι μια μεταβλητή συνάρτηση απεικόνισης μεταξύ των δεδομένων εισόδου (x) και εξόδου (G(x)). Η τιμή B εκφράζει το εύρος των τιμών για τα δεδομένα εισόδου. Στην περίπτωσή μας είναι B=255. Το σχετικό μέγεθος της τιμής A ως προς το B, καθορίζει τη μη γραμμικότητα με την οποία απεικονίζονται τα δεδομένα εισόδου στην έξοδο. Μικρές τιμές του A δίνουν ισχυρή μη γραμμική απεικόνιση μεταξύ εισόδου και εξόδου, ενώ μεγάλες τιμές συσχετίζουν γραμμικά την είσοδο με την έξοδο.

Στην παρούσα μορφή της, η εξίσωση (4.18) έχει καθολική συμπεριφορά: εφαρμόζει τον ίδιο μετασχηματισμό σε όλα τα εικονοστοιχεία της εικόνας, περίπου όπως η διόρθωση γ. Ωστόσο, μπορεί να αποκτήσει τοπική συμπεριφορά, εάν γίνουν οι παρακάτω μετατροπές:

- i. Να αντικατασταθεί ο παράγοντας Α που ελέγχει τη μη γραμμικότητα από την περιφέρεια Surround_{ij}, η οποία αποτελεί μια μέτρηση του μέσου όρου έντασης στη γειτονιά (i, j). Έτσι, θα εφαρμόζεται διαφορετική μη γραμμικότητα σε κάθε σημείο της εικόνας, ανάλογα με τη γειτονιά.
- ii. Να αντικατασταθεί η τιμή εισόδου x από την τοπική διαφορά SC_{ij} της περιφέρειας $Surround_{ij}$ από το $Center_{ij}$.

Οι παραπάνω μετατροπές περιγράφονται από τις ακόλουθες εξισώσεις.

$$SC_{ij} = Surround_{ij} - Center_{ij}$$
 (4.19)

$$G_{ij} = \begin{cases} \frac{(255 + Surround_{ij}) \cdot SC_{ij}}{Surround_{ij} + SC_{ij}} & \forall SC_{ij} > 0\\ 0 & \forall SC_{ij} \le 0 \end{cases}$$
(4.20)

Η εξίσωση (4.19) αποτελεί προσέγγιση της διαφοράς Γκαουσιανών για τα γαγγλιακά κύτταρα σκοτεινού κέντρου, που παρουσιάστηκε στην εξίσωση (4.9). Η Εικόνα 4.12 παρουσιάζει την τρισδιάστατη απεικόνιση της εξίσωσης (4.20), η οποία είναι και η βασικότερη εξίσωση της μεθόδου, αφού αυτή απεικονίζει την εικόνα σε μια μορφή στην οποία σκιάσεις και λεκέδες εξαλείφονται.



Ερμηνεία της Η εξίσωση (4.20) δέχεται σαν είσοδο την τοπική διαφορά SC_{ij} και σαν ρυθμιστικό εξίσωσης (4.20)
 (4.20) παράγοντα την περιφέρεια Surround_{ij} κάθε εικονοστοιχείου της εικόνας O', ώστε να παράγει μια απόκριση G_{ij}. Βασικός στόχος είναι η απόκριση αυτή να είναι μέγιστη για τα εικονοστοιχεία των χαρακτήρων και μηδενική για τα εικονοστοιχεία του φόντου.

148

Όταν το κέντρο του τεχνητού γαγγλιακού κυττάρου έχει μεγαλύτερη τιμή από την περιφέρειά του (*Center_{ij}* > *Surround_{ij}*, *SC_{ij}* <0), τότε το κεντρικό εικονοστοιχείο (*i*, *j*) *E* είναι πιο φωτεινό από τη γειτονιά του, οπότε η έξοδος του κυττάρου είναι μηδέν (*G_{ij}*=0). φ Ομοίως, τα γαγγλιακά κύτταρα σκοτεινού κέντρου του ΑΟΣ, δε διεγείρονται από φωτεινά σημεία σε σκούρο φόντο (Εικόνα 4.6).

Αντίθετα, όταν το κέντρο του τεχνητού γαγγλιακού κυττάρου έχει μικρότερη τιμή από την περιφέρειά του (*Surround_{ij}>Center_{ij}*, SC_{ij} >0), τότε το κεντρικό εικονοστοιχείο (*i*, *j*) είναι πιο σκούρο από τη γειτονιά του. Στην περίπτωση αυτή, η απόκριση καθορίζεται από τη μη γραμμικότητα της συνάρτησης (4.20), η οποία εξαρτάται από τις τιμές SC_{ij} και *Surround_{ij}*. Η διαφορά SC_{ij} είναι ένας δείκτης της τοπικής αντίθεσης. Όταν λαμβάνει μικρές θετικές τιμές, τότε πρόκειται για περιοχή χαρακτήρων στην οποία υπάρχει μικρή τοπική αντίθεση. Αντίθετα, όταν λαμβάνει μεγάλες τιμές, τότε τοπικής είναι του τοπικού φωτισμού. Όταν λαμβάνει μικρές τιμές τιμές τιμές, τότε πρόκειται για σκιά ή για κάποιο λεκέ του εγγράφου. Όταν αποκτάει μεγάλες τιμές, τότε πρόκειται για περιοχή με σωστό φωτισμό. Οι σχέσεις αυτές παρουσιάζονται πιο ξεκάθαρα στην Εικόνα 4.13.



Εικονοστοιχεία φόντου

Εικονοστοιχεία χαρακτήρων



Οι δύο μεταβλητές εισόδου SC_{ij} και Surround_{ij} της εξίσωσης (4.20), καθώς και φυσική τους σημασία ανάλογα με τις τιμές τους.

Ανάλογα με τους πιθανούς συνδυασμούς των τιμών SC_{ij} και $Surround_{ij}$ στην παραπάνω εικόνα, καθορίζεται και η μη γραμμικότητα της απεικόνισης. Έτσι, για μικρή τοπική αντίθεση (SC_{ij} ψ) και σκοτεινή περιοχή ($Surround_{ij}$ ψ) έχουμε τη μορφή a* (Εικόνα 4.12), η οποία δίνει μεγάλες αποκρίσεις, αυξάνοντας ουσιαστικά την ευαισθησία του τεχνητού γαγγλιακού κυττάρου για τις σκοτεινές περιοχές. Στην περίπτωση μικρής τοπική αντίθεσης (SC_{ij} ψ) και φωτεινής περιοχής ($Surround_{ij}$ ↑) έχουμε τη μορφή b* (Εικόνα 4.12), η οποία δεν πραγματοποιεί κάποια ενίσχυση στην απόκριση. Καθώς η τοπική αντίθεση αυξάνει (SC_{ij} ↑), τόσο μικρότερη γίνεται και η μη γραμμικότητα. Έτσι, σταδιακά περνάει από τη μορφή c* και για μεγάλες τιμές της SC_{ij} γίνεται τελείως γραμμική (d*).

4.4.2.3 Επιλογή παραμέτρων

Τα μεγέθη των κυττάρων M και P, καθώς και τα βάρη με τα οποία συμμετέχουν στην τελική τιμή του κέντρου *Center_{ij}* και της περιφέρειας *Surround_{ij}*, επηρεάζουν το τελικό αποτέλεσμα της προτεινόμενης μεθόδου. Η Εικόνα 4.14i παρουσιάζει την απόκριση ενός γαγγλιακού κυττάρου σκοτεινού κέντρου για διαφορετικά μεγέθη κέντρου και περιφέρειας, το οποίο έχει σαν συνάρτηση απόκρισης την εξίσωση (4.20). Το κύτταρο αυτό εφαρμόζεται σε όλες τις θέσεις μιας συνθετικής εικόνας που περιέχει διάφορα μεγέθη χαρακτήρων (από 8 ως 32 pt) και διάφορες πυκνότητες θορύβου.



Εικόνα 4.14

i. Η απόκριση ενός γαγγλιακού κυττάρου σκοτεινού κέντρου για διαφορετικά μεγέθη κέντρου (C) και περιφέρειας (S).

ii. Τα αποτελέσματα του συνδυασμού δύο γαγγλιακών κυττάρων Ρ και Μ, για διάφορα βάρη στις εξισώσεις (4.16) και (4.17).

Σύμφωνα με τα αποτελέσματα της Εικόνας 4.14i, είναι φανερό ότι το μέγεθος του κέντρου C καθορίζει την ευκρίνεια του τελικού αποτελέσματος. Όταν το κέντρο αποτελείται από ένα εικονοστοιχείο (C=0), τότε παρουσιάζει τη μεγαλύτερη ευκρίνεια, είναι ευαίσθητο στις μικρές λεπτομέρειες και επηρεάζεται σημαντικά από το θόρυβο. Όταν το κέντρο αποτελείται από ατοτελείται από μια γειτονιά 3×3 εικονοστοιχείων (C=1), τότε παρουσιάζει μεγαλύτερη ανοχή στο θόρυβο, αφού ουσιαστικά είναι ένα φίλτρο μέσης τιμής, ενώ δεν επηρεάζει κατά κάποιο εμφανή τρόπο τις μικρές γραμματοσειρές (8pt, 10pt) σε ανάλυση 300dpi. Κέντρα μεγαλύτερου μεγέθους, όπως το 5×5 (C=2) ή το 7×7 (C=3), εξαλείφουν με επιτυχία το θόρυβο, αλλά επηρεάζουν σημαντικά την ευκρίνεια των μικρών γραμματοσειρών, αφού πραγματοποιούν την πράξη του μέσου όρου σε μεγαλύτερες γειτονιές. Είναι επίσης φανερό από την Εικόνα 4.14i, ότι το μέγεθος της περιφέρειας S καθορίζει το «γέμισμα»

των χαρακτήρων. Κύτταρα με μικρό μέγεθος περιφέρειας αποκρίνονται μόνο στις μεταβάσεις από φωτεινές σε σκοτεινές περιοχές, εξάγοντας ουσιαστικά τις ακμές. Όσο μεγαλύτερο είναι το μέγεθος της περιφέρειας τόσο περισσότερο η απόκριση εκτείνεται στο εσωτερικό του χαρακτήρα. Αυτό σημαίνει ότι το μέγεθος της περιφέρειας Σ πρέπει να είναι το λιγότερο, ίδιο με το πάχος των γραμμών του μεγαλύτερου σε μέγεθος χαρακτήρα του κειμένου.

Λαμβάνοντας υπόψη τα παραπάνω συμπεράσματα, επιλέχθηκαν τα παρακάτω μεγέθη για τα γαγγλιακά κύτταρα τύπου Μ και Ρ που θα χρησιμοποιηθούν στην προτεινόμενη μέθοδο: (C^P =0, S^P =3) και (C^M =1, S^M =15). Τα μικρά κύτταρα P έχουν περιφέρεια μεγέθους 7×7 εικονοστοιχείων και κέντρο ενός εικονοστοιχείου. Κατά συνέπεια, είναι ιδανικά για τις υψηλές συχνότητες και μικρού ή μεσαίου μεγέθους χαρακτήρες. Αυτό έρχεται και σε συμφωνία με νευροφυσιολογικά δεδομένα που αναφέρουν ότι το κέντρο των Ρ γαγγλιακών κυττάρων του ΑΟΣ αποτελείται από ένα μόνο φωτοϋποδοχέα, ενώ η περιφέρειά τους είναι 4 με 8 φορές μεγαλύτερη από την ακτίνα του κέντρου τους (Dacey, 1993), (Croner, et al., 1995). Τα μεγάλα κύτταρα Μ έχουν περιφέρεια μεγέθους 31×31 εικονοστοιχείων και κέντρο 3×3 εικονοστοιχείων. Επομένως, είναι ιδανικά για σήματα χαμηλών συχνοτήτων και για χαρακτήρες μεσαίου και μεγάλου μεγέθους. Η αναλογία μεγέθους μεταξύ των Μ και των Ρ κυττάρων της μεθόδου είναι παρόμοια με αυτή των βιολογικών γαγγλιακών κυττάρων του ΑΟΣ, σύμφωνα με την οποία, τα Μ βιολογικά κύτταρα έχουν ακτίνα 2-3 φορές μεγαλύτερη από τα P (Croner, et al., 1995). Το επιλεγμένο μέγεθος της περιφέρειας των κυττάρων Μ της προτεινόμενης μεθόδου, είναι κατάλληλο για όλα τα μεγέθη χαρακτήρων μέχρι πάχους γραμμών 30 εικονοστοιχείων.

Για την ανάλυση της επίδρασης που έχουν τα βάρη συμμετοχής των κυττάρων Μ και Ρ στα αποτελέσματα της προτεινόμενης μεθόδου, έγινε η διερεύνηση που παρουσιάζεται στην Εικόνα 4.14ii. Στην εικόνα αυτή, κύτταρα Μ και Ρ με τα μεγέθη που επιλέχθηκαν στην προηγούμενη παράγραφο, συνδυάστηκαν με διάφορα βάρη, ώστε να βρεθεί ο βέλτιστος συνδυασμός τους. Όπως ήταν αναμενόμενο, το βάρος WC^P του κέντρου P είναι άμεσα συσχετισμένο με την εξαγωγή λεπτομερειών και θορύβου. Μειώνοντας το βάρος συμμετοχής του, μειώνεται ο θόρυβος και θολώνουν οι μικρές γραμματοσειρές. Από τα αποτελέσματα της Εικόνας 4.14ii, είναι φανερό ότι μια ικανοποιητική αναλογία βαρών, η οποία επιτυγχάνει ισορροπία μεταξύ λεπτομερειών στην έξοδο και μείωσης του θορύβου, είναι ακόλουθη: (WC^P =20, WS^P =33) και (WC^M =80, WS^M =66). Η αναλογία αυτή των βαρών είναι αυτή που επιλέχθηκε για την προτεινόμενη μέθοδο. Ένα άλλο σημαντικό στοιχείο, που δείχνει η Εικόνα 4.14ii, είναι ότι μικρές αλλαγές στις τιμές των βαρών δε συνεπάγονται μεγάλες αλλαγές στο τελικό αποτέλεσμα. Αυτό που είναι σημαντικό είναι ο λόγος των βαρών, ο οποίος για να αλλάξει σημαντικά πρέπει τα βάρη να μεταβληθούν πολύ. Πρακτικά αυτό σημαίνει ότι αν για παράδειγμα η τιμή του WC^M μεταβληθεί από 80 σε 90, η διαφορά στο τελικό αποτέλεσμα θα είναι ανεπαίσθητη, αφού ο λόγος $WC^{M}/(WC^{M} + WC^{P})$ που πριν ήταν 0.8, τώρα θα γίνει 0.818. Όλα τα αποτελέσματα που θα παρουσιαστούν στην ενότητα των πειραματικών αποτελεσμάτων, παράγονται χρησιμοποιώντας τα μεγέθη των κυττάρων που παρουσιάστηκαν στην προηγούμενη παράγραφο και τα βάρη συμμετοχής που παρουσιάστηκαν στην παράγραφο αυτή, χωρίς κάποια επιπλέον ρύθμιση. Μόνο σε ειδικές περιπτώσεις εγγράφων με πολύ μικρές

Επιλογή μεγέθους για τα Μ και Ρ κύτταρα.

Επιλογή των βαρών συμμετοχής. γραμματοσειρές, μικρότερες των 8 pt στα 300 dpi, στις οποίες απαιτείται πολύ μεγάλη ευκρίνεια, είναι απαραίτητο να αυξηθεί ο λόγος $WC^P/(WC^M + WC^P)$.

4.4.3 Δυαδικοποίηση Otsu

Η απόκριση των γαγγλιακών κυττάρων σκοτεινού κέντρου της προτεινόμενης μεθόδου, έχει μετασχηματίσει την αρχική εικόνα σε μια μορφή G στην οποία έχουν εξαλειφθεί οι σκιάσεις και οι λεκέδες. Στη μορφή αυτή, τιμές κοντά στο 255 αντιστοιχούν σε εικονοστοιχεία χαρακτήρων, ενώ τιμές κοντά στο 0 αντιστοιχούν σε εικονοστοιχεία του φόντου. Η μέθοδος (Otsu, 1979) δυαδικοποιεί την εικόνα G στις δύο κλάσεις, με την επιλογή ενός καθολικού κατωφλίου T_0 . Η έξοδος B αυτού του σταδίου δίνεται από την ακόλουθη εξίσωση.

$$B_{ij} = \begin{cases} 0 & \forall \ G_{ij} \ge T_O \\ 255 & \forall \ G_{ij} < T_O \end{cases}$$
(4.21)

4.4.4 Διορθωτική επεξεργασία

Τα προηγούμενα στάδια δίνουν ικανοποιητικά αποτελέσματα δυαδικοποίησης κειμένου. Εντούτοις, εφαρμόζεται μια επιπλέον διορθωτική επεξεργασία στη δυαδική εικόνα B, ώστε να ελαχιστοποιηθούν μεμονωμένα εικονοστοιχεία φόντου, που λανθασμένα έχουν τμηματοποιηθεί σαν χαρακτήρες. Όπως και με το στάδιο της προεπεξεργασίας, η συμβολή αυτού του σταδίου δεν είναι ζωτικής σημασίας, όμως αυξάνει την ευρωστία της προτεινόμενης μεθόδου, χωρίς να εισάγει μεγάλο υπολογιστικό κόστος. Τα εικονοστοιχεία που έχουν τμηματοποιηθεί ως χαρακτήρες και δε συνδέονται στην 3×3 γειτονιά τους με κάποιο άλλο εικονοστοιχείο χαρακτήρα, είναι πολύ πιθανό να προέρχονται από θόρυβο. Έτσι, το στάδιο της διορθωτικής επεξεργασίας ελέγχει τη συνδετικότητα CN_{ij} στη γειτονιά Moore για κάθε εικονοστοιχείο της κλάσης των χαρακτήρων (B_{ij} =0) και το κατατάσσει στην κλάση του φόντου (B_{ij} =255), εάν έχει μηδενική συνδετικότητα. Το τελικό αποτέλεσμα B' δίνεται από τις ακόλουθες εξισώσεις.

$$B'_{ij} = \begin{cases} 0 & \forall \ (B_{ij} = 0) \& \& (CN_{ij} > 0) \\ 255 & \forall \ (B_{ij} = 0) \& \& (CN_{ij} = 0) \\ 255 & \forall \ (B_{ij} = 255) \end{cases}$$
(4.22)

$$CN_{ij} = \left[\sum_{y=i-1}^{i+1} \sum_{x=j-1}^{j+1} \delta(B_{yx})\right] - \delta(B_{ij})$$
(4.23)

όπου δ() είναι η συνάρτηση Dirac. Η εξίσωση (4.23) υπολογίζει τη συνδετικότητα στη γειτονιά Moore του κάθε εικονοστοιχείου, ενώ η εξίσωση (4.22) δίνει το τελικό αποτέλεσμα.

Η Εικόνα 4.15 δείχνει όλα τα ενδιάμεσα στάδια επεξεργασίας της προτεινόμενης μεθόδου για μια εικόνα με μεταβλητό φόντο.



Εικόνα 4.15

Παράδειγμα των ενδιάμεσων σταδίων της προτεινόμενης μεθόδου για μια εικόνα με μεταβλητό φόντο λόγω ανομοιόμορφου φωτισμού.

4.5 Πειραματικά αποτελέσματα - συγκρίσεις

Η προτεινόμενη μέθοδος συγκρίνεται με έξι άλλες μεθόδους, καθιερωμένες στο χώρο της δυαδικοποίησης κειμένου για διάφορους βαθμούς υποβάθμισης της ποιότητας του κειμένου, όπως σκιάσεις, λεκέδες, χαμηλή αντίθεση και θόρυβος. Οι μέθοδοι που χρησιμοποιούνται στη σύγκριση είναι η μέθοδος (Otsu, 1979), η μέθοδος (Niblack, 1986), Η μέθοδος (Sauvola, et al., 2000), η ALM (Yang, et al., 2000), η μέθοδος του Γάτου (Gatos, et al., 2006) και το εμπορικό πακέτο FineReader 8.0 (Abb07). Στα πλαίσια αυτής της ενότητας, πραγματοποιήθηκαν τρία ποσοτικά πειράματα. Ο κύριος στόχος των δύο πρώτων πειραμάτων είναι η ποσοτική σύγκριση του PSNR (Peak Signal to Noise Ratio) στα αποτελέσματα όλων των αλγορίθμων. Η πειραματική διαδικασία αυτής της σύγκρισης μπορεί εύκολα να αναπαραχθεί, και παρουσιάζει τη συμπεριφορά όλων των αλγορίθμων για ένα ευρύ φάσμα υποβάθμισης των εικόνων. Αυτό επιτυγχάνεται με τη χρήση ενός εκτεταμένου συνόλου τεχνητών, και κατά συνέπεια ελεγχόμενων, υποβαθμίσεων της ποιότητας των εγγράφων, σε ένα αρχικό κείμενο που αποτελεί την ιδανική εικόνα (Ground Truth – GT). Γνωρίζοντας την αρχική μορφή της εικόνας πριν την υποβάθμιση, μπορεί εύκολα να γίνει σύγκριση του αποτελέσματος κάθε αλγορίθμου με την ιδανική περίπτωση της αρχικής εικόνας. Παρόμοια προσέγγιση έχει ακολουθηθεί και στην εργασία (Sezgin, et al., 2004), στην οποία πραγματοποιήθηκε μια εκτεταμένη ποσοτική σύγκριση των αποτελεσμάτων διαφόρων αλγορίθμων δυαδικοποίησης. Το τρίτο πείραμα είναι μια σύγκριση των ποσοστών OCR αναγνώρισης στα αποτελέσματα των αλγορίθμων σύγκρισης, όταν αυτά εφαρμοστούν σε έγγραφα υποβαθμισμένης ποιότητας.

4.5.1 Παραγωγή των υποβαθμισμένων εικόνων

Στην αρχή, δημιουργήθηκε μια ιδανική δυαδική εικόνα κειμένου, η οποία χρησιμοποιήθηκε στη σύγκριση με τα αποτελέσματα των αλγορίθμων. Η εικόνα αυτή περιλάμβανε χαρακτήρες μεγέθους από 10pt ως 48pt, τόσο κανονικούς όσο και έντονους (bold), σε ανάλυση 300dpi. Η γραμματοσειρά που επιλέχθηκε ήταν η Times New Roman, για δύο κύριους λόγους: Πρώτον, είναι μια από τις πιο διαδεδομένες γραμματοσειρές. Δεύτερον, ανήκει στην κατηγορία των γραμματοσειρών Sheriffs, οι οποίες έχουν μεταβλητό πάχος γραμμών, σε αντίθεση με άλλες γραμματοσειρές, όπως η Arial, οι οποίες έχουν σταθερό. Το μεταβλητό πάχος γραμμής μπορεί να αποτελέσει πρόβλημα κατά τη διαδικασία της δυαδικοποίησης, ιδιαίτερα στα σημεία όπου το πάχος μειώνεται, στα οποία εμφανίζονται ασυνέχειες του χαρακτήρε. Η αρχική δυαδική εικόνα (*GT*), στην οποία χρησιμοποιήθηκε σα βάση σύγκρισης με τα αποτελέσματα των αλγορίθμων, καθώς και για την παραγωγή των υποβαθμισμένων εικόνων, παρουσιάζεται στην Εικόνα 4.16.

abcdefghijklmnopqrstuvwxyzABCDEFGHIJKLMNOPQRSTUVWXYZ 1234567890-10pt abcdefghijklmnopqrstuvwxyzABCDEFGHIJKLMNOPQRSTUVWXYZ 1234567890-10pt

abcdefghijklmnopqrstuvwxyzABCDEFGHIJKLMNOPQRSTUVWXYZ 1234567890-12pt abcdefghijklmnopqrstuvwxyzABCDEFGHIJKLMNOPQRSTUVWXYZ 1234567890-12pi

Εικόνα 4.16

Η ιδανική δυαδική εικόνα (GT) που χρησιμοποιήθηκε σα βάση για την παραγωγή των υποβαθμισμένων εικόνων και για τη σύγκριση με τα αποτελέσματα των αλγορίθμων.

Παραγωγή της

ιδανικής

δυαδικής

εικόνας.

abcdefghijklmnopqrstuvwxyzABCDEFGHIJKLMNOPQRSTUVWXYZ 1234567890-16pt abcdefghijklmnopqrstuvwxyzABCDEFGHIJKLMNOPQRSTUVWXYZ 1234567890-16pt abcdefghijklmnopqrstuvwxyzABCDEFGHIJKLMNOPQRSTUVWXYZ 123456-20pt abcdefghijklmnopqrstuvwxyzABCDEFGHIJKLMNOPQRSTUVWXYZ 123456-20pt

 $a bc defghijklmnopqrstuvwxyzABCDEFGHIJKLMNOPQRSTUV-24 pt \\ a bc defghijklmnopqrstuvwxyzABCDEFGHIJKLMNOPQRST-24 pt \\$

abcdefghijklmnopqrstuvwxyzABCDEFG-36pt abcdefghijklmnopqrstuvwxyzABCDEF-36pt abcdefijABCD abcdefijABC-48pt

Για την παραγωγή των υποβαθμισμένων εικόνων, ενσωματώθηκαν στην εικόνα *GT* διάφορες πυκνότητες θορύβου, καθώς και τεχνητές σκιάσεις ή τοπικές χαμηλές αντιθέσεις.

Εισαγωγή θορύβου.

Ο θόρυβος που ενσωματώθηκε στην εικόνα *GT* ήταν λευκός Γκαουσιανός θόρυβος, με μηδενικό μέσο όρο. Η εισαγωγή του στην εικόνα πραγματοποιήθηκε με τη συνάρτηση "*imnoise()*" του Matlab. Η παρακάτω εξίσωση περιγράφει την εισαγωγή του θορύβου.

$$GT' = \left(\frac{GT}{255} + \sqrt{var} \times randn(size(GT))\right) \times 255$$
(4.24)

όπου GT' είναι η εικόνα με το θόρυβο και var, η διακύμανση του προστιθέμενου θορύβου. Η συνάρτηση randn(x) παράγει πίνακες τυχαίων αριθμών μεγέθους x, των οποίων τα στοιχεία είναι κανονικά κατανεμημένα, με μέσο όρο μηδέν και διακύμανση ίση

με 1. Τέλος, η συνάρτηση size(x) επιστρέφει το μέγεθος των διαστάσεων του x. Με τη χρήση της εξίσωσης (4.24) ενσωματώθηκαν οκτώ διαφορετικά επίπεδα θορύβου στην εικόνα GT, με αποτέλεσμα να προκύψουν οκτώ διαφορετικές εικόνες GT'. Οι διακυμάνσεις θορύβου που χρησιμοποιήθηκαν ήταν: 0 (καθόλου θόρυβος), 0.02, 0.04, 0.06, 0.08, 0.1, 0.15 και 0.2. Το όριο 0.2 επιλέχθηκε επειδή πέρα από αυτό ο θόρυβος γίνεται τόσο έντονος, που, ακόμα και ο άνθρωπος δε μπορεί να ξεχωρίσει χαρακτήρες μικρότερους από 12pt.

Για να μπορέσουμε να εξομοιώσουμε της επίδραση των σκιάσεων ή των λεκέδων, οι οκτώ εικόνες *GT* ' χρησιμοποιήθηκαν για να δημιουργηθούν υποβαθμισμένες εικόνες με σκίαση. Για κάθε εικόνα *GT* ', δηλαδή για κάθε επίπεδο θορύβου, δημιουργήθηκαν έξι διαφορετικές εικόνες με τεχνητή σκίαση. Η τεχνητή σκίαση ενσωματώθηκε στις εικόνες χρησιμοποιώντας την ακόλουθη εξίσωση.

Εισαγωγή σκίασης.

$$SH_{ij} = \begin{cases} (1 - sh) \times GT'_{ij} & \forall j \ge px/2\\ \left(1 - \frac{j}{px/2} \times sh\right) \times GT'_{ij} & \forall j < px/2 \end{cases}$$
(4.25)

όπου sh είναι ένας αριθμός στο διάστημα [0,1], ο οποίος καθορίζει την ένταση της σκίασης. SH είναι η σκιασμένη εικόνα που προκύπτει, px είναι το πλάτος της εικόνας και (i,j) οι συντεταγμένες του εκάστοτε εικονοστοιχείου. Ο πρώτος κλάδος της εξίσωσης (4.25) για $j \ge px/2$, ουσιαστικά εισάγει την σκίαση στο μισό της εικόνας, ενώ ο άλλος κλάδος, για j < px/2, παράγει την ομαλή μετάβαση προς τη σκιασμένη περιοχή. Τα έξι επίπεδα σκίασης που ενσωματώθηκαν ήταν: 0.5, 0.6, 0.7, 0.8, 0.9 και 0.95. Αυτό είχε σαν αποτέλεσμα τη δημιουργία 48 σκιασμένων εικόνων (8 επίπεδα θορύβου × 6 επίπεδα σκίασης). Το όριο 0.5 επιλέχθηκε, γιατί για μικρότερα επίπεδα σκίασης το οπτικό αποτέλεσμα ήταν στα όρια του αντιληπτού.

Η μικρή τοπική αντίθεση είναι μια άλλη αιτία υποβάθμισης της ποιότητας των εικόνων εγγράφων. Και εδώ ακολουθήθηκε μια παρόμοια προσέγγιση με την εισαγωγή της τεχνητής σκίασης. Για κάθε εικόνα θορύβου *GT'* δημιουργήθηκαν έξι εικόνας χαμηλής αντίθεσης. αντίθεσης, χρησιμοποιώντας την εξίσωση (4.26).

$$H_{ij} = \begin{cases} 255 - \left[(1-h) \times \left(255 - GT'_{ij} \right) \right] & \forall j \ge px/2 \\ 255 - \left[\left(1 - \frac{j}{px/2} \times h \right) \times \left(255 - GT'_{ij} \right) \right] & \forall j < px/2 \end{cases}$$
(4.26)

όπου h είναι ένας αριθμός στο διάστημα [0,1] που καθορίζει το επίπεδο της χαμηλής τοπικής αντίθεσης. H είναι η εικόνα χαμηλής τοπικής αντίθεσης που προκύπτει, px το πλάτος της εικόνας και (i, j) οι συντεταγμένες του εκάστοτε εικονοστοιχείου. Η εξίσωση (4.26) είναι παρόμοια με την εξίσωση (4.25), μόνο που αντί για τη μετατόπιση των τιμών προς το 0, τις μετατοπίζει προς το 255. Όπως και πριν, ο πρώτος κλάδος της εξίσωσης εφαρμόζει τη χαμηλή αντίθεση στο μισό της εικόνας, ενώ ο δεύτερος κλάδος πραγματοποιεί την ομαλή μετάβαση προς την περιοχή αυτή. Ο συνολικός αριθμός εικόνων ελέγχου που χρησιμοποιήθηκαν για την αξιολόγηση των αλγορίθμων ήταν 96 (48 σκιασμένες + 48 με χαμηλή αντίθεση), στις οποίες υπήρχαν διάφορα επίπεδα θορύβου. Η Εικόνα 4.17 δείχνει ένα παράδειγμα εικόνας για τις σκιασμένες εικόνες και τις εικόνες χαμηλής αντίθεσης.



σκίαση=0.8, διακύμανση θορύβου=0.2

αντίθεση=0.8, διακύμανση θορύβου=0.2

4.5.2 Σύγκριση του PSNR

Όλοι οι αλγόριθμοι εφαρμόστηκαν και στις 96 εικόνες ελέγχου και τα αποτελέσματά τους συγκρίθηκαν με την ιδανική δυαδική εικόνα *GT*, χρησιμοποιώντας το PSNR, όπως φαίνεται στις ακόλουθες εξισώσεις.

$$MSE = \frac{1}{py \times px} \sum_{i=1}^{py} \sum_{j=1}^{px} (GT_{ij} - K_{ij})^2$$
(4.27)

$$PSNR = 10 \times \log_{10} \left(\frac{MAX_{GT}^2}{MSE} \right)$$
db (4.28)

όπου *MSE* είναι το μέσο τετραγωνικό σφάλμα (Root Mean Square Error), *K* είναι η εικόνα εξόδου του εκάστοτε αλγορίθμου, *MAX_{GT}* είναι η μέγιστη τιμή της ιδανικής δυαδικής εικόνας *GT* και *py*, *px* οι διαστάσεις της εικόνας.

Πείραμα σε σκιασμένες εικόνες. Η Εικόνα 4.18 παρουσιάζει το αποτέλεσμα *PSNR* κάθε αλγορίθμου για τις 48 σκιασμένες εικόνες. Τα αποτελέσματα παρουσιάζονται σε τρισδιάστατη μορφή, η οποία δείχνει ξεκάθαρα τη συμπεριφορά του κάθε αλγορίθμου σε αυξανόμενα επίπεδα θορύβου και σκίασης. Τα αποτελέσματα της μεθόδου (Otsu, 1979) είναι τα χειρότερα. Αυτό δείχνει ξεκάθαρα την αδυναμία των καθολικών μεθόδων σε τέτοιου είδους υποβαθμίσεις της ποιότητας των εικόνων. Η κατηγορία αυτή των μεθόδων κατηγοριοποιεί λανθασμένα τη σκιασμένα τη



Οι τρεις αλγόριθμοι ALM, (Niblack, 1986) και (Sauvola, et al., 2000) έχουν παρόμοια συμπεριφορά. Οι τελευταίοι δύο παρουσιάζουν τα καλύτερα αποτελέσματα, όσον αφορά εικόνες χωρίς θόρυβο. Εντούτοις, η περίπτωση αυτή απέχει πολύ από την πραγματικότητα, αφού οποιαδήποτε πραγματική εικόνα, αναπόφευκτα θα έχει κάποιο ποσοστό θορύβου.

Επιπλέον, η απόδοσή τους πέφτει απότομα όταν υπάρχουν ακόμη και μικρά επίπεδα θορύβου στην εικόνα. Σε σχέση με τους δύο αλγορίθμους, ο (Sauvola, et al., 2000) παρουσιάζει καλύτερα αποτελέσματα από τον (Niblack, 1986), αφού δείχνει πιο ικανοποιητική συμπεριφορά στο θόρυβο. Η μέθοδος ALM έχει καλύτερη συμπεριφορά από τις άλλες δύο μεθόδους, για διακυμάνσεις θορύβου 0.02, 0.04 και 0.06. Για μεγαλύτερες διακυμάνσεις, η συμπεριφορά της γίνεται όμοια με των άλλων δύο. Αντίθετα, η μέθοδος (Gatos, et al., 2006) παρουσιάζει πιο βελτιωμένη συμπεριφορά στα επίπεδα του θορύβου από ότι οι προηγούμενοι αλγόριθμοι. Πιο συγκεκριμένα, η απόδοσή του μειώνεται πιο αργά καθώς αυξάνεται η διακύμανση του θορύβου, ενώ επιτυγχάνει μεγαλύτερο PSNR για τη μέγιστη διακύμανση (0.2). Το λογισμικό FineReader 8.0 παρουσιάζει καλύτερα αποτελέσματα σε όλα τα επίπεδα θορύβου, για σκιάσεις που βρίσκονται στο διάστημα [0.5, 0.8]. Εντούτοις, για πιο έντονες σκιάσεις (0.9 ή 0.95), η απόδοσή του πέφτει απότομα, στη δεύτερη χειρότερη θέση, μετά από τη μέθοδο (Otsu, 1979). Από την Εικόνα 4.18 είναι φανερό ότι την καλύτερη συμπεριφορά την παρουσιάζει η προτεινόμενη μέθοδος. Η απόδοσή της μειώνεται με το μικρότερο ρυθμό, καθώς αυξάνεται η διακύμανση του θορύβου. Επιπλέον, ο PSNR που επιτυγχάνει η προτεινόμενη μέθοδος είναι μεγαλύτερος από το διπλάσιο του αντίστοιχου της ALM, της (Niblack, 1986) και της (Sauvola, et al., 2000), 70% μεγαλύτερος από τη μέθοδο (Gatos, et al., 2006) και περίπου 30% μεγαλύτερος από του FineReader 8.0, για διακυμάνσεις θορύβου μεταξύ 0.04 και 0.1. Στην πιο υποβαθμισμένη εικόνα ελέγχου, με διακύμανση θορύβου 0.2 και σκίαση 0.95, παρουσιάζει PSNR≈13db, ενώ η μέθοδος (Gatos, et al., 2006) PSNR ≈7db, η (Sauvola, et al., 2000) και η (Niblack, 1986) PSNR ≈6db και το FineReader 8.0 *PSNR*≈2db. Το γενικότερο συμπέρασμα για τη συμπεριφορά όλων των αλγορίθμων είναι ότι τα επίπεδα σκίασης δεν επηρεάζουν σημαντικά την απόδοση των αλγορίθμων (εκτός από τη μέθοδο (Otsu, 1979) και FineReader 8.0), σε αντίθεση με τα επίπεδα του θορύβου που έχουν σημαντική επίπτωση στην απόδοσή τους. Η προτεινόμενη μέθοδος παρουσιάζει την καλύτερη απόδοση, ακολουθούμενη από τον αλγόριθμο (Gatos, et al., 2006) και FineReader 8.0. Οι μέθοδοι (Sauvola, et al., 2000), (Niblack, 1986) και ALM ακολουθούν, ενώ η μέθοδος (Otsu, 1979) αποτυγχάνει τελείως σε τέτοιου είδους εικόνες.

Πείραμα σε εικόνες με μικρή τοπική αντίθεση. Η Εικόνα 4.19 παρουσιάζει το αποτέλεσμα *PSNR* κάθε αλγορίθμου για τις 48 εικόνες με χαμηλή τοπική αντίθεση. Σε αντίθεση με τις εικόνες σκίασης, η μέθοδος (Otsu, 1979) παρουσιάζει εδώ συγκρίσιμα αποτελέσματα με τις υπόλοιπες τεχνικές, δείχνοντας ότι οι καθολικές μέθοδοι μπορούν να ανταπεξέλθουν πιο ικανοποιητικά στη χαμηλή τοπική αντίθεση, από ότι στη σκίαση. Η μέθοδος (Otsu, 1979) παρουσιάζει σχεδόν ίδια αποτελέσματα με αυτά της μεθόδου (Gatos, et al., 2006), με τη δεύτερη να παρουσιάζει ελαφρώς πιο βελτιωμένα. Το λογισμικό FineReader 8.0 έχει συμπεριφορά που γενικότερα επίσης μοιάζει με αυτή της μεθόδου (Gatos, et al., 2006). Η ALM παρουσιάζει τη χειρότερη συμπεριφορά σε σχέση με όλες τις άλλες μεθόδους. Η μέθοδος (Niblack, 1986) παρουσιάζει πάλι την καλύτερη απόδοση, για εικόνες χωρίς θόρυβο, ενώ ακολουθείται στενά από τη μέθοδο (Sauvola, et al., 2000). Όπως και πριν όμως, όταν εμφανιστεί και η παραμικρή διακύμανση θορύβου στην εικόνα, πράγμα που είναι αναπόφευκτο στις εικόνες που προέρχονται από σάρωση ή από φωτοτύπηση, η απόδοσή τους πέφτει

απότομα. Η προτεινόμενη μέθοδος παρουσιάζει την καλύτερη συμπεριφορά σε σχέση με τις άλλες μεθόδου, για τα περισσότερα επίπεδα χαμηλής αντίθεσης και θορύβου.



Συγκεκριμένα, περνάει σε απόδοση τους αλγορίθμους (Niblack, 1986) και (Sauvola, et al., 2000), για όλες τις περιπτώσεις θορύβου. Παρουσιάζει μεγαλύτερο PSNR από τη μέθοδο (Otsu, 1979) στις 42 από τις 48 εικόνες ελέγχου. Δίνει επίσης καλύτερα αποτελέσματα από τη μέθοδο (Gatos, et al., 2006) για όλες τις διακυμάνσεις θορύβου μεγαλύτερες από 0.02, και από το FineReader 8.0 για διακυμάνσεις μεγαλύτερες του 0.04. Όπως και πριν, η απόδοσή του μειώνεται με το μικρότερο ρυθμό, καθώς αυξάνεται η διακύμανση του θορύβου. Το γενικό συμπέρασμα από αυτό το πείραμα για τη συμπεριφορά όλων των αλγορίθμων είναι ότι η χαμηλή τοπική αντίθεση επηρεάζει πολύ περισσότερο τη συμπεριφορά τους σε σχέση με τη σκίαση. Η επίδραση αυτή είναι σχεδόν ίδια με την επίδραση του θορύβου. Η προτεινόμενη μέθοδος παρουσιάζει την καλύτερη απόδοση για την πλειοψηφία των 48 εικόνων ελέγχου αυτού του πειράματος, με μικρότερη όμως διαφορά σε σχέση με το πείραμα σκίασης. Δεύτεροι έρχονται το FineReader 8.0 και η μέθοδος (Gatos, et al., 2006) ακολουθούμενοι από τη μέθοδο (Otsu, 1979). Στη συνέχεια, η μέθοδος (Sauvola, et al., 2000) παρουσιάζει τα αμέσως επόμενα καλύτερα αποτελέσματα, ακολουθούμενη από τη μέθοδο (Niblack, 1986). Τέλος, στην τελευταία θέση στο πείραμα αυτό βρίσκεται η μέθοδος ALM.

4.5.3 Σύγκριση αποτελεσμάτων OCR

Οι ψηφιακές φωτογραφικές μηχανές γίνονται όλο και πιο δημοφιλείς για την ψηφιοποίηση εγγράφων και για την οπτική αναγνώριση χαρακτήρων σε αυτά. Εντούτοις, στην περίπτωση αυτή, οι συνθήκες φωτισμού μπορεί να διαφέρουν σημαντικά στις διάφορες περιοχές της εικόνας. Για το λόγο αυτό, τρία από τα έγγραφα που χρησιμοποιούνται στο πείραμα αυτό (τα έγγραφα 1, 4 και 5), ψηφιοποιήθηκαν από μια ψηφιακή φωτογραφική μηχανή (Panasonic Lumix DMC-FZ5), κάτω από διαφορετικές συνθήκες φωτισμού, οι οποίες δημιούργησαν ισχυρές σκιάσεις. Από τα υπόλοιπα έγγραφα, το δεύτερο έγγραφο έχει ένα λεκέ, ενώ το τρίτο είναι ένα ιστορικό κείμενο. Για την ποσοτικοποίηση των OCR αποτελεσμάτων, χρησιμοποιείται η μετρική που παρουσιάστηκε στην εργασία (Gupta, et al., 2007), και η οποία περιγράφεται από την ακόλουθη εξίσωση.

$$R = \frac{\sum_{n=1}^{N} \min[groundtruth(n), OCR(n)]}{\sum_{n=1}^{N} groundtruth(n)} \times 100\%$$
(4.29)

όπου *R* είναι ο βαθμός αναγνώρισης, $n \in [1, N]$ δείχνει τον αριθμό των μοναδικών λέξεων στο κείμενο, *groundtruth*(*n*) είναι οι φορές που η *n*-οστή μοναδική λέξη εμφανίζεται στην ιδανική δυαδική εικόνα, και *OCR*(*n*) είναι οι φορές που η ίδια λέξη εμφανίζεται στο OCR αποτέλεσμα του εκάστοτε αλγορίθμου. Η προσέγγιση αυτή μετράει μόνο τις λέξεις που έχουν σωστούς όλους τους χαρακτήρες τους. Η μηχανή που χρησιμοποιήθηκε για να εκτελέσει την οπτική αναγνώριση χαρακτήρων, ήταν το εμπορικό λογισμικό ABBY FineReader 8.0 (Abb07). Ο Πίνακας 4.1 δείχνει τους βαθμούς αναγνώρισης όλων των μεθόδων. Η προτεινόμενη μέθοδος παρουσιάζει τα καλύτερα αποτελέσματα σε σχέση με τις υπόλοιπες μεθόδους, ακολουθούμενη από την ALM και τη μέθοδο (Gatos, et al., 2006). Μια ενδιαφέρουσα παρατήρηση είναι ότι η μέθοδος ALM, η οποία δεν κατατάχθηκε ανάμεσα στις πρώτες στα πειράματα PSRN, ξεπέρασε και τη μέθοδο (Gatos, et al., 2006) και βρέθηκε δεύτερη στο πείραμα OCR. Αυτό οφείλεται στο γεγονός ότι η ALM είναι ευαίσθητη στο μέγεθος της γραμματοσειράς. Όταν συγκεκριμένα το μέγεθος της γραμματοσειράς είναι μεγαλύτερο από 24pt στα 300dpi (π.χ. επικεφαλίδες), δυαδικοποιεί σωστά μόνο το περίγραμμα των χαρακτήρων, παρουσιάζοντας μια συμπεριφορά που παραπέμπει σε εξαγωγή ακμών. Αυτό είναι και ο λόγος για τον οποίο δεν παρουσιάζει καλά αποτελέσματα στο πείραμα PSNR, στο οποίο η ιδανική δυαδική εικόνα *GT* είχε μεγέθη μέχρι και 48pt. Αντίθετα, τα έγγραφα των κειμένων του Πίνακα 4.1 περιέχουν μόνο μικρά μεγέθη χαρακτήρων, και έτσι η ALM παρουσιάζει βελτιωμένα αποτελέσματα. Ολοκληρώνοντας, η προτεινόμενη μέθοδος παρουσίασε σταθερά υψηλής ποιότητας αποτελέσματα και στα πέντε έγγραφα, σε αντίθεση με άλλες μεθόδους, οι οποίες παρουσίασαν καλά αποτελέσματα σε ορισμένα, ενώ χειρότερα σε άλλα.

	1	2	3	4	5			
Μέθοδοι		 YEAR CONTRACT OF THE ADDRESS OF THE AD	And the second s			Σύνολο	Πίνακας 4.1 Αποτελέσματα του	
	Χαρ.=3354	Χαρ.=858	Χαρ.=469	Χαρ.=3245	Χαρ.=2346	Χαρ.=10272	πειράματος	
	Λέξεις=574	Λέξεις =184	Λέξεις =107	Λέξεις =493	Λέξεις =417	Λέξεις =1775	ΟCR για	
	% βαθμός αναγνώρισης							
Otsu	23%	90%	46%	15%	36%	42%	φωτογραφική	
Niblack	67%	77%	83%	60%	94%	76.2%	μηχανή.	
Sauvola	94%	87%	88%	78%	95%	88.4%		
ALM	95%	93%	92%	73%	94%	89.4%		
Gatos	88%	91%	91%	71%	96%	87.4%		
FR8.0	31%	92%	68%	48%	95%	66.8%		
Προ/μενη	96%	95%	95%	79%	96%	92.2%		

4.5.4 Ποιοτική σύγκριση αποτελεσμάτων

Εκτός από τα ποσοτικά πειράματα, τα αποτελέσματα των αλγορίθμων ελέγχθηκαν και για πολλές πραγματικές εικόνες με διάφορα είδη προβλημάτων. Η Εικόνα 4.20 παρουσιάζει τα αποτελέσματά τους για μια εικόνα με λεκέ. Τα αποτελέσματα είναι σε συμφωνία με τα συμπεράσματα της προηγούμενης ενότητας. Η μέθοδος (Otsu, 1979) αποτυγχάνει να δυαδικοποιήσει επιτυχημένα το έγγραφο, και τμηματοποιεί ολόκληρη την περιοχή του λεκέ σαν χαρακτήρα. Οι μέθοδοι (Niblack, 1986) και (Sauvola, et al., 2000) επηρεάζονται πάρα πολύ από το θόρυβο. Η ALM επηρεάζεται λιγότερο από το θόρυβο σε σχέση με τις προηγούμενες δύο, όμως εξάγει αρκετό για να μειώσει σημαντικά τα αποτελέσματα στην οπτική αναγνώριση των χαρακτήρων. Οι αλγόριθμοι με την καλύτερη απόδοση είναι το FineReader 8.0, η τεχνική (Gatos, et al., 2006) και η προτεινόμενη μέθοδος. Ανάμεσά τους, η προτεινόμενη μέθοδος παρουσιάζει ελαφρώς βελτιωμένα αποτελέσματα, αφού η
μέθοδος (Gatos, et al., 2006) επηρεάζεται περισσότερο από το θόρυβο, ενώ το FineReader 8.0 εξάγει χαμηλότερης ποιότητας χαρακτήρες στην περιοχή του λεκέ.



Η ίδια περίπου συμπεριφορά φαίνεται και στην Εικόνα 4.21, στην οποία παρουσιάζονται τα αποτελέσματα των αλγορίθμων για ένα άλλο ιστορικό έγγραφο. Όπως είναι αναμενόμενο, η μέθοδος (Otsu, 1979) επηρεάζεται σημαντικά από το μεταβλητό φόντο. Οι μέθοδοι (Niblack, 1986) και (Sauvola, et al., 2000) εξάγουν αρκετό θόρυβο στα αποτελέσματά τους, ενώ η μέθοδος (Gatos, et al., 2006) και το FineReader 8.0 δυαδικοποιούν λανθασμένα ορισμένες περιοχές. Η ALM δεν επηρεάζεται από το μεταβλητό φόντο, όμως οι χαρακτήρες που εξάγει δεν είναι καλής ποιότητας. Η προτεινόμενη μέθοδος παρουσιάζει τρία βασικά χαρακτηριστικά ταυτόχρονα: τη σωστή

δυαδικοποίηση του μεταβλητού φόντου, την εξαγωγή ελάχιστου θορύβου και την καλή ποιότητα των χαρακτήρων.



Στην Εικόνα 4.22 παρουσιάζονται τα αποτελέσματα όλων των μεθόδων για ένα πραγματικό έγγραφο (απόδειξη τράπεζας) με χαμηλή αντίθεση, λόγω της έλλειψης μελανιού στον κρουστικό εκτυπωτή που το εκτύπωσε.



Αποτελέσματα των αλγορίθμων για ένα ιστορικό έγγραφο με μεταβλητό

Εικόνα 4.22

Αποτελέσματα των αλγορίθμων για πραγματικό έγγραφο με χαμηλή αντίθεση.

Όπως είναι φανερό από την προηγούμενη εικόνα, η προτεινόμενη μέθοδος εξάγει το καλύτερο αποτέλεσμα σε σχέση με όλες τις υπόλοιπες μεθόδους. Πιο συγκεκριμένα, οι μέθοδοι ALM και Gatos, αδυνατούν να δυαδικοποιήσουν σωστά τους εκτυπωμένους χαρακτήρες. Η μέθοδος Otsu τμηματοποιεί ένα μέρος του φόντου ως χαρακτήρες, με αποτέλεσμα την υποβάθμιση του τελικού αποτελέσματος. Το λογισμικό FineReader 8.0 δυαδικοποιεί τους μισούς περίπου εκτυπωμένους χαρακτήρες. Οι μέθοδοι Sauvola και Niblack εξάγουν όλους τους εκτυπωμένους χαρακτήρες. Εντούτοις, ο πρώτος παράγει χαρακτήρες υποβαθμισμένης ποιότητας, ενώ ο δεύτερος εξάγει εμφανέστατο θόρυβο. Μόνο η προτεινόμενη μέθοδο συνδυάζει καλής ποιότητας χαρακτήρες και έλλειψη θορύβου. Αυτό είναι φανερό και από τη μεγέθυνση στην περιοχή της υπογραφής.

4.5.5 Συνολική συμπεριφορά των μεθόδων

Αυτή η ενότητα συνοψίζει τα συμπεράσματα των τριών προηγούμενων ενοτήτων, ώστε να τονίσει τα πλεονεκτήματα και τα μειονεκτήματα κάθε μεθόδου. Αυτό είναι σημαντικό γιατί καθορίζει τις εφαρμογές στις οποίες μπορεί να χρησιμοποιηθεί κάθε τεχνική.

- Η μέθοδος (Otsu, 1979), όντας καθολική μέθοδος, είναι ακατάλληλη για κείμενα με μη Otsu ομοιόμορφο φόντο, όπως σκιάσεις και λεκέδες. Εντούτοις, έχει καλή απόδοση στο θόρυβο και στη χαμηλή αντίθεση. Αυτό υποδεικνύει ότι μπορεί να χρησιμοποιηθεί με επιτυχία στη δυαδικοποίηση εγγράφων καλής ποιότητας.
- Niblack &
- Οι μέθοδοι (Niblack, 1986) και (Sauvola, et al., 2000) έχουν αρκετά παρόμοια συμπεριφορά, η οποία προέρχεται από την ίδια βασική ιδέα στην οποία βασίζονται, με τη Sauvola δεύτερη να εξάγει πιο βελτιωμένα αποτελέσματα από την πρώτη. Παρουσιάζουν ιδανική συμπεριφορά για μη ομοιόμορφό φόντο, γεγονός που τις καθιστά αποτελεσματικές για κείμενα με σκιάσεις και λεκέδες. Παρόλα αυτά όμως, ακόμη και το παραμικρό ποσοστό θορύβου στην αρχική εικόνα, θα εμφανιστεί στο τελικό αποτέλεσμα, μειώνοντας την απόδοση των αλγορίθμων οπτικής αναγνώρισης χαρακτήρων. Ως εκ τούτου, είναι σημαντικό να εφαρμόζονται μόνο σε έγγραφα με πολύ μικρό ποσοστό θορύβου, στα οποία καλό θα ήταν να εφαρμοστεί πριν τη δυαδικοποίηση κάποια τεχνική απαλοιφής του θορύβου.
 - Η τεχνική ALM παρουσιάζει την πιο ευμετάβλητη συμπεριφορά από όλες τις άλλες ALM μεθόδους, αφού τα αποτελέσματά της εξαρτώνται από το μέγεθος των χαρακτήρων. Γενικότερα, είναι πολύ αποτελεσματική για κείμενα με σκιάσεις και λεκέδες, καθώς και μεγέθη χαρακτήρων ως και 24pt στα 300dpi. Για χαρακτήρες μεγαλύτερου μεγέθους, δυαδικοποιεί σωστά μόνο το περίγραμμά τους. Επιπλέον, δεν παρουσιάζει ικανοποιητικά αποτελέσματα σε έγγραφα με χαμηλή τοπική αντίθεση, ενώ εμφανίζει μέτρια συμπεριφορά στα επίπεδα του θορύβου. Συνοψίζοντας τα παραπάνω, η τεχνική ALM μπορεί να χρησιμοποιηθεί με επιτυχία σε έγγραφα με σκιάσεις ή λεκέδες, με θόρυβο μεσαίας διακύμανσης και χωρίς μεγάλες επικεφαλίδες.
- Το λογισμικό FineReader 8.0 παρουσιάζει μέτρια συμπεριφορά σε όλα τα είδη FineReader 8.0 υποβάθμισης της ποιότητας των εγγράφων (θόρυβος, χαμηλή τοπική αντίθεση και σκιάσεις). Όταν αυτές οι υποβαθμίσεις αυξηθούν, τότε η απόδοσή του μειώνεται

σημαντικά. Κατά συνέπεια, μπορεί να χρησιμοποιηθεί με επιτυχία σε όλες τις περιπτώσεις κειμένων που προήλθαν από σάρωση, αλλά όχι για έγγραφα πολύ κακής ποιότητας.

Η μέθοδος (Gatos, et al., 2006) αποτελεί μια πολύ καλή και αξιόπιστη λύση στα περισσότερα είδη υποβαθμίσεων της ποιότητας των έγγραφων. Εντούτοις, κατά τη διεξαγωγή των πειραμάτων βρέθηκε ότι επηρεάζεται σημαντικά από τις απότομες εναλλαγές μεταξύ φωτεινών και σκοτεινών περιοχών, στις οποίες έχει την τάση να εξάγει ακμές. Αυτός είναι και ο λόγος για τον οποίο κατατάχθηκε τρίτος στον πείραμα OCR. Έχει την πιο ικανοποιητική συμπεριφορά όσον αφορά το θόρυβο, σε σχέση με τις προαναφερθείσες μεθόδους. Όταν όμως ο θόρυβος αυτός αυξηθεί πολύ, τότε γίνεται εμφανής στο τελικό αποτέλεσμα. Λαμβάνοντας υπόψη τα παραπάνω, η μέθοδος (Gatos, et al., 2006) μπορεί να χρησιμοποιηθεί με επιτυχία στα περισσότερα είδη εγγράφων, χωρίς όμως έντονη εναλλαγή φωτεινών και σκοτεινών περιοχών περιοχών ή πολύ έντονα επίπεδα θορύβου.

Όλες οι προηγούμενες μέθοδοι εμφανίζουν πολύ ικανοποιητικά αποτελέσματα σε κάποιο τομέα, ενώ παρουσιάζουν προβλήματα σε άλλους. Σε αντίθεση, η προτεινόμενη μέθοδος παρουσιάζει την πιο σταθερή συμπεριφορά σε όλες τις περιπτώσεις υποβάθμισης των εγγράφων. Πιο συγκεκριμένα, έχει την καλύτερη συμπεριφορά σε μεγάλα επίπεδα θορύβου και δυαδικοποιεί με επιτυχία έγγραφα με σκιάσεις και λεκέδες. Μπορεί να χρησιμοποιηθεί με επιτυχία στα περισσότερα είδη υποβαθμισμένων εγγράφων και ιδιαίτερα σε εκείνα που έχουν μεγάλα επίπεδα θορύβου. Τα χαρακτηριστικά αυτά καθιστούν την προτεινόμενη μέθοδο ικανοποιητική για τη δυαδικοποίηση εγγράφων που έχουν ψηφιοποιηθεί με ψηφιακές φωτογραφικές μηχανές, όπου οι συνθήκες φωτισμού δε μπορούν να είναι το ίδιο ελέγξιμες όπως σε έναν σαρωτή.

4.6 Συμπεράσματα

Στο κεφάλαιο αυτό παρουσιάστηκε μια καινούρια μέθοδος δυαδικοποίησης εγγράφων, η οποία υιοθετεί μερικά από τα χαρακτηριστικά του ΑΟΣ, που αποτελούν τη βάση των αντιλήψεων του σκοτεινού και του φωτεινού. Πιο συγκεκριμένα, στα πλαίσια του κεφαλαίου αυτού παρουσιάστηκε μια σειρά από καινούρια στοιχεία:

- Μια καινούρια προσέγγιση για τη μοντελοποίηση των γαγγλιακών κυττάρων σκοτεινού κέντρου, η οποία είναι απλή, και δίνει αρκετά ικανοποιητικά αποτελέσματα σε πραγματικές εικόνες.
- Μια καινούρια συνάρτηση απόκρισης των κυττάρων, η οποία αυξάνει την ευαισθησία του κυττάρου στις σκοτεινές περιοχές, μειώνοντας έτσι τις επιπτώσεις των σκιάσεων και των λεκέδων.
- Ο συνδυασμός διαφορετικών κλιμάκων επεξεργασίας, με τη χρήση κυττάρων δύο διαφορετικών μεγεθών, τα οποία επεξεργάζονται διαφορετικές συχνότητες το καθένα.

Επιπλέον, η πειραματική διαδικασία εισήγαγε κάποια καινούρια σημαντικά χαρακτηριστικά.

Gatos

Προτεινόμενη

- Έναν τρόπο για την παραγωγή τεχνητών υποβαθμίσεων της ποιότητας των εγγράφων, ώστε να συνδυάζουν με ελεγχόμενο τρόπο σκιάσεις, θόρυβο και χαμηλή τοπική αντίθεση. Η προσέγγιση αυτή επιτρέπει την ποσοτική ανάλυση της συμπεριφοράς των αλγορίθμων.
- 2. Πραγματοποιήθηκε μια εκτεταμένη ποσοτική και ποιοτική σύγκριση των πιο σημαντικών αλγορίθμων που υπάρχουν στο πεδίο της δυαδικοποίησης εγγράφου.

Τα χαρακτηριστικά της προτεινόμενης μεθόδου, όπως προέκυψαν μέσα από τα πειραματικά αποτελέσματα, την καθιστούν κατάλληλη για να χρησιμοποιηθεί για τη δυαδικοποίηση εγγράφων με σημαντική υποβάθμιση ποιότητας.



- 203 Συμπεράσματα
- 202 Χαρακτηριστικά και επιδόσεις του κυκλώματος
- 198 Εναλλακτική υλοποίηση του κυκλώματος
- 196 Σύγκριση με τα αποτελέσματα του λογισμικού
- 170 Παρουσίαση της προτεινόμενης υλοποίησης
- 168 Περιγραφή υλοποιήσεων ανάλογων αλγορίθμων
- 168 Η ανάγκη για υλοποίηση σε υλικό

5

Υλοποίηση μεθόδου βελτίωσης αποχρώσεων σε FPGA

Το κεφάλαιο αυτό αναφέρεται στην υλοποίηση σε υλικό του αλγορίθμου βελτιστοποίησης αποχρώσεων ψηφιακών εικόνων που παρουσιάστηκε στο κεφάλαιο 3. Σκοπός της παρούσας υλοποίησης είναι η βελτιστοποίηση των αποχρώσεων σήματος βίντεο σε πραγματικό χρόνο, αφού κάτι τέτοιο δεν είναι δυνατό να πραγματοποιηθεί από την υλοποίηση σε λογισμικό. Αναφέρονται οι υπάρχουσες υλοποιήσεις σε υλικό, ανάλογων αλγορίθμων, που πραγματοποιούν βελτιστοποίηση αποχρώσεων βίντεο σε πραγματικό χρόνο. Περιγράφονται αναλυτικά τα χαρακτηριστικά της υλοποίησης, καθώς και τα πρωτότυπα εκείνα στοιχεία τα οποία εισάγονται σε αυτή την υλοποίηση. Επίσης, γίνεται αναλυτική σύγκριση των αποτελεσμάτων της υλοποίησης σε σχέση με τα αντίστοιχα του λογισμικού. Τέλος, παρατίθενται οι επιδόσεις της συγκεκριμένης υλοποίησης σε σχέση με αυτές των υπαρχόντων αλγορίθμων για βελτιστοποίηση αποχρώσεων σε πραγματικό χρόνο.

5.1 Αναγκαιότητα της υλοποίησης σε υλικό

Στο κεφάλαιο 3 παρουσιάστηκε ένας αλγόριθμος βελτίωσης αποχρώσεων για στατικές ψηφιακές εικόνες. Ο αλγόριθμος αυτός μπορεί να χρησιμοποιηθεί σαν ένα ειδικό λογισμικό επεξεργασίας εικόνας, το οποίο θα εκτελείται σε έναν προσωπικό υπολογιστή από το χρήστη, μετά τη λήψη των φωτογραφιών. Μια τέτοια προσέγγιση ωστόσο, περιορίζει σημαντικά τη λειτουργικότητα της διαδικασίας βελτίωσης, αφού θα πρέπει να γίνει μεταφορά των φωτογραφιών από την κάμερα στον υπολογιστή και στη συνέχεια, με τη μεσολάβηση του χρήστη, να γίνει η εκτέλεση του λογισμικού. Η όλη διαδικασία θα μπορούσε να απλοποιηθεί σημαντικά, αν η ίδια η κάμερα είχε τη δυνατότητα να βελτιώσει τις αποχρώσεις των εικόνων που κατέγραφε. Κάτι τέτοιο θα προϋπόθετε την ύπαρξη εξειδικευμένου υλικού μέσα στην κάμερα, το οποίο θα αναλάμβανε την εκτέλεση του αλγορίθμου βελτίωσης αποχρώσεων πάνω στα δεδομένα που κατέγραψε ο αισθητήρας.

Εκτός όμως από τη βελτίωση στατικών ψηφιακών εικόνων, ακόμη μεγαλύτερη είναι η ανάγκη ύπαρξης εξειδικευμένου υλικού για τη βελτίωση των αποχρώσεων σε σήματα βίντεο. Το βίντεο αποτελείται συνήθως από 25-30 στατικές εικόνες-πλαίσια (frames) σε κάθε δευτερόλεπτο. Πρέπει, λοιπόν, ο αλγόριθμος βελτίωσης αποχρώσεων να μπορεί να εκτελεστεί για όλες αυτές τις εικόνες, μέσα σε διάστημα το πολύ ενός δευτερολέπτου. Σε διαφορετική περίπτωση δε θα είναι δυνατή η βελτίωση των αποχρώσεων βίντεο σε πραγματικό χρόνο. Όταν ο αλγόριθμος εκτελείται από έναν προσωπικό υπολογιστή με επεξεργαστή Pentium 4 στα 3GHz, απαιτείται χρόνος εκτέλεσης 1.8 δευτερολέπτων για εικόνα ανάλυσης 640×480 εικονοστοιχείων. Είναι προφανές λοιπόν, ότι η υλοποίηση του αλγορίθμου σε λογισμικό, αδυνατεί να καλύψει τις χρονικές απαιτήσεις για βελτίωση των αποχρώσεων σήματος βίντεο σε πραγματικό χρόνο.

Από τα παραπάνω, είναι έκδηλη η αναγκαιότητα για υλοποίηση του αλγορίθμου βελτίωσης αποχρώσεων σε υλικό. Η υλοποίηση αυτή θα ήταν καλό να γίνει με τη μορφή ενός ASIC (Application Specific Integrated Circuit – Ολοκληρωμένο Κύκλωμα Ειδικής Εφαρμογής), το οποίο θα μπορεί να παρεμβληθεί μεταξύ του αισθητήρα της κάμερας και της εξόδου της, έτσι ώστε να βελτιώνει σε πραγματικό χρόνο τις αποχρώσεις των στατικών εικόνων ή του βίντεο που καταγράφεται. Επιπλέον, τα FPGA (Field-Programmable Gate Arrays) μπορούν να αποτελέσουν μια πολύ αξιόπιστη και οικονομική λύση για την υλοποίηση αυτή, αφού πρόκειται για διαδεδομένα εργαλεία τα οποία μπορούν να εύκολα να προγραμματιστούν σύμφωνα με τις απαιτήσεις του χρήστη.

5.2 Υπάρχουσες υλοποιήσεις

Όπως αναφέρθηκε στο κεφάλαιο 3, δύο είναι οι κύριοι αλγόριθμοι που εκτελούν ανάλογη βελτίωση αποχρώσεων σε σχέση με την προτεινόμενη μέθοδο: η κατηγορία των αλγορίθμων Retinex και ο αλγόριθμος ACE. Από τους δύο αυτούς αλγορίθμους, δεν έχει παρουσιαστεί μέχρι σήμερα καμία υλοποίηση σε υλικό για τον ACE, ενώ έχουν παρουσιαστεί μερικές υλοποιήσεις για κάποιες εκδόσεις των αλγορίθμων Retinex. Όπως

168

Στατικές εικόνες

Βίντεο

ήδη έχει αναφερθεί, ο αλγόριθμος MSR (Multi-Scale Retinex) είναι ο πιο διαδεδομένος της οικογένειας Retinex, καθώς αναπτύχθηκε και εξελίσσεται από τη NASA (Jobson, et al., 1997), και κυκλοφορεί σαν εμπορικό προϊόν από την εταιρεία TruView (Tru07). Μέχρι σήμερα, έχει παρουσιαστεί ουσιαστικά μια υλοποίηση σε υλικό από την ομάδα της NASA (Hines, et al., 2004), (Hines, et al., 2004). Η υλοποίηση αυτή έχει γίνει σε επεξεργαστή DSP (Digital Signal Processor) και αναφέρεται μόνο σε μια κλίμακα του Retinex (SSR – Single Scale Retinex). Τα ακριβή χαρακτηριστικά αυτής της υλοποίησης παρουσιάζονται στον ακόλουθο πίνακα.

	Επεξεργαστής	Εικόνα	Ρυθμός ανανέωσης fps	Αλγόριθμος
	Texas Instruments TMS320C6711 Floating point DSP	256×256	21	Single Scale Retinex σε αποχρώσεις του γκρι
	Texas Instruments TMS320C6713 Floating point DSP	256×256	28	Single Scale Retinex σε αποχρώσεις του γκρι
	Texas Instruments TMS320DM642 Fixed point DSP	256×256	34	Single Scale Retinex σε αποχρώσεις του γκρι

Πίνακας 5.1

Τα χαρακτηριστικά της υλοποίησης σε υλικό, του αλγορίθμου Retinex (Hines, et al., 2004), (Hines, et al., 2004).

Όπως φαίνεται και στον προηγούμενο πίνακα, η υλοποίηση του Retinex έχει μερικούς σημαντικούς περιορισμούς. Πρώτον, το μέγεθος της εικόνας είναι μικρό (256×256 εικονοστοιχεία) και αρκετά μικρότερο του μεγέθους που θεωρείται ως ελάχιστο για εφαρμογές βίντεο (640×480 εικονοστοιχεία). Επιπλέον, ο αλγόριθμος που υλοποιείται είναι ο Retinex για μια μόνο κλίμακα και μονόχρωμες εικόνες. Αυτό απέχει από τον πραγματικό MSR, ο οποίος επεξεργάζεται έγχρωμες εικόνες σε τρεις χωρικές κλίμακες. Τέλος, ο ρυθμός ανανέωσης γίνεται ικανός για επεξεργασία σε πραγματικό χρόνο, μόνο με τη χρήση πολύ δυνατών επεξεργαστών ψηφιακού σήματος DSP.

Πρόσφατα, παρουσιάστηκε μια υλοποίηση σε υλικό για FPGA (Balercia, et al., 2006), της οποίας οι συγγραφείς αναφέρουν πως είναι μια απλοποιημένη έκδοση του αλγορίθμου Retinex. Στην πραγματικότητα όμως, ο υλοποιούμενος αλγόριθμος δεν είναι ο Retinex που περιγράφηκε στο κεφάλαιο 3, αλλά μια τροποποίηση της διόρθωσης γ, ακολουθούμενη από μια επέκταση των τιμών στο διάστημα [0,255]. Δεν έχει υλοποιηθεί καθόλου ο υπολογισμός του προσπίπτοντος φωτισμού, που προκύπτει από τη συνέλιξη με μεγάλα χαμηλοπερατά φίλτρα, ούτε η αφαίρεση από την αρχική εικόνα στο λογαριθμικό πεδίο, ούτε η ανάλυση σε τρεις χωρικές κλίμακες. Η υλοποίηση αυτή πραγματοποιήθηκε για το Virtex-II της Xilinx και επιτυγχάνει επεξεργασία σε πραγματικό χρόνο εικόνων μεγέθους 640×480. Από τα παραπάνω είναι προφανές ότι αλγόριθμοι της κατηγορίας στην οποία βρίσκεται η προτεινόμενη μέθοδος είναι πολύ δύσκολο να υλοποιηθούν σε υλικό. Καμία ουσιαστική υλοποίηση δεν έχει πραγματοποιηθεί σε FPGA, ενώ οι υπάρχουσες υλοποιήσεις σε επεξεργαστές DSP έχουν σημαντικούς περιορισμούς σε σχέση με τον αρχικό αλγόριθμο που προσπαθούν να υλοποιήσουν.

5.3 Προτεινόμενη υλοποίηση

Στην παράγραφο αυτή, θα γίνει η αναλυτική περιγραφή της προτεινόμενης υλοποίησης του αλγορίθμου βελτίωσης αποχρώσεων. Η υλοποίηση στοχεύει στην επεξεργασία σήματος βίντεο σε πραγματικό χρόνο από ένα FPGA, το οποίο θα δέχεται σαν είσοδο τα αρχικά εικονοστοιχεία της κάμερας και θα εξάγει τα βελτιωμένα. Η μορφή του σήματος εισόδου, που μπορεί να δεχτεί το FPGA, μπορεί να ποικίλει ανάλογα με την εφαρμογή. Για το λόγο αυτό, θεωρήθηκε η ακόλουθη γενικευμένη μορφή σήματος βίντεο, σα σήμα εισόδου του συστήματος. Τα εικονοστοιχεία του κάθε πλαισίου έρχονται το ένα μετά το άλλο μέχρι να συμπληρωθεί ολόκληρη η εικόνα, χωρίς ενδιάμεσα σήματα ελέγχου ή συγχρονισμού. Το τελευταίο εικονοστοιχείο του κάθε πλαισίου ακολουθείται από το πρώτο εικονοστοιχείο του επόμενου πλαισίου, ενώ οι τρεις τιμές έντασης του κάθε εικονοστοιχείου (κόκκινο, πράσινο και μπλε) είναι 8-bit η κάθε μια και έρχονται παράλληλα.

Μορφή του σήματος εισόδου

5.3.1 Ανάλυση δομής του αλγορίθμου βελτίωσης αποχρώσεων

Η δομή του αλγορίθμου βελτίωσης αποχρώσεων θα καθορίσει σημαντικά τη συνολική μορφή της υλοποίησής του. Ο αλγόριθμος βελτίωσης αποχρώσεων αποτελείται από τέσσερα διακριτά βήματα, τα οποία φαίνονται στην ακόλουθη εικόνα.



Όπως φαίνεται στην προηγούμενη εικόνα, η παρούσα μορφή του αλγορίθμου δεν είναι ιδανική για υλοποίηση σε υλικό. Υπάρχουν τέσσερα διακριτά βήματα, το καθένα από τα οποία έχει χρονική εξάρτηση από τα προηγούμενα. Έτσι, για παράδειγμα, δεν είναι δυνατό να εκτελεστεί το δεύτερο βήμα παράλληλα με το πρώτο, αφού τα αποτελέσματα του πρώτου αποτελούν προϋπόθεση για το δεύτερο. Αναλόγως, το τρίτο βήμα πρέπει να εκτελεστεί υποχρεωτικά μετά το πέρας του δεύτερου, και το τέταρτο μετά το πέρας του τρίτου. Αυτό εξαλείφει τις όποιες δυνατότητες για παραλληλία. Επίσης, το γεγονός ότι απαιτούνται τρεις διαφορετικές σαρώσεις για μια εικόνα, μπορεί να μειώσει σημαντικά την ταχύτητα επεξεργασίας του κυκλώματος. Είναι λοιπόν απαραίτητο να υπάρξουν σημαντικές τροποποιήσεις στη δομή του αλγορίθμου.

5.3.2 Τροποποίηση του αλγορίθμου

Στην κλασική αντιμετώπιση, μετά την πρώτη σάρωση, κατά την οποία βρίσκεται το ελάχιστο και το μέγιστο της αρχικής εικόνας, απαιτείται μια δεύτερη σάρωση, στην οποία για κάθε εικονοστοιχείο εφαρμόζεται η ακόλουθη εξίσωση:

$$Y'_{ij} = \frac{Y_{ij} - \min[Y]}{\max[Y] - \min[Y]} \times 255$$
(5.1)

όπου Y η αρχική εικόνα, min[Y]και max[Y] η μέγιστη και η ελάχιστη τιμή της, Y_{ij} το εκάστοτε εικονοστοιχείο της και Y'_{ij} η καινούρια τιμή του εικονοστοιχείου για το διάστημα [0,255]. Οι καινούριες τιμές των εικονοστοιχείων χρησιμοποιούνται για τη δημιουργία ενός ιστογράμματος H με τρεις θέσεις (bin_{low} [0,84], bin_{middle} [85,169], bin_{high} [170,255]), το οποίο είναι απαραίτητο για τον υπολογισμό των παραμέτρων της συνάρτησης απεικόνισης (εξισώσεις (3.41)-(3.46)).

Η πρώτη σάρωση της εικόνας, στην οποία γίνεται η εύρεση της ελάχιστης και μέγιστης τιμής έντασης είναι μια αναπόφευκτη διαδικασία, η οποία δεν είναι δυνατό να παρακαμφθεί. Η δεύτερη σάρωση της εικόνας όμως, στην οποία γίνεται η έκταση των τιμών έντασης στο διάστημα [0,255], μπορεί να ενσωματωθεί στα γειτονικά της βήματα. Αυτό επιτυγχάνεται με την κατασκευή ενός πίνακα αντιστοίχισης *Stretching*, ο οποίος για κάθε πιθανή τιμή έντασης της αρχικής εικόνας θα περιέχει την ήδη προϋπολογισμένη τιμή για το διάστημα [0,255]. Με αυτόν τον τρόπο, αποφεύγεται η δεύτερη σάρωση (Εικόνα 5.1) για την εφαρμογή του μετασχηματισμού της έκτασης των τιμών (εξίσωση (5.1)), αφού κάθε τιμή έντασης της αρχικής εικόνας, εφαρμοζόμενη σαν είσοδος στον πίνακα *Stretching*, θα δίνει την αντίστοιχη τιμή για το διάστημα [0,255]. Κατά συνέπεια, αποφεύγεται η εκτέλεση της εξίσωσης (5.1) για το συνολικό αριθμό των εικονοστοιχείων, αφού θα εκτελεστεί το πολύ 256 φορές, μια για κάθε πιθανή τιμή έντασης της εικόνας. Η μαθηματική έκφραση του πίνακα *Stretching* περιγράφεται από την ακόλουθη εξίσωση.

$$Stretching(x) = \frac{x - \min[Y]}{\max[Y] - \min[Y]} \times 255 \quad \forall \ x \in [\min[Y], \max[Y]]$$
(5.2)

Δεδομένου του πίνακα *Stretching* και του πίνακα του αρχικού ιστογράμματος *OH* (Original Histogram), ο οποίος μπορεί εύκολα να κατασκευαστεί κατά τη διάρκεια της

πρώτης σάρωσης, είναι δυνατό να κατασκευαστεί γρήγορα και το ιστόγραμμα τριών θέσεων *H*, που είναι απαραίτητο για τον υπολογισμό των παραμέτρων του αλγορίθμου. Οι εξισώσεις (5.3)-(5.5) περιγράφουν τον υπολογισμό των τριών θέσεων του ιστογράμματος *H* για τις τιμές έντασης της εικόνας και είναι οι αντίστοιχες των εξισώσεων (3.41), (3.43) και (3.45), της κλασικής προσέγγισης που παρουσιάστηκε στο 3° κεφάλαιο.

$$bin_{low} = \sum_{i=0}^{K1} OH(i)$$
(5.3)

τέτοιο ώστε Stretching(K1) = 84

$$bin_{middle} = \sum_{i=K1+1}^{K2} OH(i)$$
(5.4)

τέτοιο ώστε Stretching(K2) = 169

$$bin_{high} = \sum_{i=K2+1}^{255} OH(i)$$
 (5.5)

Το χαρακτηριστικό των καινούριων εξισώσεων είναι ότι για τον υπολογισμό των τριών θέσεων του ιστογράμματος *H*, δεν είναι απαραίτητη η σάρωση όλων των εικονοστοιχείων, αλλά το άθροισμα συγκεκριμένων τμημάτων του αρχικού ιστογράμματος *OH*. Η Εικόνα 5.2 δείχνει παραστατικά το σχηματισμό του πίνακα *Stretching* και του ιστογράμματος *H* σε σχέση με το αρχικό ιστόγραμμα *OH*.



Έχοντας πραγματοποιήσει τις προηγούμενες αλλαγές στη δομή του αλγορίθμου, προκύπτει η καινούρια μορφή του, η οποία και παρουσιάζεται στην Εικόνα 5.3.



5.3.3 Τροποποιήσεις για επίτευξη διαδοχικής διοχέτευσης

Στόχος της προτεινόμενης υλοποίησης είναι να επιτευχθεί μια αρχιτεκτονική που θα καθιστά δυνατή τη διαδοχική διοχέτευση (pipeline) στην επεξεργασία των διαφόρων πλαισίων. Κάτι τέτοιο θα επέτρεπε στο κύκλωμα την εξαγωγή ενός διορθωμένου εικονοστοιχείου σε κάθε χτύπο ρολογιού. Λαμβάνοντας ως δεδομένο τη νέα μορφή του αλγορίθμου (Εικόνα 5.3), η Εικόνα 5.4 παρουσιάζει τη μορφή της διαδοχικής διοχέτευσης που θα υλοποιηθεί από το κύκλωμα.



Όπως έχει αναφερθεί στην αρχή της ενότητας 5.3, το σήμα του βίντεο θεωρείται ότι αποτελείται από τις τιμές των εικονοστοιχείων του κάθε πλαισίου, οι οποίες αποστέλλονται διαδοχικά, η μια μετά την άλλη, χωρίς κάποια ενδιάμεση καθυστέρηση. Αυτό σημαίνει ότι το τελευταίο εικονοστοιχείο ενός πλαισίου, ακολουθείται από το πρώτο εικονοστοιχείο του επόμενου πλαισίου. Κατά συνέπεια, είναι αδύνατο να επιτευχθεί η δεύτερη σάρωση της εικόνας, η οποία είναι απαραίτητη για την υλοποίηση του αλγορίθμου, χωρίς την αποθήκευση ολόκληρου του πλαισίου σε μια μνήμη. Για το λόγο αυτό είναι απαραίτητη η ύπαρξη μιας μνήμης FIFO, μεγέθους τουλάχιστο ενός πλαισίου, η οποία θα εισάγει στα δεδομένα την κατάλληλη καθυστέρηση ώστε να είναι δυνατή η πραγματοποίηση της δεύτερης σάρωσης. Έτσι, παράλληλα με την πρώτη σάρωση του πλαισίου k, οι τιμές των εικονοστοιχείων θα τροφοδοτούνται και στη μνήμη FIFO. Στη συνέχεια, όταν θα ξεκινήσει η πρώτη σάρωση του πλαισίου k+1, τα ήδη αποθηκευμένα δεδομένα του πλαισίου k θα εξέρχονται από τη μνήμη ξεκινώντας έτσι τη δεύτερη σάρωσή τους, η οποία θα δώσει και το τελικό αποτέλεσμα. Με τον τρόπο αυτό, είναι δυνατό να υλοποιηθεί η μορφή επεξεργασίας που παρουσιάζεται στην Εικόνα 5.4.

Για να μπορέσει να επιτευχθεί η διαδοχική διοχέτευση ωστόσο, είναι απαραίτητες κάποιες επιπλέον τροποποιήσεις, οι οποίες θα εξαλείψουν τις όποιες εξαρτήσεις υπάρχουν στις διάφορες φάσεις του αλγορίθμου. Μια πρώτη εξάρτηση, η οποία εμποδίζει την ομαλή διαδοχική διοχέτευση, εμφανίζεται τη στιγμή που στο ένα πλαίσιο γίνεται η ρύθμιση των παραμέτρων και στο επόμενο, η πρώτη σάρωση. Στην περίπτωση αυτή, η ρύθμιση των παραμέτρων προϋποθέτει τη χρήση του πίνακα αρχικού ιστογράμματος *OH* και της μέγιστης και ελάχιστης τιμής του πλαισίου. Όμως, εάν ταυτόχρονα εκτελείται η πρώτη σάρωση του επόμενου πλαισίου, τότε ο πίνακας *OH* και οι τιμές του μέγιστου και του ελάχιστου θα τροποποιούνται. Για το λόγο αυτό, ο πίνακας *OH* θα υλοποιείται σε δύο διαφορετικά τμήματα. Ο πίνακας *OH*1 θα χρησιμοποιείται μόνο στα περιττά πλαίσια, ενώ ο πίνακας *OH*2 μόνο στα άρτια πλαίσια. Έτσι για παράδειγμα, τη στιγμή που γίνεται η ρύθμιση των παραμέτρων του πλαισίου 1 και χρησιμοποιείται ο πίνακας *OH*1, ο πίνακας *OH*2 σχηματίζεται από την πρώτη σάρωση του πλαισίου 2. Με το πέρας της ρύθμισης των παραμέτρων του πλαισίου 1 και χρησιμοποιείται ο πίνακας *OH*1, ο πίνακας *OH*2 σχηματίζεται από την πρώτη σάρωση του πλαισίου 2. Με το πέρας της ρύθμισης των παραμέτρων του πλαισίου 1.

Ανάλογο πρόβλημα εμφανίζεται και με τον πίνακα Stretching, στη χρήση του οποίου υπάρχει επικάλυψη κατά τη δεύτερη σάρωση του ενός πλαισίου και τη ρύθμιση παραμέτρων του επόμενου. Επιπλέον, δημιουργείται και δεύτερη επικάλυψη, αφού το κέντρο και η περιφέρεια κάθε εικονοστοιχείου πρέπει να επεξεργάζονται ταυτόχρονα. Για το λόγο αυτό, ο πίνακας Stretching υλοποιείται από τέσσερα διαφορετικά στοιχεία. Οι Stretching1a και Stretching1b, κατασκευάζονται και χρησιμοποιούνται στα περιττά πλαίσια και έχουν ακριβώς τις ίδιες τιμές και χρησιμοποιούνται για τη ρύθμιση της τιμής του κέντρου και της περιφέρειας αντίστοιχα. Αντίθετα, οι Stretching2a και stretching1a και χρησιμοποιούνται στα άρτια πλαίσια και όπως και πριν, χρησιμοποιούνται αντίστοιχα για τη ρύθμιση του κέντρου και της περιφέρειας. Έτσι για παράδειγμα, όταν οι Stretching1a και Stretching1b χρησιμοποιούνται στη δεύτερη σάρωση του πρώτου πλαίσιο. Με το πέρας της δεύτερης σάρωσης του πρώτου

Μνήμη FIFO

Εξάλειψη εξαρτήσεων



πλαισίου, οι *Stretching1a* και *Stretching1b* απελευθερώνονται και μπορούν να σχηματιστούν ξανά στη διάρκεια της ρύθμισης του τρίτου πλαισίου.

Σε αντίθεση με τις προηγούμενες περιπτώσεις, η εύρεση της μέγιστης και της ελάχιστης τιμής του κάθε πλαισίου δεν απαιτεί δύο διαφορετικά κυκλώματα. Αυτό συμβαίνει, γιατί κατά τη διάρκεια της πρώτης σάρωσης του πλαισίου n γίνεται η εύρεση των τιμών min και max από το κύκλωμα μεγίστου-ελαχίστου, το οποίο στο τέλος της πρώτης σάρωσης μεταβιβάζει τα αποτελέσματά του στους καταχωρητές Ymin και Ymax. Έτσι, το κύκλωμα υπολογισμού μεγίστου-ελαχίστου, απελευθερώνεται και μπορεί να χρησιμοποιηθεί στην πρώτη σάρωση του επόμενου πλαισίου (n + 1). Παράλληλα, οι καταχωρητές Ymin και Ymax, οι οποίοι περιέχουν την ελάχιστη και τη μέγιστη τιμή του πλαισίου n, χρησιμοποιούνται για τον υπολογισμό των πινάκων Stretching1a και Stretching1b ή Stretching2a και Stretching2b, ανάλογα με το εάν πρόκειται για περιττό ή άρτιο πλαίσιο, χωρίς να επηρεάζουν το κύκλωμα μεγίστου-ελαχίστου.

Οι προηγούμενες τροποποιήσεις εξασφαλίζουν μια δομή που επιτρέπει την επίτευξη διαδοχικής διοχέτευσης από το σύστημα. Το τελικό μπλοκ διάγραμμα του προτεινόμενου συστήματος παρουσιάζεται στην Εικόνα 5.5. Η εικόνα αυτή θα αποτελέσει τον χάρτη πάνω στον οποίο θα κινηθεί όλη η παρουσίαση της περιγραφής της υλοποίησης.

5.3.4 Μετατροπές RGB→YCbCr και YCbCr→RGB

Η μετατροπή των χρωματικών χώρων είναι πρώτη και τελευταία διεργασία που πραγματοποιείται στα δεδομένα όταν αυτά εισέρχονται για επεξεργασία στο FPGA. Αυτό γίνεται ώστε να διαχωριστεί το χρώμα από τη μη χρωματική πληροφορία και να μείνει ανέπαφο κατά τη διάρκεια του μετασχηματισμού. Οι μετατροπές των χρωματικών χώρων RGB→YCbCr και YCbCr→RGB, έτσι όπως παρουσιάστηκαν στις εξισώσεις (3.23) και (3.24) του 3^{ου} κεφαλαίου, δε μπορούν να χρησιμοποιηθούν απευθείας για υλοποίηση σε υλικό, αφού περιλαμβάνουν πράξεις πλωτού σημείου. Για το λόγο αυτό, γίνεται μετατροπή αυτών των εξισώσεων, έτσι ώστε να αποτελούνται μόνο από ακέραιους αριθμούς. Οι εξισώσεις (5.6) και (5.7) περιγράφουν αυτή την μετατροπή.

$$\begin{bmatrix} Y\\ Cb\\ Cr \end{bmatrix} = \begin{bmatrix} 0.230R + 0.661G + 0.109B\\ -0.101R - 0.338G + 0.439B + 128\\ 0.439R - 0.399G - 0.040B + 128 \end{bmatrix} =$$

$$= \frac{1}{1024} \begin{bmatrix} 1024 \times (0.230R + 0.661G + 0.109B)\\ 1024 \times (-0.101R - 0.338G + 0.439B + 128)\\ 1024 \times (0.439R - 0.399G - 0.040B + 128) \end{bmatrix} =$$

$$= \frac{1}{1024} \begin{bmatrix} 236R + 677G + 112B\\ -103R - 346G + 450B + 131072\\ 450R - 409G - 41B + 131072 \end{bmatrix}$$

$$\begin{bmatrix} R\\ G\\ B \end{bmatrix} = \begin{bmatrix} 1.084Y + 1.793(Cr - 128)\\ 1.084Y - 0.534(Cr - 128) - 0.213(Cb - 128)\\ 1.084Y + 2.115(Cb - 128) \end{bmatrix} =$$

$$= \frac{1}{1024} \begin{bmatrix} 1024 \times [1.084Y + 1.793(Cr - 128)]\\ 1024 \times [1.084Y - 0.534(Cr - 128) - 0.213(Cb - 128)]\\ 1024 \times [1.084Y + 2.115(Cb - 128)] \end{bmatrix} =$$

$$= \frac{1}{1024} \begin{bmatrix} 11024 \times [1.084Y + 1.793(Cr - 128)]\\ 1024 \times [1.084Y + 2.115(Cb - 128)]\\ 1024 \times [1.084Y + 2.115(Cb - 128)] \end{bmatrix} =$$

$$= \frac{1}{1024} \begin{bmatrix} 1110Y + 1836Cr - 235012\\ 1110Y - 547Cr - 218(Cb + 97911\\ 1110Y + 2166Cb - 277217 \end{bmatrix}$$

Κύριος σκοπός της παραπάνω μετατροπής είναι να πολλαπλασιαστούν όλοι οι σταθεροί όροι των εξισώσεων με έναν κατάλληλο αριθμό, ώστε να γίνουν ακέραιοι, χωρίς να χαθεί η ακρίβεια από τα δεκαδικά τους ψηφία. Μετά την υλοποίηση των πράξεων, το αποτέλεσμα θα πρέπει διαιρεθεί με τον αριθμό που πολλαπλασιάστηκαν οι σταθεροί όροι, ώστε να προκύψει το σωστό αποτέλεσμα. Ο αριθμός αυτός θα πρέπει να είναι μια δύναμη του 2, ώστε να μπορεί εύκολα να υλοποιηθεί η διαίρεση, μέσω κατάλληλης ολίσθησης προς τα δεξιά. Όλοι οι σταθεροί όροι με τους οποίους πολλαπλασιάζονται οι τιμές *R*, *G* και *B* έχουν τρία δεκαδικά ψηφία. Για να μπορέσουν λοιπόν να μετατραπούν σε ακέραιους, πρέπει όλοι να πολλαπλασιαστούν με έναν αριθμό ίσο ή μεγαλύτερο από 1000, ώστε να μη χαθεί κανένα από τα τρία δεκαδικά τους ψηφία. Ο κοντινότερος αριθμός στο 1000, που είναι δύναμη του 2 είναι ο 1024 (2¹⁰=1024). Για το λόγο αυτό, όλοι οι σταθεροί όροι πολλαπλασιάζονται με το 1024, ενώ το τελικό αποτέλεσμα θα υποστεί δεξιά ολίσθηση κατά 10 θέσεις.

Η Εικόνα 5.6 παρουσιάζει την υλοποίηση της μετατροπής RGB→YCbCr. Μετά από τις πράξεις των όρων σε κάθε εξίσωση, εκτελείται ολίσθηση προς τα δεξιά κατά 10 θέσεις, που ισοδυναμεί με διαίρεση με τον αριθμό 1024, ο οποίος πολλαπλασιάστηκε στην αρχή με του σταθερούς, για να τους μετατρέψει σε ακέραιους. Επίσης, επειδή ο κλάδος υπολογισμού του Y δίνει το αποτέλεσμά του δύο στάδια πιο νωρίς από αυτούς του Cb και Cr, εισάγεται μια καθυστέρηση κατά δύο χτύπους ρολογιού, ώστε τα αποτελέσματα να συγχρονίζονται.



Η Εικόνα 5.7 παρουσιάζει αντίστοιχα τη μετατροπή YCbCr→RGB. Όπως και πριν, δύο κλάδοι εξάγουν νωρίτερα τα αποτελέσματά τους από τον τρίτο. Για το λόγο αυτό, χρησιμοποιούνται στοιχεία καθυστέρησης για το συγχρονισμό των αποτελεσμάτων. Επίσης, τα αποτελέσματα δέχονται μια δεξιά ολίσθηση κατά 10 θέσεις, ώστε να πραγματοποιηθεί διαίρεση με τον αριθμό 1024, με τον οποίο πολλαπλασιάστηκαν οι σταθεροί όροι στην αρχή. Υπάρχει όμως και μια βασική διαφορά σε σχέση με την προηγούμενη μετατροπή. Τα αποτελέσματα των τύπων της εξίσωσης (5.7) μπορεί να οδηγήσουν σε αριθμούς μικρότερους του μηδενός ή μεγαλύτερους του 255. Για το λόγο αυτό, τα τελικά αποτελέσματα τις μετατροπής πρέπει να ψαλιδιστούν μεταξύ του διαστήματος [0,255]. Αυτό επιτυγχάνεται με τη χρήση δύο συγκριτών και ενός πολυπλέκτη 4 σε 1. Οι συγκριτές ελέγχουν εάν το αποτέλεσμα είναι μεγαλύτερο από 255 ή μικρότερο από 0. Οι έξοδοί τους αποτελούν σήματα ελέγχου για τον πολυπλέκτη, ο οποίος επιλέγει μεταξύ τριών γραμμών (η τέταρτη δε χρησιμοποιείται): του αποτελέσματος, της τιμής 255 και τις τιμής 0. Σε περίπτωση που το αποτέλεσμα βρίσκεται στο διάστημα [0,255], επιλέγεται η γραμμή του ως έξοδος του πολυπλέκτη. Εάν το αποτέλεσμα είναι μεγαλύτερο του 255, τότε επιλέγεται η τιμή 255 για έξοδος, ενώ αν είναι μικρότερο του μηδενός, επιλέγεται η τιμή 0. Έτσι, το σήμα εξόδου περιορίζεται στο διάστημα [0,255].



Και τα δύο κυκλώματα των Εικόνων 5.6 και 5.7 περιλαμβάνουν αρκετούς πολλαπλασιασμούς, οι οποίοι εάν εκτελούνταν από μεμονωμένους πολλαπλασιαστές θα αύξαναν σημαντικά την κατανάλωση πόρων του κυκλώματος. Για το λόγο αυτό, όλοι οι πολλαπλασιασμοί υλοποιήθηκαν με αθροίσματα όρων που είναι πολλαπλάσιοι του αρχικού κατά μια δύναμη του 2. Οι πολλαπλάσιοι όροι προέκυψαν με αριστερές ολισθήσεις της αρχικής τιμής. Η Εικόνα 5.8 παρουσιάζει ένα παράδειγμα μιας τέτοιας υλοποίησης πολλαπλασιασμού. Είναι φανερό ότι η δομή που προκύπτει επιτυγχάνει διαδοχική διοχέτευση και έτσι, σε κάθε χτύπο ρολογιού μπορεί να εξάγει ένα αποτέλεσμα χρησιμοποιώντας μόνο προσθέσεις και στοιχεία καθυστέρησης. Πρέπει να σημειωθεί, ότι ανάλογα με το μέγεθος του κάθε πολλαπλασιασμού, η διαδικασία μπορεί να έχει διαφορετικό αριθμό σταδίων. Έτσι, χρησιμοποιήθηκαν επιπλέον στοιχεία καθυστέρησης, τα οποία για λόγους απεικόνισης δεν περιλαμβάνονται στις Εικόνες 5.6 και 5.7, ώστε όλοι οι πολλαπλασιασμοί να εξάγουν ταυτόχρονα τα αποτελέσματά τους.



5.3.5 Εύρεση μεγίστου-ελαχίστου

Κεφάλαιο 5

Κατά την πρώτη σάρωση του κάθε πλαισίου, μια από τις διεργασίες που πραγματοποιούνται είναι και ο υπολογισμός της μέγιστης και της ελάχιστης τιμής έντασης του πλαισίου. Η υλοποίηση αυτού του υπολογισμού παρουσιάζεται στην Εικόνα 5.9. Η βασική λειτουργία του κυκλώματος στηρίζεται σε δύο συγκριτές και δύο καταχωρητές. Ο πρώτος καταχωρητής, ο οποίος αποθηκεύει την προσωρινή τιμή του μεγίστου, έχει ως αρχική τιμή το μηδέν, ενώ ο δεύτερος, ο οποίος αποθηκεύει την προσωρινή τιμή του ελαχίστου, ξεκινάει από τη μέγιστη τιμή 255. Καθώς έρχονται διαδοχικά οι τιμές έντασης των εικονοστοιχείων, γίνονται δύο παράλληλοι έλεγχοι. Αν η τιμή του εικονοστοιχείου είναι μεγαλύτερη από την προσωρινή μέγιστη τιμή, τότε η συγκεκριμένη τιμή φορτώνεται στον πρώτο καταχωρητή ως καινούρια προσωρινή μέγιστη τιμή. Ομοίως, αν η τιμή του εικονοστοιχείου είναι μικρότερη από την προσωρινή ελάχιστη τιμή, τότε η συγκεκριμένη τιμή φορτώνεται στο δεύτερο καταχωρητή ως καινούρια προσωρινή ελάχιστη τιμή. Η αποθήκευση της προσωρινής μέγιστης και ελάχιστης τιμής στους δύο καταχωρητές γίνεται μέσω του σήματος CLEN (Clock Enable). Η είσοδος του σήματος αυτού συνδέεται με την έξοδο του αντίστοιχου συγκριτή. Όταν ικανοποιείται η συνθήκη του συγκριτή, αποστέλλεται λογικό '1' στην είσοδο CLEN του καταχωρητή, ο οποίος αποθηκεύει στον επόμενο χτύπο ρολογιού την τιμή έντασης του εικονοστοιχείου. Όταν η συνθήκη του συγκριτή δεν ικανοποιείται, τότε στέλνεται λογικό '0' στην είσοδο CLEN του καταχωρητή. Έτσι, δεν αποθηκεύονται οι νέες τιμές έντασης που έρχονται, ενώ ο καταχωρητής συνεχίζει να εξάγει την προσωρινή μέγιστη ή ελάχιστη τιμή που έχει αποθηκευμένη και η οποία καταλήγει στο συγκριτή για έλεγχο με τις καινούριες τιμές των εικονοστοιχείων. Βασικό στοιχείο του κυκλώματος είναι η επαναφορά στην αρχική κατάσταση των καταχωρητών, στην αρχή κάθε νέου πλαισίου. Αυτό επιτυγχάνεται με την είσοδο ASET (Asynchronous Set), η οποία όταν δεχτεί λογικό '1', επιτρέπει να φορτωθούν ασύγχρονα στον καταχωρητή τα δεδομένα που βρίσκονται στην είσοδο Β. Το σήμα frame, το οποίο σχηματίζεται στη μονάδα ελέγχου, αποτελείται από ένα θετικό παλμό πριν από την πρώτη τιμή εικονοστοιχείου του εκάστοτε πλαισίου και εξασφαλίζει τη φόρτωση της αρχικής κατάστασης κάθε καταχωρητή στην αρχή του κάθε πλαισίου.



Η τιμή που έχουν οι δύο καταχωρητές μετά το πέρας και του τελευταίου εικονοστοιχείου είναι η μέγιστη και η ελάχιστη τιμή του πλαισίου. Οι τιμές αυτές όμως, στους επόμενους χτύπους ρολογιού θα χαθούν, αφού το σήμα frame θα δώσει εντολή για ασύγχρονη φόρτωση των αρχικών τιμών των καταχωρητών, ώστε να προετοιμαστούν για το επόμενο πλαίσιο. Για το λόγο αυτό, πρέπει να αποθηκευτούν σε νέους καταχωρητές, οι οποίοι θα επιτρέψουν τη χρησιμοποίηση αυτών των τιμών στη φάση της ρύθμισης παραμέτρων. Οι δύο αυτοί καταχωρητές είναι οι *Ymin* και *Ymax*, οι οποίοι φαίνονται τόσο στην Εικόνα 5.9 όσο και στην Εικόνα 5.5. Οι καταχωρητές αυτοί ενεργοποιούνται μέσω του σήματος frame, το οποίο συνδέεται στην είσοδο CLEN. Έτσι, όταν το σήμα frame γίνει λογικό '1' στο τέλος του πλαισίου, οι δύο καταχωρητές φορτώνουν τις τιμές του μεγίστου και του ελαχίστου, που βρίσκονται στα προηγούμενα στοιχεία καθυστέρησης, και τη συγκρατούν έως ότου δεχτούν ξανά θετικό παλμό στο τέλος του επόμενου πλαισίου. Μέχρι τότε, οι τιμές αυτές χρησιμοποιούνται στον υπολογισμό των πινάκων *Stretching* που είναι απαραίτητοι για τα επόμενο τα επόμενου τα επόμενο τα επόμενο στάδια επεξεργασίας.

5.3.6 Εύρεση ιστογράμματος

Στη συγκεκριμένη εφαρμογή, η εύρεση του ιστογράμματος του κάθε πλαισίου είναι μια διαδικασία πολύ πιο πολύπλοκη σε σχέση με τη συνηθισμένη προσέγγιση που έχει παρουσιαστεί σε άλλες εφαρμογές. Στις εφαρμογές αυτές, η εύρεση του ιστογράμματος

Αδυναμία εφαρμογής της κλασσικής προσέγγισης υπολογισμού ιστογράμματος γίνεται με τη χρήση μιας μνήμης, η οποία όταν δέχεται την τιμή έντασης ενός εικονοστοιχείου, εκτελεί ανάγνωση, αυξάνει το περιεχόμενο της κατά ένα και στη συνέχεια, εγγράφει τη νέα τιμή στην ίδια θέση. Στην προκειμένη εφαρμογή, υπάρχουν ισχυρές εξαρτήσεις των στοιχείων που χρησιμοποιούνται, οι οποίες καθιστούν δύσκολη την κλασική προσέγγιση από πλευράς χρονισμού. Συγκεκριμένα, χρησιμοποιούνται δύο διαφορετικές μνήμες για την αποθήκευση του ιστογράμματος, μια για τα περιττά πλαίσια (OH1) και μια για τα άρτια (OH2). Τη στιγμή που σχηματίζεται το ιστόγραμμα και αποθηκεύεται στη μια μνήμη, η άλλη βρίσκεται σε κατάσταση ανάγνωσης για τον υπολογισμό των παραμέτρων του προηγούμενου πλαισίου. Στην περίπτωση της κλασικής προσέγγισης, η μνήμη στην οποία σχηματίζεται το ιστόγραμμα εκτελεί διαδοχικές αναγνώσεις και εγγραφές, ενώ η μνήμη, η οποία συμμετέχει στον υπολογισμό των παραμέτρων του πλαισίου, βρίσκεται σε μόνιμη κατάσταση ανάγνωσης. Υπάρχουν δηλαδή χρονικές στιγμές, στις οποίες και οι δύο μνήμες εκτελούν ταυτόχρονα ανάγνωση, ενώ μοιράζονται τις ίδιες αρτηρίες. Αυτό δημιουργεί αναπόφευκτα προβλήματα.

Για το λόγο αυτό, επιλέχθηκε μια εναλλακτική προσέγγιση για την κατασκευή του ιστογράμματος, η οποία εξαλείφει τις επικαλύψεις και εξασφαλίζει ότι τη στιγμή που η μια μνήμη εκτελεί ανάγνωση, η άλλη εκτελεί μόνο εγγραφή. Η προσέγγιση αυτή έχει δύο διακριτά στάδια. Το πρώτο στάδιο παρουσιάζεται στην Εικόνα 5.10 και στόχος του είναι η καταμέτρηση της συχνότητας εμφάνισης των τιμών έντασης των εικονοστοιχείων. Για κάθε εικονοστοιχείο που εμφανίζεται στη ροή δεδομένων του πλαισίου, το συγκεκριμένο στάδιο εξάγει ταυτόχρονα δύο διαφορετικές εξόδους: τον αριθμό των φορών που εμφανίστηκε μέχρι εκείνη τη στιγμή η τιμή έντασης του συγκεκριμένου εικονοστοιχείου (data[18..0]) και τη διεύθυνση μνήμης στην οποία πρέπει να αποθηκευτεί η τιμή αυτή (address[8..0]). Οι δύο αυτές τιμές οδηγούνται στην κατάλληλη μνήμη (OH1 ή OH2), ανάλογα με το εάν πρόκειται για περιττό ή άρτιο πλαίσιο. Η επιλογή της κατάλληλης μνήμης, καθώς και η απομόνωσή της από τη διαδικασία ρύθμισης παραμέτρων του προηγούμενου πλαισίου, αποτελούν διεργασίες που περιγράφονται στο δεύτερο στάδιο (Εικόνα 5.11).



Το πρώτο στάδιο αποτελείται από 256 διαφορετικούς συσσωρευτές, ένα για κάθε τιμή έντασης, οι οποίοι καταγράφουν τη συχνότητα εμφάνισης αυτών των τιμών. Πριν από τους συσσωρευτές υπάρχει ένας αποκωδικοποιητής 8→256, ο οποίος δέχεται σαν είσοδο την εκάστοτε τιμή έντασης του κάθε εικονοστοιχείου. Κάθε τέτοια τιμή ενεργοποιεί το λογικό '1' σε μια μόνο από τις 256 εξόδους του αποκωδικοποιητή. Η έξοδος αυτή αποτελεί είσοδο σε ένα συγκεκριμένο συσσωρευτή. Έτσι, μόλις ο συσσωρευτής αυτός δεχτεί στην είσοδό του το λογικό '1', το προσθέτει στην ήδη υπάρχουσα τιμή που έχει, αυξάνοντάς την κατά ένα. Όλες οι έξοδοι των καταχωρητών καταλήγουν σε έναν πολυπλέκτη 256→1, ο οποίος ανάλογα με τα δεδομένα που θα δεχτεί στην είσοδο select, θα εξάγει μια από τις 256 εξόδους των συσσωρευτών. Ως σήμα ελέγχου στην είσοδο select εισάγεται η τιμή έντασης του κάθε εικονοστοιχείου, η οποία αποτελεί ταυτόχρονα και τη διεύθυνση της μνήμης (address[8..0]) στην οποία θα αποθηκευτεί η τιμή του συσσωρευτή που θα εξαχθεί. Με τον τρόπο αυτό, όταν στη ροή δεδομένων παρουσιαστεί η τιμή έντασης ενός εικονοστοιχείου, αυτή φορτώνεται στον αποκωδικοποιητή, ενεργοποιεί έναν συγκεκριμένο συσσωρευτή, ο οποίος αυξάνει την τιμή του κατά ένα, και η καινούρια τιμή εμφανίζεται στην έξοδο του πολυπλέκτη (data[18..0]). Απαιτείται επίσης η χρήση συγκεκριμένων στοιχείων καθυστέρησης, ώστε να επιτευχθεί αρχιτεκτονική διαδοχικής διοχέτευσης και να εξασφαλιστεί ότι τα δύο αποτελέσματα του πρώτου σταδίου, data[18..0] και address[8..0], εμφανίζονται ταυτόχρονα στην έξοδο του κυκλώματος, έτσι ώστε να οδηγηθούν στο δεύτερο στάδιο για την αποθήκευση στην κατάλληλη μνήμη. Πρέπει να σημειωθεί ότι το μέγεθος των συσσωρευτών, καθώς και της εξόδου data[18..0], εξαρτάται άμεσα από το μέγεθος του πλαισίου. Στην περίπτωση της Εικόνας 5.10, τα 19 bit εξασφαλίζουν μέγιστη συχνότητα εμφάνισης 524288 φορές, η οποία καλύπτει τις ανάγκες μιας εικόνας μεγέθους 724×724 εικονοστοιχείων. Στην περίπτωση εικόνας ενός εκατομμυρίου εικονοστοιχείων, απαιτείται η επέκταση των συσσωρευτών στα 20 bit. Τέλος, σημαντικός είναι ο ταυτόχρονος μηδενισμός όλων των συσσωρευτών στο τέλος του κάθε πλαισίου. Αυτό επιτυγχάνεται με το σήμα frame, το οποίο παράγεται στη μονάδα ελέγχου και αποτελείται από ένα θετικό παλμό στο τέλος του πλαισίου. Το σήμα αυτό συνδέεται στην είσοδο ACLR (Asynchronous Clear) των συσσωρευτών και τους μηδενίζει όλους, προετοιμάζοντάς τους έτσι για τον υπολογισμό του ιστογράμματος του επόμενου πλαισίου.



Το δεύτερο στάδιο της διαδικασίας παρουσιάζεται στην Εικόνα 5.11. Στο στάδιο αυτό στόχος είναι η αποθήκευση των αποτελεσμάτων του προηγούμενου σταδίου στην κατάλληλη μνήμη και η απομόνωση αυτής από την παράλληλη διαδικασία της εύρεσης παραμέτρων του προηγούμενου πλαισίου. Τρεις είναι οι βασικές είσοδοι αυτού του σταδίου. Οι δύο, data[18..0] και address[8..0], είναι οι έξοδοι του πρώτου σταδίου εύρεσης ιστογράμματος, απαραίτητες για την εγγραφή των αποτελεσμάτων της συχνότητας των τιμών έντασης. Η τρίτη, read[7..0], είναι η διεύθυνση ανάγνωσης των παραμέτρων του ιστογράμματος, η οποία χρησιμοποιείται στη ρύθμιση των παραμέτρων του

προηγούμενου πλαισίου. Όπως παρουσιάστηκε και στην ενότητα 5.3.3, τη στιγμή που το ιστόγραμμα του ενός πλαισίου αποθηκεύεται, οπότε οι είσοδοι data[18..0] και address[8..0] χρησιμοποιούνται για την εγγραφή π.χ. στη μνήμη OH1, γίνεται η ρύθμιση των παραμέτρων του προηγούμενου πλαισίου οπότε η είσοδος read[7..0] χρησιμοποιείται για ανάγνωση στη μνήμη OH2. Στο επόμενο πλαίσιο, υπάρχει αντιστροφή των συνδέσεων: οι είσοδοι data[18..0] και address[8..0] ενεργούν στη μνήμη OH2, ενώ η read[7..0] στη μνήμη OH1. Για να επιτευχθεί αυτό, χρησιμοποιήθηκαν δύο πολυπλέκτες αρτηριών (BUS MUX), ένας στην κάθε μνήμη. Ως σήμα ελέγχου αυτών των πολυπλεκτών χρησιμοποιείται το σήμα odd, το οποίο παράγεται από τη μονάδα ελέγχου και γίνεται λογικό '1' στην πρώτη σάρωση κάθε περιττού πλαισίου. Παράλληλα, το ίδιο σήμα χρησιμοποιείται και για τον καθορισμό της κατάστασης εγγραφής ή ανάγνωσης της κάθε μνήμης (WREN - Write Enable). Έτσι, κατά τη διάρκεια της πρώτης σάρωσης ενός περιττού πλαισίου, το σήμα odd ενεργοποιεί την είσοδο A του BUS MUX1 και βάζει σε κατάσταση εγγραφής τη μνήμη OH1 για να καταγράψει το ιστόγραμμα του πλαισίου. Επίσης, μέσω μιας πύλης ΝΟΤ ενεργοποιεί την είσοδο B του BUS MUX2 και βάζει τη μνήμη OH2 σε κατάσταση ανάγνωσης, ώστε να συμμετέχει στη ρύθμιση των παραμέτρων του προηγούμενου πλαισίου. Στο επόμενο ακριβώς πλαίσιο, το οποίο είναι άρτιο, το σήμα odd γίνεται '0' και κατά συνέπεια οι συνδέσεις αντιστρέφονται: η μνήμη ΟΗ1 περνάει σε κατάσταση ανάγνωσης, ενώ λαμβάνει τις διευθύνσεις από την είσοδο read[7..0], και η μνήμη OH2 περνάει σε κατάσταση εγγραφής, λαμβάνοντας τις διευθύνσεις από τη την είσοδο address[8..0]. Το σήμα odd ελέγχει επίσης τη σύνδεση της εξόδου δεδομένων (Data Out) της κάθε μνήμης με την έξοδο out[18..0], η οποία οδηγεί στο τμήμα του FPGA που ρυθμίζει τις παραμέτρους των πλαισίων. Ο έλεγχος αυτός γίνεται μέσω του BUS MUX3, έτσι ώστε να συνδέεται μόνο εκείνη η μνήμη που βρίσκεται σε κατάσταση ανάγνωσης και να απομονώνεται η μνήμη που βρίσκεται σε κατάσταση εγγραφής.

5.3.7 Υπολογισμός των πινάκων Stretching

Σημαντικός είναι ο υπολογισμός των πινάκων Stretching, οι οποίοι περιέχουν τις νέες τιμές της αρχικής εικόνας για το διάστημα [0,255]. Ουσιαστικά πρέπει δηλαδή να υλοποιηθεί η εξίσωση (5.1) και να καταχωρηθεί σε έναν πίνακα (μνήμη RAM) που περιγράφεται από την εξίσωση (5.2). Η απευθείας υλοποίηση της εξίσωσης (5.1) είναι αρκετά απαιτητική από πλευράς υπολογιστικών πόρων, αφού περιέχει έναν διαιρέτη, ο οποίος μπορεί να αυξήσει αρκετά τις απαιτήσεις του κυκλώματος σε πόρους και να μειώσει τη συχνότητα λειτουργίας του. Όμοια, η απευθείας καταχώρηση όλων των πιθανών τιμών της εξίσωσης (5.1) σε έναν πίνακα αντιστοίχισης (Look-Up Table – LUT) οδηγεί σε απαγορευτικά μεγέθη μνήμης, αφού υπάρχουν τρεις ανεξάρτητες μεταβλητές των 8bit η καθεμιά.

Για τον παραπάνω λόγο, ακολουθήθηκε μια διαφορετική προσέγγιση υλοποίησης της εξίσωσης (5.2), η οποία αποφεύγει τον διαιρέτη ή το συνολικό πίνακα αντιστοίχισης. Όπως μπορεί να φανεί και στο παράδειγμα της Εικόνας 5.2, οι τιμές του πίνακα *Stretching*, μπορούν να υπολογιστούν ακολουθιακά σύμφωνα με την ακόλουθη εξίσωση.

Αδυναμία εφαρμογής της απευθείας υλοποίησης

Προτεινόμενη προσέγγιση

Stretching(x) = Stretching(x - 1) +
$$\frac{255}{max - min}$$
 $\forall x \in [min, max]$
 $\mu \varepsilon$ Stretching(min) = 0 (5.8)

~ - -

Η παραπάνω εξίσωση ουσιαστικά δηλώνει ότι όλες οι τιμές του πίνακα μπορούν να προκύψουν από τις προηγούμενες προσθέτοντας την ποσότητα 255/(max - min). Εφόσον όμως οι τιμές max και min είναι τιμές έντασης, τότε υποχρεωτικά θα βρίσκονται στο διάστημα [0,255]. Κατά συνέπεια, ανεξάρτητα από το συνδυασμό των πιθανών τιμών του max και min, η διαφορά τους θα είναι πάντα μια τιμή που θα βρίσκεται στο διάστημα [0,255], θα έχει δηλαδή 256 διαφορετικές πιθανές τιμές. Είναι δυνατό, λοιπόν, να αποθηκευθούν σε έναν πίνακα αντιστοίχισης όλες οι πιθανές τιμές της ποσότητας 255/(max – min), που, όπως είδαμε, είναι στο σύνολό τους 256. Έχοντας αποθηκευμένες όλες τις πιθανές ποσότητες αύξησης, είναι δυνατό να υπολογιστούν ακολουθιακά με τη βοήθεια ενός συσσωρευτή, όλες οι πιθανές τιμές του πίνακα Stretching μεταξύ των max και min. Στον προηγούμενο συλλογισμό υπάρχει ένα αδύνατο σημείο το οποίο καθιστά μη δυνατή την άμεση εφαρμογή του. Το σημείο αυτό είναι η ακρίβεια του υπολογισμού της ποσότητας αύξησης 255/(max – min). Σε περίπτωση που τα αποτελέσματα είναι αποθηκευμένα σε λέξεις των 8 bit, τότε θα υπάρξει σημαντικό σφάλμα, το οποίο θα μεγαλώνει, καθώς υπολογίζονται οι υψηλότεροι όροι της ακολουθίας. Για παράδειγμα, αν min=100 και max=200, τότε η ποσότητα αύξησης θα είναι 2.55, η οποία αποθηκευμένη σε μια λέξη 8 bit θα δώσει 2. Έτσι, η διαφορά του 0.55 που προκύπτει, θα αθροίζεται σε κάθε όρο της ακολουθίας, οδηγώντας σε σημαντικά σφάλματα. Αυτό το πρόβλημα μπορεί να παρακαμφθεί, εάν ο πίνακας αντιστοίχισης, καθώς και ο συσσωρευτής, ενεργούν στις αρχικές τιμές αύξησης, πολλαπλασιασμένες κατά έναν μεγάλο σταθερό όρο. Με τον τρόπο αυτό, δεσμεύονται περισσότερα bits στον υπολογισμό της ακολουθίας, με αποτέλεσμα τη μεγαλύτερη ακρίβεια. Αυτό παρουσιάζεται στην επόμενη σχέση.

$$accumulator(x) = accumulator(x - 1) + \frac{255 \cdot 1024}{max - min} \quad \forall x \in [min, max]$$

$$\mu \varepsilon \ accumulator(min) = 0 \tag{5.9}$$

$$Stretching(x) = \frac{accumulator(x)}{1024}$$

Το τελικό αποτέλεσμα της τιμής του πίνακα *Stretching* θα είναι το εκάστοτε αποτέλεσμα του συσσωρευτή, διαιρεμένο κατά το σταθερό όρο με τον οποίο είναι πολλαπλασιασμένες οι τιμές αύξησης. Ο σταθερός όρος, με τον οποίο πολλαπλασιάζονται οι αποθηκευμένες τιμές αύξησης και διαιρούνται τα αποτελέσματα του συσσωρευτή, πρέπει να είναι αρκετά μεγάλος, ώστε να δίνει την απαραίτητη ακρίβεια υπολογισμών και ταυτόχρονα να είναι δύναμη του 2, ώστε να επιτρέπει την εύκολη υλοποίηση της διαίρεσης μέσω ολισθήσεων προς τα δεξιά. Ο αριθμός αυτός επιλέχθηκε να είναι ο 1024, αφού εισάγει 10 επιπλέον bit, τα οποία αντιστοιχούν περίπου σε ακρίβεια τριών δεκαδικών ψηφίων. Η τελική υλοποίηση της σχέσης (5.9) παρουσιάζεται στην Εικόνα 5.12i.



Αρχικά υπολογίζεται η διαφορά του μεγίστου με το ελάχιστο, η οποία αποτελεί την είσοδο για τον πίνακα αντιστοίχισης, που περιλαμβάνει όλες τις πιθανές τιμές αύξησης της ακολουθίας. Η έξοδος του πίνακα αντιστοίχισης είναι η τιμή αύξησης που προκύπτει για τη συγκεκριμένη ακολουθία που καθορίζουν οι τιμές Yman και Ymin του εκάστοτε πλαισίου. Ένας μετρητής των 8 bit ξεκινάει να μετράει από το 0 έως το 255, ώστε να υπολογιστούν όλες οι πιθανές τιμές των πινάκων Stretching. Υπενθυμίζουμε ότι υπάρχουν συνολικά τέσσερεις πίνακες Stretching, των οποίων οι τιμές υπολογίζονται κατά ζεύγη. Οι πίνακες Stretching~a και Stretching~b υπολογίζονται πάντα ταυτόχρονα και έχουν ακριβώς τις ίδιες τιμές. Αντίθετα, οι πίνακες Stretching1~ χρησιμοποιούνται στα περιττά πλαίσια, ενώ οι πίνακες Stretching2~ στα άρτια πλαίσια. Η επιλογή μεταξύ των μνημών Stretching1a, 1b και Stretching2a, 2b γίνεται με τη βοήθεια ενός κυκλώματος παρόμοιο με εκείνο το οποίο χρησιμοποιήθηκε για την επιλογή μεταξύ των μνημών ΟΗ1 και ΟΗ2 (Εικόνα 5.11). Για να υπολογιστούν σωστά οι τιμές, πρέπει ο συσσωρευτής να εκτελέσει ακριβώς τόσες προσθέσεις όσοι και οι όροι της ακολουθίας, οι οποίοι βρίσκονται στο διάστημα (Ymin, Ymax]. Ο όρος Ymin δεν περιλαμβάνεται στην ακολουθία, αφού πάντα το Ymin αντιστοιχίζεται στην τιμή μηδέν. Για το λόγο αυτό, χρησιμοποιούνται δύο συγκριτές, οι οποίοι σε συνδυασμό με μια πύλη AND, υπολογίζουν τους όρους για τους οποίους ο συσσωρευτής πρέπει εκτελέσει προσθέσεις. Έτσι, όταν οι τιμές του μετρητή

Ερμηνεία του κυκλώματος βρεθούν στο διάστημα (Ymin, Ymax], τότε δίνεται σήμα CLEN στον συσσωρευτή και με κάθε χτύπο ρολογιού πραγματοποιεί και μια άθροιση. Σε κάθε αποτέλεσμα του συσσωρευτή επιλέγονται μόνο τα 8 πιο σημαντικά ψηφία από τα 18 συνολικά. Αυτό ισοδυναμεί με δεξιά ολίσθηση κατά 10 θέσεις και ουσιαστικά σε διαίρεση με το 1024. Με τον τρόπο αυτό, αναιρείται η πολλαπλασιασμένη με το 1024 τιμή των αποθηκευμένων τιμών του πίνακα αντιστοίχισης. Έτσι, προκύπτει το σωστό αποτέλεσμα, το οποίο φορτώνεται στον κατάλληλο πίνακα *Stretching* στη διεύθυνση που καθορίζει η τιμή του μετρητή. Στο τέλος της όλης διαδικασίας, ο πίνακας *Stretching* θα έχει μηδενικές τιμές στο διάστημα [0, Ymin], τις υπολογισμένες από την ακολουθία τιμές στο διάστημα [Ymin + 1, Ymax] και την τιμή 255 στο διάστημα [Ymax + 1,255].

5.3.8 Υπολογισμός ιστογράμματος Η

Κατά τη ρύθμιση των παραμέτρων για την επεξεργασία του κάθε πλαισίου, απαιτείται ο υπολογισμός του ιστογράμματος Η τριών θέσεων, όπως παρουσιάζει και η Εικόνα 5.2. Η υλοποίηση αυτού του ιστογράμματος παρουσιάζεται στην Εικόνα 5.12ii και ουσιαστικά αποτελεί εφαρμογή των εξισώσεων (5.3)-(5.5), οι οποίες προέκυψαν από τους μετασχηματισμούς που εφαρμόστηκαν στον αρχικό αλγόριθμο της μεθόδου. Οι τιμές των πινάκων Stretching (Y'), οι οποίες αποτελούν τις τιμές της εικόνας εισόδου για το διάστημα [0,255], χωρίζονται σε τρία όμοια διαστήματα από δύο συγκριτές. Ο πρώτος συγκριτής δίνει λογικό '1' για τις τιμές Y' που ανήκουν στο διάστημα [0,85]. Το σήμα αυτό ενεργοποιεί τον κατάλληλο συσσωρευτή, ο οποίος αθροίζει τις τιμές του ιστογράμματος που είναι αποθηκευμένες στους πίνακες ΟΗ1 ή ΟΗ2, ανάλογα με το εάν πρόκειται για περιττό ή άρτιο πλαίσιο. Έτσι, πραγματοποιείται η υλοποίηση της εξίσωσης (5.3) και υπολογίζεται η τιμή της πρώτης θέσης του ιστογράμματος Η (bin_{low}). Με ανάλογο τρόπο υλοποιείται και ο υπολογισμός του bin_{high}, χρησιμοποιώντας κατάλληλο συγκριτή ώστε να δίνει λογικό '1' για τις τιμές Υ' που ανήκουν στο διάστημα [170,255]. Τέλος, το bin_{middle} μπορεί να προκύψει από την αφαίρεση των δύο άλλων τιμών από το συνολικό μέγεθος της εικόνας. Πρέπει να σημειωθεί ότι από την έξοδο των συσσωρευτών χρησιμοποιούνται μόνο τα 8 πιο σημαντικά ψηφία. Αυτό δεν έχει κάποια επίπτωση στα τελικά αποτελέσματα του αλγορίθμου, αφού τα bin_{low} , bin_{middle} και bin_{high} δε χρησιμοποιούνται σαν απόλυτες τιμές, αλλά σε σχέση με το μέγεθος της εικόνας. Επίσης, όπως θα δειχθεί σε επόμενη ενότητα, η μείωση της ακρίβειας δεν επηρεάζει τα αποτελέσματα.

5.3.9 Υπολογισμός παραμέτρων

Το επόμενο απαραίτητο βήμα στην υλοποίηση του αλγορίθμου είναι ο υπολογισμός των παραμέτρων του κάθε πλαισίου για τη σωστή προσαρμογή του τελικού μετασχηματισμού. Αυτό ουσιαστικά συνεπάγεται τον υπολογισμό των M_{dark} , M_{bright} και lobe, που περιγράφονται στις εξισώσεις (3.42), (3.44) και (3.46), αντίστοιχα. Η ακρίβεια στον υπολογισμό των τριών παραμέτρων περιορίστηκε σκοπίμως στα 2 bit. Αυτό σημαίνει ότι κάθε παράμετρος μπορεί να έχει τέσσερεις διαφορετικές τιμές. Η επιλογή της χαμηλής ακρίβειας έγινε για λόγους οικονομίας μνήμης στο μέγεθος του τελικού πίνακα αντιστοίχισης, ο οποίος περιγράφεται σε επόμενη ενότητα. Παρόλα αυτά, όπως θα δειχθεί και στο τέλος αυτού του κεφαλαίου, οι διαφορές που προκύπτουν στο τελικό αποτέλεσμα, σε σύγκριση με το αντίστοιχο του λογισμικού, κυμαίνονται στα όρια του αντιληπτού. Η Εικόνα 5.13 παρουσιάζει την προτεινόμενη υλοποίηση για τον υπολογισμό των τριών παραμέτρων του αλγορίθμου.



Όπως φαίνεται στην παραπάνω εικόνα, οι τρεις παράμετροι του αλγορίθμου εξαρτώνται από τις ποσοστιαίες τιμές του ιστογράμματος τριών θέσεων Η $(bin_{low}, bin_{middle}$ και bin_{high}). Κατά συνέπεια, οι τέσσερεις στάθμες κάθε παραμέτρου πρέπει να κατανεμηθούν ανάλογα με το ποσοστό των εντάσεων της εικόνας που βρίσκεται στην αντίστοιχη θέση του ιστογράμματος Η. Πιο συγκεκριμένα, πρέπει να διαπιστωθεί σε ποιο από τα ακόλουθα διαστήματα ανήκουν οι τιμές bin_{low} , bin_{middle} και bin_{high} : [0, size/4), [size/4, size], όπου size ο συνολικός αριθμός εικονοστοιχείων της

εικόνας. Για το λόγο αυτό, πρέπει να γίνουν τρεις έλεγχοι για τον καθορισμό της τιμής κάθε παραμέτρου. Έτσι, αρχικά υπολογίζονται οι ποσότητες size/4 και size/2, οι οποίες προκύπτουν εύκολα με δεξιά ολίσθηση κατά δύο και μια θέση αντίστοιχα, ενώ η ποσότητα 3size/4 προκύπτει από το άθροισμά των δύο προηγούμενων. Οι ποσότητες αυτές χρησιμεύουν ως σημεία αναφοράς στους τρεις συγκριτές της κάθε παραμέτρου. Κάθε συγκριτής εξάγει λογικό '1' όταν ικανοποιείται η συνθήκη του. Τα λογικά αποτελέσματα των συγκριτών αθροίζονται, έτσι ώστε να προκύψει η τελική στάθμη της παραμέτρου. Ανάλογα με το ποιες συνθήκες ικανοποιούνται, προκύπτει και η κατάλληλη στάθμη, όπως φαίνεται στον Πινάκα 5.2.

Πίνακας 5.2

Η αντιστοιχία του διαστήματος στο οποίο βρίσκονται οι τιμές bin_{low}, bin_{middle} και bin_{high} σε σχέση με τη στάθμη των αντίστοιχων παραμέτρων τους.

Διάστημα	Στάθμη παραμέτρου		
[3size/4, size]	00		
[size/2, 3size/4)	01 10		
[size/4, size/2)			
[0, size/4)	11		

Εάν, για παράδειγμα, η τιμή μιας θέσης του ιστογράμματος Η βρίσκεται στο διάστημα [0,size/4), τότε θα ικανοποιούνται οι συνθήκες και των τριών συγκριτών, δίνοντας την τελική τιμή $11_B=3_D$ για τη στάθμη της αντίστοιχης παραμέτρου. Παρομοίως, εάν η τιμή βρίσκεται στο διάστημα [3size/4, size], τότε δεν ικανοποιείται καμία συνθήκη των συγκριτών και έτσι προκύπτει η στάθμη 0. Πρέπει να σημειωθεί ότι οι συγκριτές και οι αθροιστές δέχονται κατάλληλο σήμα μηδενισμού των τιμών τους στην αρχή κάθε διαδικασίας ρύθμισης παραμέτρων του εκάστοτε πλαισίου. Τέλος, όπως παρουσιάζεται και στην Εικόνα 5.5, οι τιμές των παραμέτρων αποθηκεύονται οι επικαλύψεις κατά τον υπολογισμό και τη χρήση των τιμών των παραμέτρων.

5.3.10 Σερπαντίνα και υπολογισμός της περιφέρειας

Με την έναρξη της δεύτερης σάρωσης του κάθε πλαισίου και έχοντας ήδη υπολογίσει τις ρυθμίσεις των παραμέτρων, ξεκινάει η εφαρμογή του τελικού μετασχηματισμού της εικόνας. Όπως περιγράφηκε και στο κεφάλαιο 3, η εφαρμογή του μετασχηματισμού σε κάθε εικονοστοιχείο προϋποθέτει τον υπολογισμό της περιφέρειάς του σε τρεις διαφορετικές κλίμακες. Επειδή, όμως, οι τιμές έντασης των εικονοστοιχείων εξέρχονται σειριακά από τη μνήμη FIFO, είναι απαραίτητη μια αρχιτεκτονική που θα επιτρέπει την παράλληλη επεξεργασία των περιφερειακών εικονοστοιχείων γύρω από το κεντρικό. Αυτή η αρχιτεκτονική είναι γνωστή ως *σερπαντίνα* (serpentine), (Benedetti, et al., 1998). Η Εικόνα 5.14 δείχνει τη μορφή της σερπαντίνας για μια μάσκα 3×3 εικονοστοιχείων.



Όπως φαίνεται στην προηγούμενη εικόνα, για να σχηματιστεί η μάσκα 3×3 απαιτούνται 9 συνολικά καταχωρητές και μνήμες FIFO κατάλληλου μεγέθους, το οποίο εξαρτάται από το πλάτος της εικόνας. Στην περίπτωση της προτεινόμενης εφαρμογής όμως, απαιτείται ένα πολύ μεγαλύτερο μέγεθος μάσκας για το σωστό υπολογισμό της περιφέρειας κάθε εικονοστοιχείου. Η Εικόνα 5.15 παρουσιάζει τη μορφή της μάσκας αυτής.



Εικόνα 5.15

Η μορφή της απαιτούμενης μάσκας στη σερπαντίνα της προτεινόμενης υλοποίησης. Περιγραφή προτεινόμενης μάσκας Η μάσκα της προτεινόμενης υλοποίησης έχει μέγεθος 61×61 εικονοστοιχεία, που σημαίνει ότι στην αρχιτεκτονική σερπαντίνας απαιτούνται 61×61 καταχωρητές και 60 μνήμες FIFO μεγέθους size_y-61, όπου size_y το πλάτος της εικόνας. Στην προτεινόμενη μάσκα της Εικόνας 5.15 υπάρχουν 5 διαφορετικές ομάδες καταχωρητών. Σε κάθε χτύπο ρολογιού ο κεντρικός καταχωρητής περιέχει την τιμή του κεντρικού εικονοστοιχείου. Τα λευκά τετράγωνα της Εικόνας 5.15 αντιστοιχούν σε καταχωρητές οι οποίοι παραμένουν αχρησιμοποίητοι. Γύρω από το κεντρικό εικονοστοιχείο υπάρχουν σε ομόκεντρη διάταξη τρεις επιπλέον ομάδες καταχωρητών, μια για κάθε κλίμακα περιφέρειας (*S*1, *S*2, *S*3). Κάθε κλίμακα αποτελείται από τέσσερα όμοια τεταρτημόρια. Στόχος είναι ο υπολογισμός της μέσης έντασης των εικονοστοιχείων στις τρεις χωρικές κλίμακες. Αυτό προϋποθέτει το άθροισμα και τη διαίρεση όλων των εικονοστοιχείων γύρω από το κεντρικό για κάθε κλίμακα. Για το λόγο αυτό, το μέγεθος των κλιμάκων και των τεταρτημόριων τους έχει επιλεχθεί έτσι ώστε να είναι πάντα δύναμη του δύο. Αυτό απλοποιεί εξαιρετικά την απαραίτητη πράξη της διαίρεσης, αφού ουσιαστικά την ανάγει σε δεξιά ολίσθηση. Οι τρεις κλίμακες, καθώς και οι μέσες τιμές έντασής τους παρουσιάζονται στις επόμενες εξισώσεις.

51 52 53	= S1a + S1b + S1c + S1= S1 + S2a + S2b + S2c= S2 + S3a + S3b + S3c	d + S2d + S3d	(128 εικονοστοιχεία) (1024 εικονοστοιχεία) (2048 εικονοστοιχεία)	(5.10)
	Surround1 = S1/128	δεξιά ολ	λίσθηση 7 θέσεων	
	Surround2 = S2/1024	δεξιά ολ	λίσθηση 10 θέσεων	(5.11)
	Surround3 = S3/2048	δεξιά ολ	λίσθηση 11 θέσεων	

Ιδιαίτερο ενδιαφέρον παρουσιάζει η μεγάλη κλίμακα S3. Σε αντίθεση με τις άλλες δύο κλίμακες, οι καταχωρητές της δεν καταλαμβάνουν μια ομοιόμορφη περιοχή, αλλά είναι διατεταγμένοι εναλλάξ σε συνδυασμό με μη χρησιμοποιούμενους καταχωρητές. Αυτό πραγματοποιείται για να καλυφθεί όσο το δυνατό περισσότερος χώρος γύρω από το κεντρικό εικονοστοιχείο με το μισό αριθμό καταχωρητών. Η τεχνική αυτής της διάταξης χρησιμοποιείται μόνο στην περιφέρεια της μάσκας, γιατί τα περιφερειακά εικονοστοιχεία ασκούν μικρότερη επίδραση στο κεντρικό σε σχέση με τα άμεσα γειτονικά του. Παρόμοια προσέγγιση έχει παρουσιαστεί για τον υπολογισμό της περιφέρειας της μεθόδου Retinex, σε επίπεδο λογισμικού όμως, στην εργασία (Provenzi, et al., 2007). Στην περίπτωση εκείνη, χρησιμοποιείται μια μορφή «σπρέι» (τυχαίες θέσεις εικονοστοιχείων γύρω από το κεντρικό) για την επιλογή εκείνων των εικονοστοιχείων που θα συμμετάσχουν στον υπολογισμό της περιφέρειας.

Κύκλωμα παράλληλης άθροισης Η Εικόνα 5.16 παρουσιάζει τον τρόπο με τον οποίο γίνεται το άθροισμα των όρων των τριών κλιμάκων, καθώς και ο υπολογισμός της μέσης τιμής τους. Βασικός περιορισμός στην παράλληλη άθροιση των όρων των τριών κλιμάκων είναι το γεγονός ότι ο μέγιστος αριθμός παράλληλων λέξεων που μπορούν να αθροιστούν ταυτόχρονα είναι 128. Για το λόγο αυτό, οι όροι των τριών κλιμάκων χωρίζονται σε υποσύνολα των 128, τα οποία αθροίζονται από έναν παράλληλο αθροιστή 128 θέσεων.



Οι όροι της μικρής κλίμακας είναι στο σύνολό τους 128, οπότε αθροίζονται από έναν παράλληλο αθροιστή 128 θέσεων. Οι όροι της μεσαίας κλίμακας είναι συνολικά 896, οπότε για την παράλληλη άθροισή τους απαιτούνται 7 αθροιστές. Τέλος, οι όροι της μεγάλης κλίμακας είναι 1024, οπότε στην περίπτωσή τους απαιτούνται 8 αθροιστές. Συνολικά, το κύκλωμα παράλληλης άθροισης αποτελείται από 16 παράλληλους αθροιστές 128 θέσεων. Επίσης, πρέπει να σημειωθεί ότι οι μεγάλες κλίμακες περιλαμβάνουν και τις μικρότερες από αυτές. Έτσι, η κλίμακα *S2* περιλαμβάνει και τις τιμές της *S1*, ενώ η *S3* τις τιμές της *S2*. Συνεπώς, για να ολοκληρωθεί η άθροιση των όρων της μικρότερης από αυτήν κλίμακας. Για το λόγο αυτό, προστίθενται τα απαραίτητα στοιχεία καθυστέρησης, ώστε οι τρεις μέσες τιμές να εξάγονται ταυτόχρονα.

Για να μπορέσει, τέλος, να εφαρμοστεί ο τελικός μετασχηματισμός, πρέπει να υπολογιστεί ένα σταθμισμένο άθροισμα των τριών μέσων τιμών, καθώς και οι αντίστοιχες τιμές του κέντρου και της περιφέρειας, στο διάστημα [0,255], χρησιμοποιώντας τους κατάλληλους πίνακες *Stretching*. Η Εικόνα 5.17 παρουσιάζει την υλοποίηση αυτών των υπολογισμών. Το σταθμισμένο άθροισμα των τριών μέσων τιμών που υπολογίζεται είναι το ακόλουθο.

$$Surround = \frac{2S1 + S2 + S3}{4}$$
(5.12)

όπου Surround είναι η τελική τιμή της περιφέρειας πριν την εφαρμογή του πίνακα Stretching. Στο κλάσμα της εξίσωσης (5.12) η μικρή κλίμακα έχει διπλάσιο βάρος από τις

άλλες δύο. Αυτό γίνεται για δύο λόγους. Πρώτον, τα κοντινότερα εικονοστοιχεία στο κεντρικό ασκούν μεγαλύτερη επιρροή. Δεύτερον, διευκολύνεται η διαίρεση, αφού αρκούν δύο δεξιές ολισθήσεις. Τέλος, όπως παρουσιάζει και η Εικόνα 5.17, για την τιμή της περιφέρειας χρησιμοποιείται η μνήμη Stretching~b, ενώ για την τιμή του κέντρου η μνήμη Stretching~a. Και οι δύο μνήμες έχουν ακριβώς τις ίδιες τιμές, όπως περιγράφηκε στην ενότητα 5.3.7.



Εικόνα 5.17

Ο υπολογισμός της τιμής του κέντρου και της περιφέρειας πριν την εφαρμογή του τελικού μετασχηματισμού.

5.3.11 Πίνακας απεικόνισης και τελικός μετασχηματισμός

Ο τελικός μετασχηματισμός της μεθόδου, ο οποίος παρουσιάστηκε στις εξισώσεις (3.26)-(3.29) αποτελείται από πράξεις που έχουν πολύ μεγάλο κόστος σε πόρους (πολλαπλασιασμοί, διαίρεση, ύψωση σε δύναμη) και οι οποίες μπορούν να μειώσουν σημαντικά τη συχνότητα του ρολογιού του συστήματος. Για το λόγο αυτό, επιλέχθηκε η χρήση ενός πίνακα αντιστοίχισης, ο οποίος θα περιλαμβάνει όλες τις πιθανές τιμές του μετασχηματισμού. Ο πίνακας αυτός θα είναι ουσιαστικά μια μνήμη ROM, η οποία θα δέχεται ως διευθύνσεις τις τιμές των παραμέτρων εισόδου του μετασχηματισμού και θα εξάγει σε έναν χτύπο ρολογιού το αποτέλεσμα, αποφεύγοντας τις πράξεις μεγάλου υπολογιστικού κόστους.

Τέσσερεις είναι οι παράμετροι που καθορίζουν το αποτέλεσμα του μετασχηματισμού: δύο καθολικοί (M_{dark} ή M_{bright} και lobe) και δύο τοπικοί παράμετροι ($Center'_{ii}$ και Surround';). Οι καθολικοί παράμετροι, όπως ήδη έχει περιγραφεί, καθορίζονται από την κατανομή των εντάσεων κάθε εικόνας και διατηρούνται σταθεροί για όλα τα εικονοστοιχεία της. Στόχος της ύπαρξής τους είναι να προσαρμόσουν την επιφάνεια απεικόνισης (βλέπε Εικόνα 3.18 – κεφ. 3) στα χαρακτηριστικά της εκάστοτε εικόνας. Αντίθετα, οι τοπικοί παράμετροι είναι διαφορετικοί για κάθε εικονοστοιχείο και παράμετροι καθορίζουν μια συγκεκριμένη θέση πάνω στην επιφάνεια απεικόνισης, η οποία προσδιορίζει τη βελτιωμένη τιμή του εικονοστοιχείου. Η ακρίβεια των παραμέτρων, έτσι όπως παρουσιάστηκε στο κεφάλαιο 3 για την υλοποίηση σε λογισμικό είναι: 8 bits για τις τοπικές παραμέτρους, 8 bits για την παράμετρο M_{dark} ή M_{bright} και 5 bits για την παράμετρο *lobe*. Σε περίπτωση που και οι τέσσερεις παράμετροι χρησιμοποιούνταν χωρίς καμία αλλαγή στην ακρίβειά τους, τότε το μέγεθος της απαιτούμενης μνήμης ROM θα έπρεπε να είναι 2^{29} ×8bit = 512MB, μέγεθος που είναι απαγορευτικό σύμφωνα με την παρούσα τεχνολογία. Για το λόγο αυτό πρέπει να γίνει επιλεκτική μείωση της ακρίβειας των παραμέτρων. Η Εικόνα 5.18 παρουσιάζει την επίδραση αυτής της μείωσης στα αποτελέσματα της μεθόδου.



Όπως φαίνεται και από την προηγούμενη εικόνα, υπάρχει σημαντική διαφορά μεταξύ της μείωσης της ακρίβειας των τοπικών και των καθολικών παραμέτρων. Οι τοπικοί παράμετροι καθορίζουν τις τοπικές σχέσεις έντασης μεταξύ των γειτονικών εικονοστοιχείων. Αυτό σημαίνει ότι έχουν μεγάλη επίδραση στο τελικό αποτέλεσμα, αφού ακόμα και μικρή μείωση της ακρίβειάς τους δημιουργεί εμφανή υποβάθμιση της τελικής εικόνας. Ήδη από την ακρίβεια των 6 bits (τρίτη σειρά στην Εικόνα 5.18), παρουσιάζεται εμφανής κβαντισμός των αποχρώσεων. Για το λόγο αυτό, η μείωση της ακρίβειας των τοπικών παραμέτρων πρέπει να είναι περιορισμένη. Έτσι, πραγματοποιήθηκε διερεύνηση για τον ακριβή καθορισμό του μέγιστου σφάλματος που επιφέρει η εκάστοτε μείωση της ακρίβειάς της διερεύνησης παρουσιάζονται στην Εικόνα 5.19.



Όπως είναι φανερό, η μείωση της ακρίβειας των τοπικών παραμέτρων από 8 σε 7 bits, επιφέρει μέγιστο σφάλμα τριών βαθμίδων έντασης. Αυτό σημαίνει ότι εάν X είναι η τιμή έντασης ενός εικονοστοιχείου, όπως αυτή υπολογίζεται από το λογισμικό, η τιμή που θα εξάγεται από το υλικό θα βρίσκεται εντός του διαστήματος [X-3, X+3]. Τέτοιες διαφορές είναι πολύ δύσκολο να γίνουν αντιληπτές από τον άνθρωπο, ειδικά όταν πρόκειται για βίντεο. Για το λόγο αυτό η ακρίβεια των τοπικών παραμέτρων στην παρούσα υλοποίηση επιλέχθηκε να είναι 7 bits.

Καθολικοί παράμετροι

Αντίθετα με τις τοπικές παραμέτρους, οι καθολικοί παράμετροι καθορίζουν το βαθμό της βελτίωσης που θα εφαρμοστεί στις υπερεκτεθειμένες ή τις υποεκτεθειμένες περιοχές. Όπως φαίνεται και στην Εικόνα 5.18 η μείωση της ακρίβειας των καθολικών παραμέτρων έχει σαν αποτέλεσμα μια μικρή αύξηση της φωτεινότητας της σκοτεινής περιοχής, η οποία προέρχεται από το κβαντισμό του βαθμού βελτίωσης. Η αύξηση αυτή της φωτεινότητας είναι καθολική για όλες τις υποεκτεθειμένες περιοχές της εικόνας και δε συνοδεύεται από τοπικές αλλοιώσεις των εντάσεων των εικονοστοιχείων, όπως συμβαίνει με τις τοπικές παραμέτρους. Αυτό σημαίνει ότι οι καθολικοί παράμετροι μπορούν να υποστούν μεγαλύτερη μείωση στην ακρίβειά τους, σε σχέση με τις τοπικές παραμέτρους, χωρίς αυτό να δημιουργεί σημαντικές αλλαγές στο τελικό αποτέλεσμα. Για το λόγο αυτό η ακρίβεια των καθολικών παραμέτρων στην παρούσα υλοποίηση επιλέχθηκε να είναι 2 bits.

Ο Πίνακας 5.3 παρουσιάζει τις παραμέτρους της συνάρτησης απεικόνισης, καθώς και τα bits της ακρίβειας που χρησιμοποιήθηκαν στο λογισμικό και στην υλοποίηση σε υλικό.

	Τοπικές Παράμετροι		Καθολικές παράμετροι		Σύνολο	Πίνακας 5.3
	Κέντρο (Y _{ij})	Περιφέρεια (Surround' _{ij})	M _{dark} ή M _{bright}	lobe		Η ακρίβεια των παραμέτρων της συνάρτησης απεικόνισης για το
Λογισμικό (bits)	8	8	8	5	29	
Υλικό (bits)	7	7	2	2	18	λογισμικό και την υλοποίηση σε υλικό.

Το συνολικό μέγεθος της μνήμης ROM μετά τη μείωση της ακρίβειας των παραμέτρων εισόδου είναι 2¹⁸×8bit = 256KB, μέγεθος που, αν και μεγάλο, είναι εντός των δυνατοτήτων των σημερινών συσκευών FPGA.

Η διευθυνσιοδότηση της μνήμης ROM φαίνεται στην ακόλουθη εικόνα.



Όπως φαίνεται στον Πίνακα 5.3 και στην Εικόνα 5.20, οι παράμετροι M_{dark} ή M_{bright} δε χρησιμοποιούνται ποτέ ταυτόχρονα. Αυτό προκύπτει από τις εξισώσεις της μεθόδου στο κεφάλαιο 3. Η επιλογή μεταξύ αυτών των δύο παραμέτρων εξαρτάται από το πιο σημαντικό ψηφίο της τιμής της περιφέρειας. Όταν η περιφέρεια έχει τιμή μεγαλύτερη από 128, τότε χρησιμοποιείται το M_{bright} . Διαφορετικά χρησιμοποιείται το M_{dark} . Για το λόγο αυτό, υπάρχει ειδικό κύκλωμα, βασισμένο στο πιο σημαντικό ψηφίο της περιφέρειας και σε έναν πολυπλέκτη, το οποίο επιλέγει μεταξύ των δύο αυτών παραμέτρων ώστε να αποσταλούν στην αρτηρία διευθύνσεων της μνήμης ROM.

5.4 Σύγκριση με τα αποτελέσματα του λογισμικού

Η μείωση της ακρίβειας των παραμέτρων στη συνάρτηση απεικόνισης αποχρώσεων έγινε με κύριο γνώμονα την όσο το δυνατό μικρότερη αισθητική επίδραση στο τελικό αποτέλεσμα. Εντούτοις, η σύγκριση των αποτελεσμάτων μεταξύ υλικού και λογισμικού είναι επιβεβλημένη. Η Εικόνα 5.21 παρουσιάζει αυτή τη σύγκριση.


Σύγκριση των αποτελεσμάτων υλικού και λογισμικού.



Όπως είναι φανερό, οι διαφορές των αποτελεσμάτων μεταξύ υλικού και λογισμικού κυμαίνονται στα όρια του αισθητού για τον άνθρωπο. Έτσι, είναι σχεδόν αδύνατο να εντοπιστούν οι όποιες διαφορές με την απευθείας παρατήρηση των δύο αποτελεσμάτων. Για το λόγο αυτό, στην τελευταία γραμμή της Εικόνας 5.21 παρατίθεται η αφαίρεση μεταξύ των αποτελεσμάτων υλικού και λογισμικού, καθώς και το ιστόγραμμά της. Όπως ήταν αναμενόμενο, οι διαφορές που προκύπτουν οφείλονται στο διαφορετικό βαθμό βελτίωσης που εφαρμόζεται στις υποεκτεθειμένες ή υπερεκτεθειμένες περιοχές λόγω της μείωσης στης ακρίβεια των καθολικών παραμέτρων. Εντούτοις όμως, ο βαθμός βελτίωσης που πρόκυπτουν στα αποτελέσματα του υλικού περιοχή είναι κάτι το υποκειμενικό και δεν είναι δυνατό να καθοριστεί με αντικειμενικό τρόπο. Εφόσον λοιπόν οι διαφορές που προκύπτουν στα αποτελέσματα του υλικού περιορίζονται μόνο στο βαθμό βελτίωσης των υποεκτεθειμένων ή υπερεκτεθειμένων περιοχών και όχι σε τοπικές αλλοιώσεις των εντάσεων των εικονοστοιχείων, η μείωση της ακρίβειας των στην ποιότητα του τελικού αποτελέσματος.

5.5 Εναλλακτική υλοποίηση του κυκλώματος

Η υλοποίηση που παρουσιάστηκε στις προηγούμενες ενότητες έχει υψηλές απαιτήσεις σε μνήμη, λόγω της μνήμης FIFO η οποία απαιτείται για την πραγματοποίηση της δεύτερης σάρωσης της εικόνας. Αυτό μπορεί να αποτελέσει σημαντικό πρόβλημα αφού για εικόνες μεγάλου μεγέθους η εσωτερική μνήμη ενός FPGA δεν είναι επαρκής και αναπόφευκτα οδηγούμαστε στη χρήση εξωτερικής μνήμης. Μια τέτοια επιλογή καθιστά το κύκλωμα κατασκευαστικά ακριβότερο και την απόδοσή του εξαρτώμενη από τη μέγιστη ταχύτητα μετάδοσης της μνήμης. Για το λόγο αυτό στην ενότητα αυτή παρουσιάζεται μια εναλλακτική μορφή της προηγούμενης υλοποίησης, η οποία παρακάμπτει την αναγκαιότητα της μνήμης FIFO. Έτσι, το τελικό σύστημα γίνεται φθηνότερο, γρηγορότερο και απλούστερο.

Η βασική ιδέα στην οποία βασίζεται η εναλλακτική αυτή υλοποίηση είναι ότι σε ένα σήμα βίντεο οι διαφορές μεταξύ των γειτονικών πλαισίων συνήθως δεν είναι μεγάλες. Η διαπίστωση αυτή απαλείφει την αναγκαιότητα για δύο διαφορετικές σαρώσεις ενός πλαισίου, αφού μπορούν να χρησιμοποιηθούν οι παράμετροι που υπολογίστηκαν σε προηγούμενα πλαίσια. Η Εικόνα 5.22 παρουσιάζει τη νέα μορφή της διαδοχικής διοχέτευσης για την εναλλακτική υλοποίηση, σύμφωνα με την οποία στη διάρκεια του πλαισίου k, πραγματοποιούνται τρεις παράλληλες διεργασίες:

- Η επεξεργασία των αποχρώσεων του πλαισίου k, χρησιμοποιώντας τον πίνακα Stretching1 με τις παραμέτρους του πλαισίου k-2 (parameters1), τα οποία έχουν ήδη υπολογιστεί στη διάρκεια του πλαισίου k-1.
- 2. Η καταγραφή της μέγιστης και ελάχιστης τιμής του πλαισίου k καθώς και του ιστογράμματός του (*OH*1).
- Ο υπολογισμός του πίνακα Stretching2 και των παραμέτρων του πλαισίου k-1 (parameters2), τα οποία θα χρησιμοποιηθούν στην επεξεργασία του πλαισίου k+1.

Ερμηνεία της διαδοχικής διοχέτευσης



Η νέα αυτή διαδικασία απαλείφει την ανάγκη για ύπαρξη εξωτερικής μνήμης. Έτσι, το κύκλωμα αποκτά το μπλοκ διάγραμμα που παρουσιάζεται στην Εικόνα 5.23.



Η εναλλακτική υλοποίηση που παρουσιάστηκε στην ενότητα αυτή, βασίζεται στην παραδοχή ότι τα γειτονικά πλαίσια έχουν παρόμοια χαρακτηριστικά. Παρόλα αυτά, είναι επιβεβλημένη η διερεύνηση των σφαλμάτων που μπορούν να προκύψουν από αυτή την προσέγγιση για όλες τις πιθανές περιπτώσεις. Η Εικόνα 5.24 παρουσιάζει αυτή τη υλοποίησης διερεύνηση για διαφορετικές περιπτώσεις του πλαισίου k-2, του οποίου οι παράμετροι χρησιμοποιούνται για τη διόρθωση του πλαισίου k. Εκτός από τις εικόνες, παρατίθενται τα ιστογράμματά τους, οι τρεις θέσεις του ιστογράμματος Η $(bin_{low}, bin_{middle}$ και $bin_{high})$ καθώς και οι παράμετροι του τελικού μετασχηματισμού (M_{dark} ή M_{bright} και lobe).

Διερεύνηση σφαλμάτων της εναλλακτικής

Η πρώτη περίπτωση του πλαισίου k-2, αποτελεί και τη συνηθέστερη: εφόσον δεν υπάρχει αλλαγή σκηνής, οι διαφορές που μπορούν να προκύψουν μεταξύ δύο διαδοχικών πλαισίων σε 2/25 = 0.08 οκτώ εκατοστά του δευτερολέπτου είναι πολύ μικρές. Κατά συνέπεια, τα δύο πλαίσια (το k και το k-2) διαφέρουν ελάχιστα. Μάλιστα, στην περίπτωση 1 της Εικόνας 5.24 η διαφορές μεταξύ των δύο πλαισίων είναι τόσο μικρές που δεν εμφανίζονται στις τρεις θέσεις του ιστογράμματος Η. Έτσι, η διόρθωση του πλαισίου k γίνεται ακριβώς με τις ίδιες παραμέτρους που θα υπήρχαν εάν γινόταν δύο σαρώσεις του πλαισίου k.

Η δεύτερη περίπτωση της Εικόνας 5.24, αποτελεί μια δύσκολη περίπτωση, αφού το πλαίσιο k και το πλαίσιο k-2 δεν έχουν καμία απολύτως ομοιότητα. Τα ιστογράμματά τους είναι τελείως διαφορετικά, και αυτό εκφράζεται στις τρεις θέσεις του ιστογράμματος Η. Έτσι, οι παράμετροι που προκύπτουν για τη ρύθμιση του μετασχηματισμού είναι διαφορετικοί. Εντούτοις, η διαφορά στο τελικό αποτέλεσμα είναι κάτω από τα όρια του αντιληπτού. Αυτό φαίνεται ξεκάθαρα στα ιστογράμματα των τελικών αποτελεσμάτων, τα οποία είναι σχεδόν ίδια. Αυτό αποδεικνύει ότι ακόμη και όταν υπάρχει αλλαγή σκηνής η διαφορά στο τελικό αποτέλεσμα είναι ανεπαίσθητη.

Οι δύο επόμενες περιπτώσεις αποτελούν τη χειρότερη περίπτωση που μπορεί να συμβεί στην αλληλουχία των πλαισίων: την ύπαρξη ενός εντελώς λευκού και ενός μαύρου πλαισίου. Στης περιπτώσεις αυτές, οι οποίες δεν είναι δυνατό να εμφανιστούν σε πραγματικές συνθήκες (εκτός εξαιρετικά σπάνιες περιπτώσεις), όλο το ιστόγραμμα του πλαισίου k-2 είναι συγκεντρωμένο σε μια και μόνο θέση. Οι παράμετροι που προκύπτουν είναι αρκετά διαφορετικές σε σχέση με τις προηγούμενες περιπτώσεις. Εντούτοις, αυτό δεν εκφράζεται και στο τελικό αποτέλεσμα. Η διαφορά που προκύπτει για την περίπτωση 3 κυμαίνεται επίσης στα όρια του αντιληπτού, όπως δείχνει και το τελικό ιστόγραμμα. Η μεγαλύτερη απόκλιση παρουσιάζεται στην περίπτωση 4. Ακόμη και αυτή όμως, δεν είναι αρκετή για να γίνει αντιληπτή από τον ανθρώπινο παρατηρητή, αφού θα προβληθεί για 1/25=0.04 (τέσσερα εκατοστά του δευτερολέπτου). Έχοντας υπόψη ότι αυτή η περίπτωση είναι και η χειρότερη που μπορεί να υπάρξει είναι φανερό από την παραπάνω διερεύνηση ότι η εναλλακτική αυτή υλοποίηση του αλγορίθμου μπορεί να χρησιμοποιηθεί με επιτυχία για βελτίωση αποχρώσεων σήματος βίντεο χωρίς να δημιουργεί προβλήματα στο τελικό αποτέλεσμα.



Εικόνα 5.24

Διερεύνηση της επίδρασης του πλαισίου k-2, στην ποιότητα του τελικού αποτελέσματος.

5.6 Χαρακτηριστικά του κυκλώματος και της συσκευής FPGA

Οι δύο υλοποιήσεις που περιγράφηκαν στις προηγούμενες ενότητες σχεδιάστηκαν στο περιβάλλον Quartus II 5.1 της ALTERA και προσομοιώθηκε για μια συσκευή Stratix II GX: EP2SGX130GF1508C5. Η συγκεκριμένη συσκευή επιλέχθηκε λόγω των αυξημένων απαιτήσεων του κυκλώματος σε εσωτερική μνήμη. Οι αυξημένες αυτές απαιτήσεις προέρχονται κυρίως από τον τελικό πίνακα αντιστοίχισης, το μεγάλο μέγεθος της σερπαντίνας και της μνήμης FIFO. Από όλες τις διαθέσιμες συσκευές FPGA της εταιρείας ALTERA, η σειρά Stratix ΙΙ είναι αυτή της οποίας τα ονομαστικά χαρακτηριστικά ικανοποιούν τις απαιτήσεις του κυκλώματος. Μικρότερες σε δυνατότητα συσκευές, όπως αυτές της σειράς Cyclone, δεν πληρούν τα απαραίτητα χαρακτηριστικά.

Τα ακριβή χαρακτηριστικά του κυκλώματος, έτσι όπως εξήχθησαν από το λογισμικό Quartus II παρουσιάζονται στον Πίνακα 5.4.

Χαρακτηριστικά του κυκλώματος για συσκευή Stratix II GX: EP2SGX130GF1508C5					
	Ακριβής Υλοποίηση του αλγορίθμου			Εναλλακτική	
	Χωρίς Εξωτερική Μνήμη		Με Εξωτερική Μνήμη	Υλοποίηση	
Είδος εικόνας	Έγχρωμη (24bit)	Αποχρώσεις του γκρι (8bit)	Έγχρωμη (24bit)	Έγχρωμη (24bit)	_
Μέγιστο μέγεθος εικόνας	400×400 εικονοστοιχεία	640×480 εικονοστοιχεία	Εξαρτώμενο από την ταχύτητα μετάδοσης της μνήμης	2.5 εκατομμύρια εικονοστοιχεία	Πίν Τα 1 χαρ
Μέγιστος ρυθμός ανανέωσης	25 πλαίσια / δευτερόλεπτο	25 πλαίσια / δευτερόλεπτο	Εξαρτώμενος από την ταχύτητα μετάδοσης της μνήμης	25 πλαίσια / δευτερόλεπτο	του προ κυκ
Συνολικά ALUTs	56.675/106.032 (53%)	56.672/106.032 (53%)	56.431/106.032 (53%)	56.431/106.032 (53%)	
Συνολικοί καταχωρητές	48.111	48.111	48.031	48.031	
Συνολικά bits μνήμης	5.487.088 / 6.747.840 (81%)	5.093.872 / 6.747.840 (75%)	2.603.504 / 6.747.840 (39%)	2.603.504 / 6.747.840 (39%)	
Μέγιστη συχνότητα	66,66 MHz				

ς 5.4

η *γριστικά* όμενου ατος.

Όπως φαίνεται από τον προηγούμενο πίνακα, η ακριβής υλοποίηση του αρχικού αλγορίθμου παρέχει τη δυνατότητα για επεξεργασία σε πραγματικό χρόνο (25 πλαίσια ανά δευτερόλεπτο) έγχρωμων εικόνων μεγέθους 400×400 εικονοστοιχείων, είτε εικόνων σε αποχρώσεις του γκρι μεγέθους 640×480 εικονοστοιχείων. Στην πρώτη περίπτωση καταλαμβάνεται το 81% της συνολικής εσωτερικής μνήμης του FPGA, ενώ στη δεύτερη

Ακριβής υλοποίηση περίπτωση το 75%. Σε περίπτωση που απαιτείται η επεξεργασία μεγαλύτερων σε μέγεθος εικόνων, τότε είναι αναγκαία η χρήση εξωτερικής μνήμης η οποία θα λειτουργεί σαν FIFO για την επίτευξη της δεύτερης σάρωσης των πλαισίων. Αυτό φυσικά εισάγει σημαντικά προβλήματα, όπως αύξηση του κόστους και εξάρτηση από την ταχύτητα μετάδοσης της εξωτερικής μνήμης. Οι πιθανές μνήμες που μπορούν να χρησιμοποιηθούν για την περίπτωση αυτή, σύμφωνα με την ALTERA, περιγράφονται στην ιστοσελίδα (ALTERA, 2007). Ενδεικτικά αναφέρεται ότι η χρήση εξωτερικής μνήμης DDR, η οποία είναι και η απλούστερη από τον κατάλογο που παρατίθεται, επιτρέπει την επεξεργασία έγχρωμων εικόνων μεγέθους 640×480 εικονοστοιχείων σε πραγματικό χρόνο.

Εναλλακτική υλοποίηση

Αντίθετα με τα παραπάνω, η εναλλακτική υλοποίηση που παρουσιάστηκε παρέχει τη δυνατότητα για επεξεργασία έγχρωμων εικόνων μεγέθους έως και 2,5 εκατομμυρίων εικονοστοιχείων σε πραγματικό χρόνο. Όπως είναι αναμενόμενο, η μείωση ενός από τα τρία παραπάνω χαρακτηριστικά μπορεί να αυξήσει τα υπόλοιπα δύο. Έτσι, το μέγεθος της εικόνας μπορεί να αυξηθεί εάν η επεξεργασία γίνεται για εικόνες σε αποχρώσεις του γκρι ή αν μειωθεί ο ρυθμός ανανέωσης. Οι απαιτήσεις σε μνήμη της υλοποίησης αυτής είναι οι μικρότερες και καλύπτουν το 39% της συνολικής εσωτερικής μνήμης της συσκευής.

Η μέγιστη καθυστέρηση και για τις δύο υλοποιήσεις εμφανίστηκε στους παράλληλους αθροιστές οι οποίοι υπολογίζουν την περιφέρεια του εκάστοτε εικονοστοιχείου σε τρεις κλίμακες. Το μικρότερο δυνατό ρολόι συστήματος που παρουσίασε το λογισμικό Quartus II, ώστε οι παράλληλοι αθροιστές να λειτουργούν χωρίς προβλήματα, ήταν στα 13.6ns. Τέτοια περίοδος ρολογιού αντιστοιχεί σε συχνότητα 73,52MHz. Εντούτοις, για λόγους ασφάλειας το ρολόι του συστήματος επιλέχθηκε να έχει περίοδο 15ns, το οποίο αντιστοιχεί σε συχνότητα 66,66MHz.

5.7 Συμπεράσματα

Στο κεφάλαιο αυτό παρουσιάστηκαν δύο υλοποιήσεις σε FPGA της μεθόδου βελτίωσης αποχρώσεων (κεφαλαίο 3). Η προτεινόμενες υλοποιήσεις βασίζονται σε μια συσκευή Stratix II GX: EP2SGX130GF1508C5 της ALTERA και έχουν μέγιστη συχνότητα λειτουργίας 66.66MHz. Η πρώτη υλοποίηση επιτρέπει την επεξεργασία σε πραγματικό χρόνο έγχρωμων εικόνων μεγέθους 400×400 εικονοστοιχείων, είτε εικόνων σε αποχρώσεις του γκρι μεγέθους 640×480 εικονοστοιχείων. Μεγαλύτερου μεγέθους εικόνες επιβάλουν την ύπαρξη εξωτερικής μνήμης. Η δεύτερη υλοποίηση έχει μικρότερες απαιτήσεις σε μνήμη και μπορεί να επεξεργαστεί έγχρωμες εικόνες μεγέθους 2.5 εκατομμυρίων εικονοστοιχείων, με ρυθμό ανανέωσης 25 πλαίσια/δευτερόλεπτο. Οι δύο υλοποιήσεις υπερτερούν σε δυνατότητες από άλλες υπάρχουσες υλοποιήσεις παρόμοιων αλγορίθμων, αφού επεξεργάζονται έγχρωμες εικόνες με πολλαπλάσια ανάλυση.

- Συμπεράσματα διατριβής Προτάσεις για μελλοντική έρευνα



6

Συμπεράσματα & Προτάσεις για Μελλοντική Έρευνα

Το κεφάλαιο αυτό συνοψίζει τα συμπεράσματα που προέκυψαν στην παρούσα διδακτορική διατριβή. Συγκεκριμένα, περιγράφονται συνοπτικά τα πεδία στα οποία εστίασε η διατριβή σε κάθε κεφάλαιο, τα κυριότερα αποτελέσματα που εξήχθησαν, καθώς και η συνεισφορά τους στους τομείς της επιστήμης. Επίσης, παρουσιάζονται προτάσεις για μελλοντική έρευνα, ξεχωριστά για κάθε πεδίο με το οποίο πραγματεύτηκε η παρούσα διατριβή, οι οποίες μπορούν να αποτελέσουν εξέλιξη για τις μεθόδους και τα αποτελέσματα που αναπτύχθηκαν στα προηγούμενα κεφάλαια.

6.1 Συμπεράσματα διατριβής

Τα προβλήματα, που καλούνται να αντιμετωπίσουν τα τεχνητά και τα βιολογικά συστήματα όρασης, είναι στη βάση τους ίδια. Κατά συνέπεια, πολλά από τα ζητήματα, τα οποία αποτελούν αντικείμενο έρευνας στο πεδίο της όρασης μηχανών και της επεξεργασίας εικόνας, έχουν ήδη αντιμετωπιστεί από την εξέλιξη. Βασικός στόχος της παρούσας διατριβής ήταν η διερεύνηση και ανάπτυξη νέων τεχνικών στα δύο αυτά ερευνητικά πεδία, οι οποίες έχουν εμπνευστεί από τον τρόπο λειτουργίας του οπτικού συστήματος των πρωτευόντων θηλαστικών.

Πιο συγκεκριμένα, παρουσιάστηκε μια νέα μέθοδος για την εξαγωγή εξεχόντων περιγραμμάτων, βασισμένη στο είδος διασύνδεσης των νευρώνων κλίσης του πρωτοταγούς οπτικού φλοιού. Στα πλαίσια αυτής της μεθόδου, σχεδιάστηκε ένα καινούριο σύνολο μασκών για τη μοντελοποίηση των νευρώνων κλίσης, το οποίο βελτιώνει κατά 20 φορές τη μέχρι τώρα υπάρχουσα προσέγγιση, η οποία βασίζεται στη συνέλιξη με φίλτρα Gabor. Επιπλέον, ο καινούριος τρόπος υπολογισμού της κλίσης των ακμών ευνοεί την παραλληλία και ταιριάζει περισσότερο στο διακριτό πλέγμα υπερστηλών του πρωτοταγούς οπτικού φλοιού. Η προτεινόμενη μέθοδος εξαγωγής εξεχόντων περιγραμμάτων βασίζεται σε μια νέα μορφή νευρωνικού δικτύου. Το δίκτυο αυτό δέχεται ως εισόδους τα αποτελέσματα των μασκών κλίσης σε κάθε σημείο της εικόνας, ενώ με την πάροδο των επαναλήψεων αυξάνει τις τιμές εκείνων των μασκών που κωδικοποιούν εξέχοντα περιγράμματα. Για την αλληλεπίδραση των μασκών κλίσης, χρησιμοποιήθηκαν κανόνες οι οποίοι ευνοούν το σχηματισμό εκθετικών καμπυλών. Η προσέγγιση αυτή διαπιστώθηκε ότι δίνει βελτιωμένα αποτελέσματα σε σχέση με τη μέχρι τώρα προσέγγιση, η οποία βασίζεται στο σχηματισμό κυκλικών καμπυλών. Παρουσιάστηκε επίσης μια καινούρια συνάρτηση διέγερσης των νευρώνων του δικτύου, η οποία βασίζεται στην ομοιότητα της διέγερσης των δύο λοβών τους. Χαρακτηριστικό της νέας αυτής συνάρτησης είναι ότι ευνοεί τη σωστή συνέχεια των εξεχόντων περιγραμμάτων. Το προτεινόμενο δίκτυο έχει τη δυνατότητα να τονίζει συγκεκριμένους βαθμούς καμπυλότητας και κατά συνέπεια, μπορεί να χρησιμοποιηθεί για το φιλτράρισμα των εξεχόντων περιγραμμάτων, ώστε να αναδειχθεί μια μόνο κατηγορία, όπως ευθείες ή κύκλοι. Τέλος, η μέθοδος παρουσιάζει ικανοποιητικά αποτελέσματα για μια ευρεία γκάμα πραγματικών και συνθετικών εικόνων, σε συνδυασμό πάντα με τους συντομότερους χρόνους εκτέλεσης που έχουν αναφερθεί στην υπάρχουσα βιβλιογραφία.

Μια άλλη πτυχή της έρευνας που, διερευνήθηκε κατά τη διάρκεια της διατριβής, ήταν και η βελτίωση των αποχρώσεων εικόνων υποβαθμισμένης ποιότητας, λόγω του υψηλού δυναμικού εύρους της σκηνής από την οποία προήλθαν. Στα πλαίσια της έρευνας αυτής, αναπτύχθηκε και προτάθηκε μια νέα μέθοδος για τη βελτίωση εκείνων των περιοχών των εικόνων που έχουν υποστεί υποέκθεση ή υπερέκθεση. Κύριος στόχος της μεθόδου είναι να βελτιώσει τα τμήματα που παρουσιάζουν το συγκεκριμένο πρόβλημα, χωρίς να επηρεάσει τις σωστά εκτεθειμένες περιοχές. Στα πλαίσια αυτά, παρουσιάστηκε μια νέα συνάρτηση απεικόνισης αποχρώσεων, βασισμένη στη διακλαδισμένη αναστολή των δικτύων ανταγωνισμού κέντρου-περιφέρειας του ΑΟΣ. Η συνάρτηση αυτή χρησιμοποιεί ρυθμιζόμενες μη γραμμικότητες, οι οποίες μεταβάλλουν την τιμή έντασης κάθε εικονοστοιχείου, βάσει της ευρύτερης περιφέρειάς του και των γενικότερων χαρακτηριστικών της εικόνας. Παρουσιάστηκε επίσης ένας καινούριος τρόπος για τον υπολογισμό της περιφέρειας κάθε εικονοστοιχείου, ο οποίος βασίζεται σε μια περιοχή η οποία προσαρμόζει το σχήμα της στις τοπικές ακμές της εικόνας. Η προσαρμοστικότητα του σχήματος αυτής της περιοχής βασίζεται στα κύτταρα κλίσης του πρωτοταγούς οπτικού φλοιού. Η προτεινόμενη μέθοδος παρουσιάζει βελτιωμένα αποτελέσματα σε σχέση με αυτά της υπάρχουσας βιβλιογραφίας. Πιο συγκεκριμένα, σε αντίθεση με τις υπάρχουσες μεθόδους, δεν εξάγονται φαινόμενα άλω στις ισχυρές μεταβάσεις εντάσεως, οι σωστά εκτεθειμένες περιοχές δεν επηρεάζονται, ενώ βελτιώνονται μόνο οι υποεκτεθειμένες περιοχές, αλλά και οι υπερεκτεθειμένες. Ο χρόνος εκτέλεσης της προτεινόμενης μεθόδου είναι τουλάχιστο μια τάξη μεγέθους μικρότερος από τον αντίστοιχο των άλλων μεθόδων, ενώ ψυχοφυσιολογική μελέτη που έγινε σε ανθρώπινους παρατηρητές ανέδειξε τα αποτελέσματα της προτεινόμενης μεθόδου ως τα πιο «φυσικά» σε σχέση με τα υπόλοιπα. Το πεδίο εφαρμογής της εκτείνεται στην όραση μηχανών, σε ιατρικές εφαρμογές (π.χ. ψηφιακές ακτινογραφίες, τομογραφία κ.τ.λ.) έως την αισθητική βελτίωση φωτογραφιών.

Η χρήση χαρακτηριστικών στοιχείων των βιολογικών οπτικών συστημάτων επεκτάθηκε και στο πεδίο της δυαδικοποίησης εγγράφων. Πιο συγκεκριμένα, παρουσιάστηκε μια νέα μέθοδος δυαδικοποίησης εγγράφων, η οποία υιοθετεί τους βιολογικούς εκείνους μηχανισμούς που αποτελούν τη βάση αντίληψης του σκοτεινού και του φωτεινού. Οι βιολογικοί αυτοί μηχανισμοί είναι τα γαγγλιακά κύτταρα σκοτεινού κέντρου του ΑΟΣ. Η προτεινόμενη μέθοδος χρησιμοποιεί δομές ανταγωνισμού κέντρου-περιφέρειας ανάλογες με τα γαγγλιακά κύτταρα αυτού του τύπου. Εισάγεται μια καινούρια συνάρτηση απόκρισης των κυττάρων αυτών η οποία αυξάνει την ευαισθησία του κυττάρου στις σκοτεινές περιοχές, μειώνοντας έτσι τις επιπτώσεις των σκιάσεων και των λεκέδων. Έτσι, η εφαρμογή αυτών των κυττάρων στην εικόνα τη μετατρέπει σε μια ενδιάμεση μορφή ικανή να δυαδικοποιηθεί με επιτυχία από τις κλασικές μεθόδους. Παράλληλα, υιοθετείται ο συνδυασμός διαφορετικών χωρικών κλιμάκων επεξεργασίας με τη χρήση κυττάρων δύο διαφορετικών μεγεθών, τα οποία επεξεργάζονται πληροφορίες διαφορετικής συχνότητας το καθένα. Τα πειραματικά αποτελέσματα έδειξαν ότι η προτεινόμενη μέθοδος υπερτερεί των ήδη υπαρχόντων της βιβλιογραφίας, ειδικότερα για έγγραφα με σημαντική υποβάθμιση. Αυτό συμβαίνει γιατί παρουσιάζει πολύ καλύτερη απόκριση στο θόρυβο και στις σκιάσεις. Για το λόγο αυτό, μπορεί να χρησιμοποιηθεί ιδιαίτερα σε ιστορικά έγγραφα και έγγραφα τα οποία ψηφιοποιήθηκαν με φωτογραφική μηχανή.

Τέλος, παρουσιάστηκαν δύο υλοποιήσεις σε FPGA της μεθόδου βελτίωσης αποχρώσεων. Η πρώτη αναπαράγει πιστά τον αρχικό αλγόριθμο και επιτρέπει την επεξεργασία σε πραγματικό χρόνο έγχρωμων εικόνων μεγέθους 400×400 εικονοστοιχείων, είτε εικόνων σε αποχρώσεις του γκρι μεγέθους 640×480 εικονοστοιχείων. Μεγαλύτερου μεγέθους εικόνες επιβάλουν την ύπαρξη εξωτερικής μνήμης. Η δεύτερη υλοποίηση χρησιμοποιεί τα χαρακτηριστικά προηγούμενων πλαισίων για τη βελτίωση των αποχρώσεων του εκάστοτε πλαισίου. Η υλοποίηση αυτή έχει μικρότερες απαιτήσεις σε μνήμη και επιτρέπει την επεξεργασία σε πραγματικό χρόνο έγχρωμων εικόνων μεγέθους εικόνων μεγέθους στου γκρι μεγέθους στου για τη βελτίωση των αποχρώσεων του εκάστοτε πλαισίου. Η υλοποίηση αυτή έχει μικρότερες απαιτήσεις σε μνήμη και επιτρέπει την επεξεργασία σε πραγματικό χρόνο έγχρωμων εικόνων μεγέθους έως 2,5 εκατομμυρίων εικονοστοιχείων. Η συσκευή που χρησιμοποιήθηκε και στις δύο υλοποιήσεις είναι η Stratix

II GX: EP2SGX130GF1508C5 της ALTERA. Οι επιδόσεις των δύο κυκλωμάτων υπερτερούν σε σχέση με τις υπάρχουσες υλοποιήσεις ανάλογων αλγορίθμων, αφού επεξεργάζεται εικόνες με πολλαπλάσια ανάλυση και χρώμα, σε αντίθεση με τις υπάρχουσες οι οποίες περιορίζονται στη βελτίωση αποχρώσεων του γκρι.

6.2 Προτάσεις για μελλοντική έρευνα

Τα χαρακτηριστικά των βιολογικών οπτικών συστημάτων, τα οποία μπορούν να υιοθετηθούν στο πεδίο της όρασης μηχανών και της επεξεργασίας εικόνας, είναι στην παρούσα χρονική περίοδο ακόμη πολλά. Μερικά από αυτά τα χαρακτηριστικά παρουσιάζονται παρακάτω και μπορούν να αποτελέσουν προτάσεις για μελλοντική έρευνα και εξέλιξη της παρούσας διατριβής.

Η μέθοδος εξαγωγής εξεχόντων περιγραμμάτων μπορεί να εξελιχθεί ως εξής:

Έχει βρεθεί ότι το ΑΟΣ εφαρμόζει τους αλγορίθμους επεξεργασίας του ταυτόχρονα σε πολλές χωρικές κλίμακες. Η προτεινόμενη μέθοδος μπορεί να επεκταθεί σε αυτή την κατεύθυνση, χρησιμοποιώντας γκαουσιανή πυραμίδα. Πιο συγκεκριμένα, από την αρχική εικόνα μπορούν να προκύψουν με διαδοχικές σμικρύνσεις πολλές μικρότερες εικόνες, κάθε μια έχοντας χαρακτηριστικά διαφορετικής κλίμακας. Σε κάθε εικόνα θα εφαρμόζεται η προτεινόμενη μέθοδος. Εκτός από τη διέγερση και τον ανταγωνισμό μεταξύ της ίδιας εικόνας, που ήδη υπάρχει, θα υπάρχουν αντίστοιχα χαρακτηριστικά διέγερσης και ανταγωνισμού μεταξύ των μασκών διαφορετικών χωρικών κλιμάκων.

Για τη μέθοδο βελτίωσης αποχρώσεων μπορούν να γίνουν τα ακόλουθα:

- Να γίνει επέκταση της μεθόδου για βελτίωση της τοπικής αντίθεσης της εικόνας.
 Αυτό μπορεί να επιτευχθεί χρησιμοποιώντας τις ρυθμιζόμενες μη γραμμικότητες για έλεγχο της διαφοράς μεταξύ κέντρου και περιφέρειας στα κύτταρα πλευρικού ανταγωνισμού.
- Η μέθοδος να επεκταθεί και στη διόρθωση χρώματος, κάνοντας κατάλληλες τροποποιήσεις, ώστε να μπορεί να εφαρμοστεί ταυτόχρονα και στα τρία χρωματικά κανάλια. Επίσης, μπορούν να υιοθετηθούν νευροφυσιολογικά χαρακτηριστικά της περιοχής V4 του οπτικού φλοιού, τα οποία έχουν συνδεθεί με τη χρωματική σταθερότητα του ΑΟΣ.

Η μέθοδος δυαδικοποίησης μπορεί να επεκταθεί χρησιμοποιώντας τα ακόλουθα στοιχεία:

Όλες σχεδόν οι οπτικές περιοχές του εγκεφάλου δέχονται ισχυρά σήματα ανάδρασης από ανώτερες περιοχές επεξεργασίας. Το ίδιο μπορεί να γίνει και στη δυαδικοποίηση εγγράφων. Εκτελώντας ταυτόχρονα με τη δυαδικοποίηση και οπτική αναγνώριση χαρακτήρων, το αποτέλεσμα της αρχικής ταξινόμησης μπορεί να χρησιμοποιηθεί σαν ανάδραση για τη βελτίωση της διαδικασίας δυαδικοποίησης του αρχικού σήματος. Έτσι, μπορεί να αυξάνεται αυτόματα η ευαισθησία της προτεινόμενης μεθόδου σε περιοχές που υπάρχει μεγάλη

πιθανότητα για ύπαρξη χαρακτήρων, ενώ να μειώνεται σε περιοχές που δεν αναγνωρίζονται χαρακτήρες.

Για την υλοποίηση σε FPGA μπορούν να γίνουν τα ακόλουθα:

- Να γίνουν κατάλληλες τροποποιήσεις για την επίτευξη χαμηλής κατανάλωσης ισχύος, κάτι που θα ήταν ιδιαίτερα χρήσιμο για συσκευές εμπορικού ενδιαφέροντος, οι οποίες αναγκαστικά βασίζονται στη χρήση μπαταριών.
- Να περιοριστεί το μέγεθος του τελικού πίνακα αντιστοίχισης, ο οποίος καταλαμβάνει 256KB μνήμης. Μια πιθανή προσέγγιση θα ήταν η περαιτέρω μείωση της ακρίβειας ορισμένων παραμέτρων, η οποία όμως θα συνοδεύεται από ένα επιπλέον στάδιο παρεμβολής (interpolation). Σε περίπτωση, για παράδειγμα, που προκύψει η ανάγκη για εύρεση μιας τιμής που δεν είναι καταχωρημένη στον πίνακα αντιστοίχισης, το στάδιο αυτό θα πραγματοποιεί παρεμβολή μεταξύ των γειτονικών καταγεγραμμένων τιμών, ώστε να προκύψει η καινούρια τιμή. Μια τέτοια μείωση του μεγέθους του πίνακα αντιστοίχισης σε συσκευή μικρότερων δυνατοτήτων (π.χ. Cyclone II) και κατά συνέπεια θα οδηγούσε στη μείωση του κόστους.

Βιβλιογραφία

ACE - an Automatic Color Equalization algorithm. (n.d.). Retrieved September 2007, from Laboratorio di Eidomatica: http://eidomatica.dico.unimi.it/eng/research/ace.html

Alleysson, D., Süsstrunk, S., & Hérault, J. (2005). Linear Demosaicing inspired by the Human Visual System. *IEEE Transactions on Image Processing*, 14 (4), 439-449.

Alter, T. D., & Basri, R. (1998). Extracting Salient Curves from Images: An Analysis of the Saliency Network. *International Journal of Computer Vision, 27* (1), 51-69.

ALTERA. (2007). *Altera's External Memory Solutions Center*. Retrieved November 2007, from ALTERA: http://www.altera.com/technology/memory/mem-index.jsp

Balercia, T., Zitti, A., Francesconi, H., Orcioni, S., & Conti, M. (2006). FPGA Implementations of a Simplified Retinex Image Processing Algorithm. *13th IEEE International Conference on Electronics, Circuits and Systems ICECS '06.* (pp. 176-179). Nice, France: IEEE.

Battiato, S., Castorina, A., & Mancuso, M. (2003). High dynamic range imaging for digital still camera: an overview. *Journal of Electronic Imaging*, *12*, 459-469.

Bayer, B. (1976). Patent No. 3,971,065. US.

Benedetti, A., Prati, A., & Scarabottolo, N. (1998). Image convolution on FPGAs: the implementation of a multi-FPGA FIFO structure. *24th Euromicro Conference*, *1*, pp. 123-130. Vasteras, Sweden.

Ben-Shahar, O., Huggins, P. S., Izo, T., & Zucker, S. W. (2003). Cortical Connections and Early Visual Function: Intra- and Inter-Columnar Processing. *Journal of Physiology Paris, 97*, 191-208.

Bloomfield, S. (1974). Arithmetical operations performed by nerve cells. *Brain Research, 69,* 115-124.

Bovik, A. (2004). Handbook of Image and Video Processing. Academic Press.

Braun, J. (1999). On the Detection of Salient Contours. Spatial Vision, 12 (2), 211-225.

Čadík, M., Wimmer, M., Neumann, L., & Artusi, A. (2006). Image Attributes and Quality for Evaluation of Tone Mapping Operators. *14th Pacific Conference on Computer Graphics and Applications* (pp. 35-44). Pacific Graphics 2006.

Calkins, D. (2001). Seeing with S cones. *Progress in Retinal and Eye Research, 20* (3), 255-287.

Choe, Y., & Miikkulainen, R. (2004). Contour Integration and Segmentation with Selforganized Lateral Connections. *Biological Cybernetics*, *90*, 75-88.

CIE. (1986). *Publication CIE No 15.2, Colorimetry* (Vols. Publication CIE No 15.2,). Vienna: Central Bureau of the Commission Internationale de L'Éclairage.

Croner, L., & Kaplan, E. (1995). Receptive Fields of P and M Ganglion Cells Across the Primate Retina. *Vision Research*, *35* (1), 7-24.

Dacey, D. (1999). Primate Retina: Cell Types, Circuits and Color Opponenc. *Progress in Retinal and Eye Research*, *18* (6), 737-763.

Dacey, D. (1993). The Mosaic of Midget Ganglion Cells in the Human Retina. *Journal of Neuroscience*, *13* (12), 5334-4355.

Debevec, P. E., & Malik, J. (1997). Recovering high dynamic range radiance maps from photographs. *SIGGRAPH'97*, (pp. 369-378).

Devlin, K. (2002). *A review of tone reproduction techniques.* University of Bristol, CSTR-02-005.

Drago, F., Martens, W., & Chiba, N. (2003). Design of a tone mapping operator for high dynamic range images based upon psychophysical evaluation and performance mapping. In P. I. Imaging (Ed.), *The Human Vision and Electronic Imaging VIII*, (pp. 321-331). San Jose, USA.

Field, D. J., & Hayes, A. (2004). Contour Integration and the Lateral Connections of V1 Neurons. In L. M. Chalupa, & J. S. Werner, *The Visual Neurosciences* (pp. 1069-1079). MIT Press.

Field, D. J., Hayes, A., & Hess, R. F. (1993). Contour Integration by the Human Visual System: Evidence for a local "Association Field". *Vision Research*, *33*, 173-193.

FineReader. (n.d.). Retrieved 2007, from Abbyy: http://www.abbyy.com/finereader_ocr/

Fiorentini, A. (2004). Brightness and Lightness. In L. Chalupa, & J. Werner, *The Visual Neurosciences* (Vol. 2, pp. 881-891). MIT Press.

Fiorentini, A., Baumgartner, G., Magnussen, S., Schiller, P. H., & Thomas, J. P. (1990). The Perception of Brightness and Darkness: Relations to Neuronal Receptive Fields. In L. Spillmann, & J. S. Werner, *Visual Perception: The Neurophysiological Foundations* (pp. 129-161). Academic Press.

Fitzpatrick, D. (1996). The Functional-Organization of Local Circuits in Visual Cortex - Insights from the study of tree shrew striate cortex. *Cerebral Cortex, 6*, 329-341.

Gatos, B., Pratikakis, I., & Perantonis, S. (2006). Adaptive degraded document image binarization. *Pattern Recognition*, *39*, 317-327.

Gatta, C., Rizzi, A., & Marini, D. (2006). Local linear LUT method for spatial colour-correction algorithm speed-up. *IEE Proceedings Vision Image Signal Processing*, *153* (3), 357-363.

Geisler, W. S., Perry, J. S., Super, B. J., & Gallogly, D. P. (2001). Edge Co-occurrence in Natural Images Predicts Contour Grouping Performance. *Vision Research*, *41*, 711-724.

Gilbert, C. D., Das, A., Ito, M., Kapadia, M., & Westheimer, G. (1996). Spatial Integration and Cortical Dynamics. *Proceedings of the National Academy of Sciences USA*, *93*, 615-622.

Gilchrist, A., Kossyfidis, C., Bonato, F., Agostini, T., Cataliotti, J., Li, X., et al. (1999). An Anchoring Theory of Lightness Perception. *Psychological Review*, *106* (4), 795-834.

Goshtasby, A. A. (2005). Fusion of multi-exposure images. *Image and Vision Computing, 23,* 611-618.

Grigorescu, C., Petkov, N., & Westenberg, M. A. (2003). Contour Detection Based on Nonclassical Receptive Field Inhibition. *IEEE Transactions on Image Processing*, *12*, 729-739.

Grossberg, S. (1994). 3-d Vision and Figure-Ground Separation by Visual Cortex. *Perception and Psychophysics*, 55, 48-121.

Grossberg, S. (2004). Visual Boundaries and Surfaces. Στο L. M. Chalupa, & J. S. Werner, *The Visual Neurosciences* (σσ. 1624-1639). MIT Press.

Grossberg, S., & Mingolla, E. (1985). Neural Dynamics of Perceptual Grouping: textures, boundaries, and emergent segmentations. *Perceptual Psychophysics, 38*, 141-171.

Grossberg, S., & Todorovic, D. (1988). Neural dynamics of 1-d and 2-d brightness perception: A unified model of classical and recent phenomena. *Perception & Psychophysics, 43*, 241-277.

Grossberg, S., Mingolla, E., & Williamson, J. (1995). Synthetic Aperture Radar Processing by a Multiple Scale Neural System for Boundary and Surface Representation. *Neural Networks*, *8*, 1005-1028.

Gupta, M., Jacobson, N., & Garcia, E. (2007). OCR binarization and image pre–processing for searching historical documents. *Pattern Recognition, 40*, 389–397.

Helmholtz, H. (1866). *Helmholtz's Treatise on Physiological Optics*. New York: Optical Society of America.

Herault, L., & Horaud, R. (1993). Figure-Ground Discrimination: A combinatorial optimisation approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *15*, 899-914.

Hering, E. (1874, 1964). *Outlines of a Theory of the Light Sense.* Cambridge, MA: Harvard University Press.

Hess, R. F., Hayes, A., & Field, D. J. (2003). Contour Integration and Cortical Processing. *Journal of Physiology Paris*, *97*, 105-119.

Hines, G., Rahman, Z., Jobson, D., & Woodell, G. (2004). DSP Implementation of the Retinex Image Enhancement Algorithm. *Visual Information Processing XIII, Proc. SPIE 5438.* SPIE. Hines, G., Rahman, Z., Jobson, D., & Woodell, G. (2004). Single-Scale Retinex Using Digital Signal Processors. *Global Signal Processing Expo (GSPx)*.

Hubel, D. H., & Wiesel, T. N. (1965). Receptive Fields, Binocular Interaction and Functional Architecture in Nonstriate Areas (188 and 19) of the Cat. *Journal of Neurophysiology, 28*, 229-289.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive Fields, Binocular Interaction and Functional Architecture of Monkey Striate Cortex. *Journal of Physiology*, *195*, 215-243.

Jobson, D. J., Rahman, Z.-U., & Woodell, G. A. (1997). A Multiscale Retinex for Bridging the Gap Between Color Images and the Human Observation of Scenes. *IEEE Transactions on Image Processing*, *6* (7), 965-976.

Jobson, D. J., Rahman, Z.-U., & Woodell, G. A. (1997). Properties and performance of a center/surround Retinex. *IEEE Transactions on Image Processing*, *6*, 451-462.

Jones, J. P., & Palmer, L. A. (n.d.). An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex.

Jones, J. P., & Palmer, L. A. (1987). An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, *58*, 1233–1258.

Kapadia, M. K., Ito, M., Gilbert, C. D., & Westheimer, G. (1995). Improvement in Visual Sensitivity by Changes in Local Context: Parallel studies in human observers and in V1 of alert monkeys. *Neuron*, *15*, 843-856.

Kapadia, M. K., Westheimer, G., & Gilbert, C. D. (2000). Spatial Distribution of Contextual Interactions in Primary Visual Cortex and in Visual Perception. *Neurophysiology, 84*, 2048-2062.

Kovács, I., & Julesz, B. (1993). A Closed Curve is Much More than an Incomplete One: Effect of closure in figure-ground segmentation. *Proceedings of National Academy of Sciences USA*, 90, 7495-7497.

Kovács, I., & Julesz, B. (1994). Perceptual Sensitivity Maps within Globally Defined Visual Shapes. *Nature*, *370*, 644-646.

Kuang, J., Yamaguchi, H., Johnson, G., & Fairchild, M. (2004). Testing HDR image rendering algorithms. In P. IS&T/SID (Ed.), *Twelfth Color Imaging Conference: Color Science, Systems, and Application,*, (pp. 315-320). Scottsdale,.

Land, E. (1986). An alternative technique for the computation of the designator in the Retinex theory of color vision. *Proceedings of National Academy of Sciences of the United States of America*, *83*, 3078-3080.

Land, E. (1964). The Retinex. 52 (2), 247-264.

Land, E. (1977). The retinex theory of color vision. *Scientific American, 237*, 108-128.

Ledda, P., Chalmers, A., Troscianko, T., & Seetzen, H. (2005). Evaluation of tone mapping operators using high dynamic range display. *Annual Conference on Computer Graphics* (pp. 640-648). Los Angeles, USA: Proc ACM SIGGRAPH.

Lee, B. (1996). Receptive Field Structure in the Primate Retina. *Vision Research, 36* (5), 631-644.

Li, Z. (1998). A Neural Model of Contour Integration in the Primary Visual Cortex. *Neural Computation*, *10*, 903-940.

Mahamud, S. (1999). *Results*. Ανάκτηση 2007, από Segmenting Multiple Closed Contours: http://www-2.cs.cmu.edu/~mahamud/grouping/grouping.html

Mahamud, S., Williamns, L. R., Thornber, K. K., & Xu, K. (2003). Segmentation of Multiple Salient Closed Contours from Real Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *25*, 433-444.

Mantiuk, G., Krawczyk, K., & Myszkowski, S. (2004). Perception Motivated High Dynamic Range Video Encoding. *SIGGRAPH 2004*, (pp. 733-741).

Marr, D. (1982). Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. San Francisco: Freeman.

Meylan, L. (2006). Tone Mapping for High Dynamic Range Images. EPFLThesis No. 3588.

Meylan, L., & Süsstrunk, S. (2006). High Dynamic Range Image Rendering With a Retinex-Based Adaptive Filter. *IEEE Transactions on Image Processing*, *15* (9), 2820-2830.

Meylan, L., Alleysson, D., & Susstrunk, S. (2007). A Model of Retinal Local Adaptation for the Tone Mapping of Color Filter Array Images. *Journal of the Optical Society of America A, 24* (9), 2807-2816.

Mundhenk, N. (2005). *CINNIC images*. Retrieved 2007, from CINNIC: http://cinnic.org/images/12_orientations_jpg/index.html

Mundhenk, T. N., & Itti, L. (2003). CINNIC, a New Computational Algorithm for the Modeling of Early Visual Contour Integration in Humans. *Neurocomputing*, *52-54*, 599-604.

Mundhenk, T. N., & Itti, L. (2005). Computational Modelling and Exploration of Contour Integration for Visual Saliency. *Biological Cybernetics*, *93*, 188-212.

Nelson, R., & Kolb, H. (2004). ON and OFF Pathways in the Vertebrate Retina and Visual System. In L. Chalupa, & J. Werner, *The Visual Neurosciences*, (Vol. 1, pp. 260–278). MIT Press.

Neural and Computational Models of Vision. (n.d.). Retrieved 2007, from Boston University: http://cns.bu.edu/~ennio/cn530-f-2006/index.html

Newsome, W. T., Britten, K. H., & Movshon, J. A. (1989). Neuronal Correlates of a Perceptual Decision. *Nature*, *341*, 52-54.

Niblack, W. (1986). *An Introduction to Digital Image Processing*. Englewood Cliffs, NJ: Prentice-Hall.

Otsu, N. (1979). A thresholding selection method from gray–scale histogram. *IEEE Transactions on Systems, Man and Cybernetics, 9,* 62–66.

Paradiso, M., & Nakayama, K. (1991). Brightness Perception and Filling-in. *Vision Research,* 31 (7), 1221-1236.

Parent, P., & Zucker, S. W. (1989). Trace Inference, Curvature Consistency, and Curve Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *11*, 823-839.

Pettet, M. W., McKee, S. P., & Grzywacz, N. M. (1998). Constrains on Long-Range Interactions Mediating Contour-Detection. *Vision Research*, *38*, 865-879.

Polat, U. K., Mizobe, K., Pettet, M. W., Kasamatsu, T., & Norcia, A. M. (1998). Collinear Stimuli Regulate Visual Responses Depending on Cell's Contrast Threshold. *Nature, 391*, 580-584.

Poynton, C. (1996). A Technical Introduction to Digital Video. John Wiley & Sons.

Provenzi, E., Fierro, M., Rizzi, A., Carli, L. D., Gadia, D., & Marini, D. (2007). Random Spray Retinex: A New Retinex Implementation to Investigate the Local Properties of the Model. *IEEE Transactions on Image Processing*, *16* (1), 162-171.

Purves, D., Williams, M. S., Nundy, S., & Lotto, B. R. (2004). Perceiving the Intensity of Light. *Psychological Review*, *111* (1), 142-158.

Qiu, G., Duan, J., & Finlayson, G. D. (2007). Learning to display high dynamic range images. *Pattern Recognition*, 40, 2641-2655.

Reinhard, E., & Devlin, K. (2005). Dynamic Range Reduction Inspired by Photoreceptor Physiology. *IEEE Transactions on Visualization and Computer Graphics, 11* (1), 13-24.

Retinex Image Processing. (n.d.). Retrieved 2007, from NASA: http://dragon.larc.nasa.gov/

Rizzi, A., Gatta, C., & Marini, D. (2003). A new algorithm for unsupervised global and local color correction. *Pattern Recognition Letters, 24*, 1663-1677.

Rizzi, A., Gatta, C., & Marini, D. (2004). From Retinex to Automatic Color Equalization: issues in developing a new algorithm for unsupervised color equalization. *Journal of Electronic Imaging*, *13* (1), 75-84.

Rockland, K. S., & Lund, J. S. (1983). Intrinsic Laminar Lattice Connections in Primate Visual Cortex. *Journal of Comparative Neurology*, *216*, 303-318.

Rockland, K. S., & Lund, J. S. (1982). Widespread Periodic Intrinsic Connections in the Tree Shrew Visual Cortex. *Science*, *215*, 532-534.

Rodieck, R. (1965). Quantitative analysis of cat retinal ganglion cell response to visual stimuli. *Vision Research*, *5*, 583-601.

Rodieck, R. (1998). The First Steps in Seeing. Sinauer Associates.

Ross, W. D., & Mingolla, E. (1998). Recent Progress in Modeling Neural Mechanisms of Form and Color Vision. *Image and Vision Computing*, *16*, 447-472.

Sato, K. (2006). Image-Processing Algorithms. In J. Nakamura, *Image Sensors and Signal Processing for Digital Still Cameras*. CRC Press.

Sauvola, J., & Pietikainen, M. (2000). Adaptive document image binarization. *Pattern Recognition*, *33*, 225–236.

Sezgin, M., & Sankur, B. (2004). Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electron Imaging*, *13* (1), 146–165.

Sha'ashua, A., & Ullman, S. (1988). Structural saliency: The detection of globally salient structures using a locally connected network. *2nd ICCV*, (pp. 321–327).

Sincich, L. C., & Blasdel, G. G. (2001). Oriented Axon Projections in Primary Visual Cortex of the Monkey. *Journal of Neuroscience*, *21*, 4416-4426.

Singer, W., & Gray, C. M. (1995). Visual Feature Integration and the Temporal Correlation Hypothesis. *Annual Review of Neuroscience*, *18*, 555-586.

Stockman, A., MacLeod, D., & Johnson, N. (1993). Spectral sensitivities of the human cones. *Journal of the Optical Society of America, 10,* 2491-2521.

Swindale, N. (1996, July). Visual cortex: Looking into a Klein bottle. *Current Biology, 6* (7), 776-779.

TruView Imaging Company. (n.d.). Retrieved 2007, from http://www.truview.com/

Wandell, B. (1995). Foundations of Vision. Sunderland, MA: Sinauer Associates.

Wang, S., Kubota, T., Siskind, J. M., & Wang, J. (2005). Salient Closed Boundary Extraction with Ratio Contour. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *27*, 546-560.

Watamaniuk, S. N., & Sekuler, R. (1992). Temporal and Spatial Integration in Dynamic Random-Dot Stimuli. *Vision Research*, *32*, 2341-2347.

Williams, L. R., & Thornber, K. K. (2000). A Comparison of Measures for Detecting Natural Shapes in Cluttered Backgrounds. *International Journal of Computer Vision, 34*, 81-96.

Yamada, T. (2006). CCD Image Sensors. In J. Nakamura, *Image Sensors and Signal Processing for Digital Still Cameras*. CRC Press.

Yang, Y., & Yan, H. (2000). An adaptive logical method for binarization of degraded document images. *Pattern Recognition*, *33*, 787-807.

Yen, S. C., & Finkel, L. H. (1998). Extraction of Perceptually Salient Contours by Striate Cortical Networks. *Vision Research*, *38*, 719-741.

Yoshida, A., Blanz, V., Myszkowski, K., & Seidel, H. (2005). Perceptual evaluation of tone mapping operators with real-world scenes. *The Human Vision and Electronic Imaging X* (pp. 192-203). San Jose, USA: Proc. IS&T/SPIE Electronic Imaging.

Zeki, S. (1993). A Vision of the Brain. Blackwell Scientific Publications.

Δημοσιεύσεις βασισμένες στην παρούσα έρευνα

Δημοσιεύσεις σε διεθνή επιστημονικά περιοδικά με κριτές

- 1. Vonikakis, V., Andreadis, I., & Gasteratos, A. (2008). Fast centre-surround contrast modification. *IET Image processing*, 2(1), 19-34.
- Vonikakis, V., Gasteratos, A., & Andreadis, I. (2006). Enhancement of Perceptually Salient Contours using a Parallel Artificial Cortical Network. *Biological Cybernetics*, 94(3), 192-214.
- Vonikakis, V., Andreadis, I., & Papamarkos, N. (2007). Robust document binarization with OFF center-surround cells. *Pattern Analysis and Applications*, accepted for publication.
- 4. Vonikakis, V., Andreadis, I., & Gasteratos, A. (2005). Recognition of 2-d Shapes Inspired by the Human Visual System. *Journal of Computational Methods in Science and Engineering*, accepted for publication.
- 5. Iakovidou, C., Vonikakis, V., & Andreadis, I. (2007). Real-time biologically inspired image enhancement. Submitted for publication

Δημοσιεύσεις σε διεθνή επιστημονικά συνέδρια με κριτές

- 1. Vonikakis, V., Andreadis, I., & Gasteratos, A. (2004). A New Approach to Machine Contour Perception. *Int. Conf. on Computational Methods in Sciences and Engineering*. (pp. 566-568). Athens, Greece.
- Vonikakis, V., Andreadis, I., & Gasteratos, A. (2005). Simple Shape Classification Based on the Human Visual System. *IASTED Int. Conf. on Visualization, Imaging and Image Processing*. (pp. 162-167). Benidorm, Spain.
- 3. Vonikakis, V., Andreadis, I., & Gasteratos, A. (2006). Extraction of Salient Contours in Color Images. *4th Panhellenic Conference of Artificial Intelligence (SETN 2006), Lecture Notes in Computer Science*. (pp. 400-410), Vol. 3955. Heraklion, Greece.
- Panitsidis, G., Konstandinidis, K., Vonikakis, V., Andreadis, I., & Gasteratos, A. (2006). Fast Image Retrieval Based on Attributes of the Human Visual System. *IEEE NORSIG'06*. (pp. 206-209). Reykjavik, Iceland.

- 5. Vonikakis, V., Andreadis, I., & Gasteratos, A. (2007). Fast Dynamic Range Compression for Grey Scale Images. *Int. Workshop on Advanced Imaging Technology*. (pp. 35-39). Bangkok, Thailand.
- 6. Vonikakis, V., Andreadis, I., Papamarkos, N., & Gasteratos, A. (2007). Adaptive Document Binarization: A Human Vision Approach. *Int. Conference on Computer Vision Theory and Applications*. (pp. 104-110). Barcelona, Spain.
- Vonikakis, V., & Andreadis, I. (2007). Fast Automatic Compensation of Under/Over Exposured Image Regions. *IEEE Pacific Rim Symposium on Image Video and Technology (PSIVT 2007), Lecture Notes in Computer Science*. (pp. 510-521), Vol. 4872. Santiago, Chile. Awarded "Best Paper in Computer Vision Applications".

Κατάλογος Αγγλικών Όρων

Αγγλικός Όρος	Μετάφραση
Affinity Function	Συνάρτηση Συσχέτισης
Anchoring Problem	Πρόβλημα Αγκύρωσης (για αποχρώσεις)
Aperture Condition	Συνθήκη διαφράγματος
Association Field	Πεδίο συσχέτισης
Beam Splitter	Διαχωριστής ακτίνας
Bipole Cells	Δίπολα κύτταρα
Brightness	Λαμπρότητα
Center-Surround	Κέντρο-Περιφέρεια (για κύτταρα και φίλτρα)
Character Stroke	Πάχος χαρακτήρα (γραμματοσειράς)
Chromatic Aberration	Χρωματική εκτροπή
Co-circular	Συγκυκλικός
Co-exponential	Συνεκθετικός
Color Constancy	Χρωματική σταθερότητα
Color Contrast	Χρωματική αντίθεση
Complex Cells	Σύνθετα κύτταρα (Πρωτοταγούς Οπτικού Φλοιού)
Cones	Κωνία
Dark Current	Ρεύμα Σκότους
Dark Noise	Θόρυβος Σκότους
Demosaicing	Απομωσαϊκοποίηση
Divisive Inhibition	Διαιρετική αναστολή
Dominant Wavelength	Επικρατών μήκος κύματος
Dynamic Range	Δυναμικό Εύρος
Excitatory	Διεγερτικός
Exposure	Έκθεση (φωτοευαίσθητης επιφάνειας σε μια φωτεινή πηγή)

Fast Plasticity Term	Όρος ταχείας πλαστικότητας	
Filling-in	Γέμισμα	
Ganglion cells	Γαγγλιακά κύτταρα	
Gray World rule	Κανόνας του Γκρίζου Κόσμου	
Ground Truth Image	Ιδανική Εικόνα	
Halo effects - artifacts	Εφέ άλω (γεν. της λέξης άλως)	
Histogram Equalization	Εξισορρόπηση ιστογράμματος	
Hypercolumn	Υπερστήλη	
Illuminance	Φωτισμός	
Illusionary Contour	Ψευδαισθητικό περίγραμμα	
Induced Color	Προκαλούμενο χρώμα	
Infero-temporal Cortex	Κάτω Κροταφικός Λοβός	
Inhibitory	Ανασταλτικός	
Intensity	Ένταση	
Inter-columnar Competition	Ανταγωνισμός μεταξύ υπερστηλών	
Intra-columnar Competition	Ανταγωνισμός εντός της υπερστήλης	
Lateral Inhibition	Πλευρική αναστολή	
Leaky Integrator Neurons	Διαρρέοντες νευρώνες ενοποίησης	
Lightness	Φωτεινότητα	
Luminance	Φωτοβολία	
Luminous flux	Φωτεινή ροή	
Luminous intensity	Φωτεινή ένταση	
Magni	Μεγάλο	
Mesopic	Μεσαία επίπεδα φωτισμού	
Neurophysiology	Νευροφυσιολογία	
Neuroscience	Νευροεπιστήμη	
OFF-center Ganglion Cells	Γαγγλιακά κύτταρα σκοτεινού κέντρου	
ON-center Ganglion Cells	Γαγγλιακά κύτταρα φωτεινού κέντρου	
Overexposed	Υπερεκτεθειμένος	
Parvi	Μικρό	
Photopic	Υψηλά επίπεδα φωτισμού	
Photoreceptors	Φωτοϋποδοχείς	

Primal Sketch	Βασικό σκιαγράφημα
Primary Visual Cortex	Πρωτοταγής Οπτικός Φλοιός
Receptive Field	Υποδεκτικό Πεδίο
Reflectance	Ανακλαστικότητα
Retina	Αμφιβληστροειδής
Retinex	Οικογένεια αλγορίθμων για τον υπολογισμό της αντιληπτής φωτεινότητας
Retinotopical	Σύμφωνος με την οργάνωση του αμφιβληστροειδούς
Rods	Ραβδία
Salient Contours	Εξέχοντα Περιγράμματα
Saturation Voltage	Τάση Κορεσμού
Scotopic	Χαμηλά επίπεδα φωτισμού
Serpentine Memory	Μνήμη σερπαντίνας
Shunting Inhibition	Διακλαδισμένη αναστολή
Simple Cells	Απλά κύτταρα (Πρωτοταγούς Οπτικού Φλοιού)
Solid angle	Στερεά γωνία
Tone mapping	Απεικόνιση Αποχρώσεων
Underexposed	Υποεκτεθειμένος
White Patch rule	Κανόνας του Λευκού Σημείου
Zero Padding	Πλήρωση με μηδενικά

Κατάλογος Ακρωνυμίων

Ακρωνύμιο	Ολοκληρωμένη Έκφραση	
ACE	Automatic Color Equalization	
AE	Automatic Exposure	
ALM	Adaptive Logical Method	
ASIC	Application Specific Integrated Circuit	
BCS	Boundary Contour System	
CCD	Charge-Coupled Device	
CIE	Commission Internationale de l'Eclairage	
CINNIC	Carefully Implemented Neural Network for Integrating Contours	
DoG	Difference of Gaussians	
DSP	Digital Signal Processor	
FIFO	First-In-First-Out (Memory)	
FPGA	Field-Programmable Gate Arrays	
GT	Ground Truth	
GW	Gray World (rule)	
HDR	High Dynamic Range (>255:1)	
IT	Infero-temporal Cortex	
JPEG	Joint Photographic Experts Group	
LLL	Local Linear LUT	
LMS	Long – Medium – Short (photoreceptors)	
LoG	Laplacian of Gaussian	
LUT	Look-Up Table	
MLP	Multi Layer Perceptron	
MSR	Multi-Scale Retinex	

MT	Middle-temporal Cortex
OCR	Optical Character Recognition
OM	Orientation Map
PSNR	Peak Signal-to-Noise Ratio
RGB	Red – Green – Blue
SDR	Standard Dynamic Range (255:1)
SoG	Sum of Gaussians
SSR	Single-Scale Retinex
V1	Visual Area 1
Vx	Visual Area x
WP	Wight Patch (rule)
ΑΟΣ	Ανθρώπινο Οπτικό Σύστημα
ΜΔΕ	Μεγάλο Δυναμικό Εύρος (>255:1)
ΣΔΕ	Συνηθισμένο Δυναμικό Εύρος (255:1)