

Εκτίμηση αντιληπτής ποιότητας εικόνας πλήρους αναφοράς  
με οπτικές περιοχές βάσει στατιστικών κατανομών

Η Μεταπτυχιακή Διπλωματική Εργασία

υποβάλλεται στην ορισθείσα

από τη Συνέλευση

του Τμήματος Μηχανικών Η/Υ και Πληροφορικής

Εξεταστική Επιτροπή

από την

Ελένη Παρασκευά

ως μέρος των υποχρεώσεων για την απόκτηση του

ΔΙΠΛΩΜΑΤΟΣ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ  
ΣΤΗ ΜΗΧΑΝΙΚΗ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΥΠΟΛΟΓΙΣΤΙΚΩΝ  
ΣΥΣΤΗΜΑΤΩΝ ΜΕ ΕΙΔΙΚΕΥΣΗ  
ΣΤΗΝ ΕΠΙΣΤΗΜΗ ΚΑΙ ΜΗΧΑΝΙΚΗ ΔΕΔΟΜΕΝΩΝ

Πανεπιστήμιο Ιωαννίνων

Πολυτεχνική Σχολή

Ιωάννινα 2024

Εξεταστική επιτροπή:

- **Κόντης Λυσίμαχος-Παύλος**, Καθηγητής Τμήμα Μηχανικών Η/Υ και Πληροφορικής, Πανεπιστήμιο Ιωαννίνων (Επιβλέπων)
- **Νίκου Χριστόφορος**, Καθηγητής Τμήμα Μηχανικών Η/Υ και Πληροφορικής, Πανεπιστήμιο Ιωαννίνων
- **Παρσόπουλος Κωνσταντίνος**, Καθηγητής Τμήμα Μηχανικών Η/Υ και Πληροφορικής, Πανεπιστήμιο Ιωαννίνων

## ΑΦΙΕΡΩΣΗ

---

Αφιερώνω αυτή τη διπλωματική εργασία στην οικογένειά μου.

## ΕΥΧΑΡΙΣΤΙΕΣ

---

Ολοκληρώνοντας τον κύκλο των σπουδών μου με την παρούσα διπλωματική εργασία που υλοποιήθηκε στο πλαίσιο του μεταπτυχιακού προγράμματος σπουδών «Μηχανική Δεδομένων και Υπολογιστικών Συστημάτων» του Τμήματος Μηχανικών Ηλεκτρονικών Υπολογιστών και Πληροφορικής, στην Πολυτεχνική Σχολή του Πανεπιστημίου Ιωαννίνων, θα ήθελα να εκφράσω τις ευχαριστίες μου σε όσους με στήριξαν, τόσο πρακτικά όσο και ψυχολογικά.

Πρώτιστα, ευχαριστώ εγκάρδια τον επιβλέποντα καθηγητή μου, κ. Λυσίμαχο Κόντη, Καθηγητή του Τμήματος Μηχανικών Ηλεκτρονικών Υπολογιστών και Πληροφορικής για την ανάθεση του θέματος, καθώς και για την πολύτιμη καθοδήγηση και τις συμβουλές του.

Επίσης, ευχαριστώ θερμά τον κ. Αντώνιο Μαϊργιώτη, Διδάκτορα Πληροφορικής, για την συνεργασία και την ανεκτίμητη βοήθειά του καθ' όλη τη διάρκεια της εκπόνησης αυτής της διπλωματικής εργασίας, καθώς η συνεισφορά του ήταν καθοριστική.

Τέλος, εκφράζω την ευγνωμοσύνη μου στην οικογένειά μου και στους φίλους μου, για την αδιάλειπτη ψυχολογική στήριξη που μου παρείχαν καθ' όλη τη διάρκεια των σπουδών μου.

# ΠΕΡΙΕΧΟΜΕΝΑ

Αφιέρωση	2
Ευχαριστίες	3
Κατάλογος Σχημάτων	iv
Κατάλογος Πινάκων	vi
Περίληψη	viii
Extended Abstract	ix
<b>ΚΕΦΑΛΑΙΟ 1 Εισαγωγή</b>	<b>1</b>
1.1 Αντικείμενο της Διπλωματικής .....	1
1.2 Κατηγορίες IQA αλγορίθμων .....	3
1.3 MOS (Mean Opinion Score) - DMOS (Differential Mean Opinion Score).....	9
1.4 Διακριτός Μετασχηματισμός Wavelet (DWT).....	10
1.5 Οπτική Ευκρίνεια (Visual Saliency) .....	12
1.6 Δομή της Διατριβής .....	14
<b>ΚΕΦΑΛΑΙΟ 2 Κοινά Σύνολα Δεδομένων Και Μετρικές Αξιολόγησης</b>	<b>16</b>
2.1 Πειράματα και Βάσεις Δεδομένων Εικόνων .....	16
2.2 Μετρικές .....	22
2.2.1 MSE (Mean Square Error) .....	22
2.2.2 PSNR (Peak Signal to Noise Ratio).....	23
2.2.3 MDSI.....	24
2.2.4 ESSIM .....	25
2.2.5 SG_ESSIM .....	27
2.2.6 VCGS.....	29
2.2.7 VSI.....	31
2.2.8 SSIM .....	33
2.2.9 MS_SSIM .....	35

2.2.10	GMSD .....	36
2.2.11	Πίνακας Συγκέντρωσης Όλων Των Μετρικών .....	38
2.3	Μετασχηματισμοί Fourier και wavelet για την αξιολόγηση ποιότητας εικόνας.....	39
2.4	Κριτήρια αξιολόγησης (Αποτίμησης) .....	42
<b>ΚΕΦΑΛΑΙΟ 3</b>		<b>45</b>
<b>Μετρικές Αξιολόγησης Πλήρους Αναφοράς</b>		<b>45</b>
3.1	Μετρικές αξιολόγησης πλήρους αναφοράς .....	45
3.2	Προσεγγίσεις βασισμένες σε φυσικές εικόνες (natural images) - NSS - Στατιστική φυσικών σκηνών.....	48
3.3	Μετρική βασισμένη στην αντιληπτή οπτική ποιότητα έγχρωμων εικόνων	48
3.4	Στατιστική περιγραφή συντελεστών κυματιδίων.....	52
3.5	Πρώτη ομάδα πειραμάτων - χρήση νέων χαρακτηριστικών υφής.....	55
3.6	Δεύτερη ομάδα πειραμάτων .....	58
3.7	Τρίτη ομάδα πειραμάτων .....	62
3.8	Τέταρτη ομάδα πειραμάτων - Υπολογισμός visual saliency βασισμένος σε μη-χαμηλοπερατή πληροφορία με πολλά επίπεδα μετασχηματισμού.....	66
3.8.1	Χρήση πληροφορίας λεπτομέρειας - μηδενισμός πληροφορίας προσέγγισης .....	68
3.8.2	Χρήση πληροφορίας λεπτομέρειας με στάθμιση - μηδενισμός πληροφορίας προσέγγισης - χρήση σιγμοειδούς συνάρτησης.....	71
3.8.3	Χρήση πληροφορίας λεπτομέρειας - μηδενισμός πληροφορίας προσέγγισης - προσαρμογή με Student't-t κατανομή .....	73
<b>ΚΕΦΑΛΑΙΟ 4</b>		<b>76</b>
<b>Εικόνες με περιεχόμενο οθονών (Screen Content Images - SCI)</b>		<b>76</b>
4.1	Screen Content Images .....	76
4.2	Βάση Δεδομένων SCID.....	78
4.3	Πειράματα για SCI.....	79

4.3.1 Τροποποίηση της μεθοδολογίας VSSC με πρόσθεση χαρακτηριστικών υφής .....	79
4.3.2 Μεθοδολογία με χρήση πρώτων διαφορών - πληροφορία κλίσης.....	81
4.3.3 Χρήση πληροφορίας λεπτομέρειας με στάθμιση - μηδενισμός πληροφορίας προσέγγισης.....	83
4.3.4 Χρήση πληροφορίας λεπτομέρειας με στάθμιση - μηδενισμός πληροφορίας προσέγγισης - χρήση σιγμοειδούς συνάρτησης.....	84
4.4 Πειράματα για SCI .....	85
<b>ΚΕΦΑΛΑΙΟ 5</b>	<b>86</b>
<b>Συμπεράσματα</b>	<b>86</b>
5.1 Συμπεράσματα 3 <sup>ου</sup> Κεφαλαίου και Μελλοντική εργασία.....	88
5.2 Συμπεράσματα 4 <sup>ου</sup> Κεφαλαίου και Μελλοντική εργασία .....	89
<b>Βιβλιογραφία</b>	<b>91</b>
<b>Σύντομο Βιογραφικό</b>	<b>1</b>

## ΚΑΤΑΛΟΓΟΣ ΣΧΗΜΑΤΩΝ

---

Εικόνα 1.1. Γενικό πλαίσιο των μεθοδολογιών πλήρους αναφοράς.....	4
Εικόνα 1.2. Γενικό πλαίσιο των μεθοδολογιών χωρίς αναφορά στην αρχική εικόνα .....	7
Εικόνα 1.3. Γενικό πλαίσιο των μεθοδολογιών μειωμένης πρόσβασης σε πληροφορίες της αρχικής εικόνας.....	9
Εικόνα 1.4. Ένα επίπεδο, δισδιάστατου μετασχηματισμού κυματιδίων. Αρχικά, ένας μονοδιάστατος DWT εφαρμόζεται κατά μήκος των γραμμών. Στη συνέχεια, ένας μονοδιάστατος DWT εφαρμόζεται κατά μήκος των στηλών του αποτελέσματος του πρώτου σταδίου, δημιουργώντας τέσσερις περιοχές υποζώνης στον μετασχηματισμένο χώρο: LL, LH, HL και HH.....	11
Εικόνα 2.1. Οπτική αναπαράσταση της συχνότητας εμφάνισης των τιμών DMOS της βάσης δεδομένων LIVE.....	18
Εικόνα 2.2. Δείγμα εικόνων της βάσης δεδομένων LIVE .....	18
Εικόνα 2.3. Οπτική αναπαράσταση της συχνότητας εμφάνισης των τιμών MOS της βάσης δεδομένων LIVE In the Wild .....	19
Εικόνα 2.4. Δείγμα εικόνων της βάσης δεδομένων LIVE In the Wild.....	19
Εικόνα 2.5. Οπτική αναπαράσταση της συχνότητας εμφάνισης των τιμών DMOS της βάσης δεδομένων KADID-10K.....	20
Εικόνα 2.6 : Δείγμα εικόνων της βάσης δεδομένων KADID-10K .....	20
Εικόνα 2.7 : Οπτική αναπαράσταση της συχνότητας εμφάνισης των τιμών MOS της βάσης δεδομένων TID2013. ....	21
Εικόνα 2.8: Δείγμα εικόνων της βάσης δεδομένων TID2013 .....	21
Εικόνα 2.9 : (α) Μετασχηματισμός Fourier ενός σήματος.(b) Μετασχηματισμός wavelet ενός σήματος .....	41
Εικόνα 3.1 : Σκορ ποιότητας μεταξύ της αρχικής και της παραμορφωμένης εικόνας .....	46
Εικόνα 3.2 : Η μεθοδολογία VCGS.....	50



Εικόνα 3.3 : Ροή της VSSC για τη δημιουργία χαρτών χαρακτηριστικών της οπτικής περιοχής, της κλίσης και της χρωματικότητας .....	50
Εικόνα 3.4. Το διάγραμμα ροής για τη δημιουργία των χαρτών ομοιότητας της οπτικής ευκρίνειας, της κλίσης και της χρωματικότητας, του τελικού τοπικού χάρτη ποιότητας και της σταθμισμένης συγκέντρωσης με βάση την οπτική ευκρίνεια για την εξαγωγή της συνολικής βαθμολογίας ποιότητας .....	51
Εικόνα 3.5 : Αρχική εικόνα, εικόνα BDIP και BVLC εικόνα .....	56
Εικόνα 3.6: Αρχική εικόνα από τη βάση δεδομένων TID2013, ανακατασκευασμένες εικόνες για το κανάλι a και το κανάλι b, έχοντας κρατήσει τους συντελεστές προσέγγισης (cA) χωρίς τροποποίηση.....	61
Εικόνα 3.7: Αρχική εικόνα από τη βάση δεδομένων TID2013, ανακατασκευασμένες εικόνες για το κανάλι a και το κανάλι b, έχοντας μηδενίσει τους συντελεστές προσέγγισης cA. ....	61
Εικόνα 3.8 : Αρχική εικόνα από τη βάση δεδομένων KADID-10K, ανακατασκευασμένες εικόνες για το κανάλι a και το κανάλι b, έχοντας κρατήσει τους συντελεστές λεπτομέρειας (cA) .....	62
Εικόνα 3.9: Αρχική εικόνα από τη βάση δεδομένων KADID-10K, ανακατασκευασμένες εικόνες για το κανάλι a και το κανάλι b, έχοντας μηδενίσει το cA .....	62
Εικόνα 3.10: Εφαρμογή δισδιάστατου μετασχηματισμού κυματιδίων 4 επιπέδων .....	66
Εικόνα 3.11: Visual saliency της προτεινόμενης μεθοδολογίας με χρήση του μετασχηματισμού κυματιδίων στην εικόνα βάθους Dab. (α) αρχική εικόνα (β) visual saliency.....	69
Εικόνα 3.12: Σιγμοειδής συνάρτηση .....	71
Εικόνα 3.13: Visual saliency της προτεινόμενης μεθοδολογίας με εφαρμογή της σιγμοειδούς συνάρτησης στους συντελεστές ενδιαφέροντος με χρήση της εικόνας βάθους Dab .....	73
Εικόνα 3.14: Visual saliency της προτεινόμενης μεθοδολογίας με εφαρμογή της προσαρμοσμένης Student's-t στατιστικής κατανομής και τη δημιουργία visual	

saliency μέσω των 'hidden' μεταβλητών κατά την εκτίμηση των παραμέτρων .....	74
Εικόνα 4.1 : Παραδείγματα εικόνων SCI (Screen Content Images) .....	77
Εικόνα 4.2 : Οπτική αναπαράσταση της συχνότητας εμφάνισης των τιμών MOS της βάσης δεδομένων SCID .....	78
Εικόνα 4.3 : Δείγμα εικόνων της βάσης δεδομένων SCID .....	78

## ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

---

Πίνακας 2.1: Συνοπτική πληροφορία των Βάσεων Δεδομένων εικόνων για λόγους αξιολόγησης της ποιότητας των εικόνων .....	17
Πίνακας 2.2: Κατηγορίες Αντικειμενικών Τεχνικών Πλήρους Αναφοράς .....	38
Πίνακας 3.1: Επίδοση μεθοδολογίας με προσθήκη χαρακτηριστικών υψής στη μεθοδολογία της VSSC.....	57
Πίνακας 3.2: Επίδοση μεθοδολογίας με τροποποίηση συντελεστών κυματιδίων και επιλογή συγκεκριμένου καναλιού του χρωματικού μοντέλου CIELAB .....	60
Πίνακας 3.3: Επίδοση μεθοδολογίας βασισμένης στις πρώτες διαφορές σε διάφορες κατευθύνσεις. Χρήση διαφορετικών τιμών κατωφλίωσης $T=\{0.1, 1, 175, 1000, 10000\}$ .....	64
Πίνακας 3.4: Επίδοση μεθοδολογίας βασισμένης στις πρώτες διαφορές σε διάφορες κατευθύνσεις αφαιρώντας τη μέση τιμή. Χρήση διαφορετικών τιμών κατωφλίωσης $T=\{0.1, 1, 175, 1000, 10000\}$ .....	65
Πίνακας 3.5: Επίδοση μεθοδολογίας βασισμένης στην εφαρμογή 4 επιπέδων ανάλυσης μέσω εφαρμογής του δισδιάστατου μετασχηματισμού κυματιδίων, μηδενισμό των συντελεστών προσέγγισης και διατήρηση των συντελεστών λεπτομέρειας.....	70

Πίνακας 3.6: Επίδοση μεθοδολογίας βασισμένης στην εφαρμογή 4 επιπέδων ανάλυσης μέσω εφαρμογής του διδιάστατου μετασχηματισμού κυματιδίων, μηδενισμό των συντελεστών προσέγγισης και διατήρηση των συντελεστών λεπτομέρειας. Στο τελευταίο στάδιο εφαρμόζεται σιγμοειδής συνάρτηση.....	72
Πίνακας 3.7: Επίδοση μεθοδολογίας βασισμένης στην εφαρμογή 4 επιπέδων ανάλυσης μέσω εφαρμογής του διδιάστατου μετασχηματισμού κυματιδίων, μηδενισμό των συντελεστών προσέγγισης και διατήρηση των συντελεστών λεπτομέρειας. Στο τελευταίο στάδιο εφαρμόζεται προσαρμογή (fit) της Student's-t και εν συνεχεία προτείνονται οι visual salient περιοχές.....	75
Πίνακας 4.1: Επίδοση γνωστών μεθοδολογιών στην σχετική με εικόνες που εμφανίζουν περιεχόμενο οθονών (SCI) .....	79
Πίνακας 4.2: Επίδοση μετρικών της εργασίας στην σχετική με εικόνες που εμφανίζουν περιεχόμενο οθονών (SCI) .....	80
Πίνακας 4.3: Επίδοση της μεθοδολογίας που χρησιμοποιεί πληροφορία κλίσης χρησιμοποιώντας πρώτες διαφορές στην σχετική με εικόνες που εμφανίζουν περιεχόμενο οθονών .....	82
Πίνακας 4.4: Επίδοση της μεθοδολογίας που χρησιμοποιεί πληροφορία λεπτομερειών μετά την εφαρμογή του μετασχηματισμού κυματιδίων 4 επιπέδων .....	84
Πίνακας 4.5: Επίδοση της μεθοδολογίας που χρησιμοποιεί πληροφορία λεπτομερειών μετά την εφαρμογή του μετασχηματισμού κυματιδίων 4 επιπέδων. Κατά την αντιστροφή στο τελευταίο επίπεδο χρησιμοποιείται η σιγμοειδής συνάρτηση .....	85

# ΠΕΡΙΛΗΨΗ

---

Ελένη Παρασκευά, Δ.Μ.Σ. στη Μηχανική Δεδομένων και Υπολογιστικών Συστημάτων, Τμήμα Μηχανικών Η/Υ και Πληροφορικής, Πολυτεχνική Σχολή, Πανεπιστήμιο Ιωαννίνων, Ιούνιος 2024

Εκτίμηση αντιληπτής ποιότητας εικόνας πλήρους αναφοράς με οπτικές περιοχές βάσει στατιστικών κατανομών

Επιβλέπων: Κόντης Λυσίμαχος-Παύλος Καθηγητής

Στη συγκεκριμένη εργασία εξετάζεται η ανάπτυξη αλγορίθμων για την αξιολόγηση της ποιότητας των εικόνων, προσεγγίζοντας την ανθρώπινη αντίληψη. Λόγω του κόστους και της χρονοβόρας διαδικασίας αξιολόγησης της ποιότητας με τη βοήθεια ανθρώπινων παρατηρητών, δημιουργούνται υπολογιστικά μοντέλα που μας δίνουν νέες μετρικές. Έτσι προτείνονται αλγόριθμοι για την αξιολόγηση της ποιότητας με χρήση των χαρακτηριστικών όπως η κλίση, ή η χρωματικότητα.

Εργαζόμαστε με αλγορίθμους Full-Reference (FR), ενώ βασική υπόθεση είναι ότι διαθέτουμε την αρχική εικόνα και την εικόνα που έχει δεχτεί κάποια παραμόρφωση. Έτσι, διερευνούμε την εφαρμογή γνωστών μετασχηματισμών αλλά και στατιστικών κατανομών στο πλαίσιο της δημιουργίας οπτικά εξεχουσών περιοχών και την περαιτέρω πρόταση νέων μετρικών. Πιο συγκεκριμένα, προτείνονται νέες μεθοδολογίες για την ανίχνευση οπτικά εξεχουσών περιοχών με την εφαρμογή μετασχηματισμού κυματιδίων (wavelet) πολλών επιπέδων, τη χρήση συναρτήσεων για την τροποποίηση των συντελεστών και την προσαρμογή των στατιστικών κατανομών. Στο ίδιο πλαίσιο εξετάζουμε την επίδοσή τους με χρήση χαρακτηριστικών όπως η υφή. Η επίδοση των αλγορίθμων της εργασίας εξετάζεται σε γνωστές βάσεις δεδομένων.

Επιπλέον, διερευνούμε το πεδίο των εικόνων που προκύπτουν ως περιεχόμενο οθονών γνωστές και ως screen content images. Τέτοιου είδους εικόνες έχουν επαναλαμβανόμενα μοτίβα και λεπτές γραμμές. Η επίδοση των προτεινόμενων αλγορίθμων εξετάζεται σε εξειδικευμένες βάσεις δεδομένων.

## EXTENDED ABSTRACT

---

Eleni Paraskeua, M.Sc. in Data and Computer Systems Engineering, Department of Computer Science and Engineering, School of Engineering, University of Ioannina, Greece, June 2024

Estimation of perceptual quality of a full reference image with visual areas based on statistical distributions

Advisor: Lysimachos Paulos Kontis

This thesis delves into the development and evaluation of algorithms designed to assess image quality in a manner that aligns closely with human perception. Given the significant costs and time demands associated with traditional quality assessment methods involving human observers, there is a strong impetus to develop computational models that can generate reliable and objective metrics. These algorithms aim to emulate human visual assessment by leveraging specific image features such as gradient and chromaticity.

Our focus is on Full-Reference (FR) algorithms, predicated on the availability of both the original, undistorted image and its distorted counterpart. This approach allows for a direct comparison between the original and the altered images. In this context, we investigate the application of well-established transformations and statistical distributions to identify and enhance visually salient regions within images. The study proposes innovative methodologies for detecting these salient regions by employing multi-level wavelet transformations, modifying coefficients through specific functions, and adapting statistical distributions accordingly. Additionally, we explore the role of texture and other image characteristics in improving the performance of these algorithms.

To validate the effectiveness of the proposed algorithms, we conduct extensive testing using renowned image databases. This rigorous evaluation ensures that the algorithms are robust and reliable across a diverse range of image types and quality variations.

Moreover, our research extends into the domain of screen content images (SCIs), which are images typically associated with digital screens and displays. These images often feature repetitive patterns, fine lines, and text, presenting unique challenges for quality assessment. By tailoring our algorithms to address the specific characteristics of SCIs, we aim to enhance their accuracy and applicability in this specialized area. The performance of these algorithms is rigorously tested on specialized databases designed for SCIs, ensuring that they meet the distinct demands of this image category.

Overall, this study contributes to the advancement of image quality assessment by developing sophisticated algorithms that bridge the gap between computational efficiency and human perceptual accuracy. The proposed methodologies have the potential to significantly impact various applications, from digital imaging and broadcasting to medical imaging and beyond.

# ΚΕΦΑΛΑΙΟ 1

## ΕΙΣΑΓΩΓΗ

### 1.1 Αντικείμενο της Διπλωματικής

Η ψηφιακή απεικόνιση αναδεικνύει έναν εντελώς καινούργιο τρόπο επικοινωνίας, καθώς η τεχνολογική πρόοδος διευκολύνει τη δημιουργία και την κοινοποίηση ψηφιακού περιεχομένου σε εκατομμύρια ανθρώπους με λίγα μόνο κλικ και δευτερόλεπτα. Αυτή η ευκολία παραγωγής ψηφιακού περιεχομένου ενθαρρύνει επίσης την ανάγκη για αποτελεσματική διαχείρισή του.

Μία από τις κύριες έννοιες που καθορίζουν πολλές τεχνικές διαχείρισης ψηφιακών εικόνων είναι η ομοιότητα εικόνων. Ωστόσο, η έννοια αυτή είναι πολύπλευρη και ασαφής, επομένως υπόκειται σε διάφορες ερμηνείες και ορισμούς: μπορεί να οριστεί με αντικειμενικά κριτήρια, αλλά επίσης μπορεί να αναφέρεται σε υποκειμενικές ερμηνείες. Η ομοιότητα εικόνων μπορεί να χρησιμοποιηθεί για να ποσοτικοποιήσει την αντιληπτή οπτική ποιότητα των ψηφιακών εικόνων, συχνά χρησιμοποιούμενη στη συμπίεση σήματος και τις τηλεπικοινωνίες. Επιπλέον, μπορεί να χρησιμεύσει ως βάση για την αναζήτηση εικόνων και την ομαδοποίηση τους, καθώς και να βοηθήσει στη διαχείριση προσωπικών φωτογραφικών συλλογών και την εντοπισμό παραβιάσεων πνευματικών δικαιωμάτων και παράνομης χρήσης περιεχομένου [1].

Στις μέρες μας, η τεχνολογική εξέλιξη των συσκευών που σχετίζονται με τη σύλληψη των εικόνων (image acquisition), την αποθήκευσή τους (storage) και το

σημαντικότερο την εμφάνισή τους (display), έχει οδηγήσει σε έναν τεράστιο όγκο δεδομένων σε μορφή εικόνας ή βίντεο. Τέτοιου είδους οπτικά δεδομένα παράγονται στο πλαίσιο εφαρμογών που χρησιμοποιούνται στη καθημερινότητά μας αλλά και σε διάφορες εφαρμογές κρίσιμες και χρήσιμες σε διάφορα πεδία της ζωής. Έτσι αντιλαμβανόμαστε ότι υπάρχει ανάγκη για ανάλυση των οπτικών δεδομένων αναφέροντας λόγου χάριν την πρόβλεψη καιρικών φαινομένων, την επίβλεψη καλλιεργείων, την αλλαγή χρήσης γης, τους γαλαξίες και τα αστρικά φάσματα (stellar spectra), την παρακολούθηση της αποψίλωσης των δασών, τις στρατιωτικές παρακολουθήσεις, την αναγνώριση δακτυλικών αποτυπωμάτων και ίριδας καθώς εν τέλει την αντικατάσταση της ανθρώπινης επίβλεψης (μέσω όρασης) στις βιομηχανίες (εφαρμογές βιομηχανικής επιθεώρησης). Αρκετές εφαρμογές μπορεί να αναφέρει κανείς, αυτό που είναι σημαντικό όμως είναι η ξεκάθαρη και ακριβής απεικόνιση της πληροφορίας ενδιαφέροντος. Κάτι τέτοιο καθοδηγεί τους ερευνητές να δώσουν περισσότερη προσοχή στην ανάλυση εικόνας, την ταξινόμησή τους ή εν γένει την επεξεργασία τους [2]. Συνέπεια τούτου είναι η επεξεργασία, η αποθήκευση και μετάδοση του τεράστιου όγκου δεδομένων εικόνων αναδεικνύουν την ανάγκη για αποτίμηση της ποιότητάς τους. Οι πρόσφατες εξελίξεις λόγου χάριν στην υπολογιστική νέφους (cloud computing) και στην κινητή υπολογιστική (mobile computing) έχουν προσελκύσει ευρύ ενδιαφέρον και νέες προκλήσεις στην κωδικοποίηση της αντίληψης (perceptual coding) και την αξιολόγηση της ποιότητας (quality assessment) [3].

Ακόμη περισσότερο, η ανάπτυξη των κοινωνικών δικτύων και η συνεπαγόμενη αύξηση των υπηρεσιών γύρω από τις εικόνες, παράγει ένα τεράστιο μέγεθος οπτικών δεδομένων που καταλήγει στους καταναλωτές/χρήστες αυτών. Οι ψηφιακές εικόνες όμως, αναπόφευκτα, γίνονται αντικείμενο υποβάθμισης στη διαδικασία από την αρχή της δημιουργίας περιεχομένου μέχρι την "κατανάλωση" των δεδομένων. Οι οποιεσδήποτε παραμορφώσεις συμβαίνουν έχουν μελετηθεί αρκετά την τελευταία εικοσαετία για λόγους συντήρησης, ελέγχου ή ενίσχυσης της λαμβανόμενης ποιότητας εικόνας.

Στο πλαίσιο της παρούσας διπλωματικής, εξετάζουμε αλγόριθμους αξιολόγησης της ποιότητας των εικόνων. Σκοπός αυτών των αλγορίθμων είναι να μετρήσουν πόσο καλή είναι η ποιότητα μιας εικόνας σε σχέση με την ανθρώπινη αντίληψη. Το



ανθρώπινο οπτικό σύστημα (HVS) είναι ένας ιδανικός δέκτης οπτικών πληροφοριών και είναι επίσης ο πιο αξιόπιστος τρόπος αξιολόγησης της ποιότητας της εικόνας. Είναι ανέφικτο σε όλες τις έρευνες να χρησιμοποιούνται ανθρωπίνοι παρατηρητές, οι οποίοι θα βαθμολογούν την ποιότητα κάθε εικόνας ξεχωριστά. Είναι μια διαδικασία τόσο χρονοβόρα όσο και δαπανηρή, γι' αυτό οι ερευνητές στοχεύουν στην δημιουργία μοντέλων, τα οποία θα προσεγγίζουν την ανθρώπινη κρίση. Αυτοί οι αλγόριθμοι βασίζονται σε διάφορες μετρικές και μοντέλα που αξιολογούν την ευκρίνεια, τη φωτεινότητα, τον χρωματισμό και άλλες πτυχές της εικόνας. Συνοπτικά, τα μοντέλα αξιολόγησης ποιότητας εικόνας (IQA) σχεδιάστηκαν για να αναπαράγουν τον τρόπο με τον οποίο οι άνθρωποι αντιλαμβάνονται και αξιολογούν την ποιότητα μιας εικόνας, χρησιμοποιώντας υπολογιστικές μετρικές και αλγόριθμους.

Οι IQA αλγόριθμοι είναι χρήσιμοι σε πολλές εφαρμογές, όπως στη βελτιστοποίηση της ποιότητας της εικόνας κατά τη μετάδοση σε δίκτυα, στην αξιολόγηση της ποιότητας σε φωτογραφίες και βίντεο, καθώς και σε ιατρικές εφαρμογές όπως την απεικόνιση μαγνητικής τομογραφίας (MRI), την αξονική τομογραφία (CT), την απεικόνιση με υπερήχους [4].

## 1.2 Κατηγορίες IQA αλγορίθμων

Οι αλγόριθμοι αποτίμησης της ποιότητας των εικόνων μπορούν να χωριστούν σε διάφορες κατηγορίες, έχοντας ως κριτήριο τη διαθεσιμότητα της αρχικής (ας θεωρήσουμε μη παραμορφωμένης) εικόνας. Επομένως, η διάκριση γίνεται με ευθεία αναφορά στην πληρότητα πρόσβασης στην αρχική, στη μειωμένη διαθεσιμότητα της αρχικής μέσω π.χ. κάποιων παραμέτρων, ή στην καθόλα ανύπαρκτη πρόσβαση στα δεδομένα της αρχικής εικόνας. Περιληπτικά οι κατηγορίες είναι οι εξής:

1. **FR (Full-Reference) IQA Αλγόριθμοι:** Οι αλγόριθμοι της συγκεκριμένης κατηγορίας συγκρίνουν την υπό εξέταση εικόνα με την αρχική εικόνα. Αξιολογούν πόσο αλλάζει η ποιότητα της εικόνας έπειτα από την επεξεργασία ή από τις αυθεντικές παραμορφώσεις που έχει υποστεί. Οι μέθοδοι της κατηγορίας αυτής, αποσκοπούν στην αποτίμηση, προσομοιώνοντας το ανθρώπινο οπτικό σύστημα

(HVS) και λαμβάνοντας υπόψη παράγοντες πέραν των απλών διαφορών κατά εικονοστοιχεία, όπως την αντίθεση, την φωτεινότητα και την δομική ομοιότητα [5].



**Εικόνα 1.1:** Γενικό πλαίσιο των μεθοδολογιών πλήρους αναφοράς

Ο τομέας έχει εξελιχθεί από τις πιο βασικές μετρικές όπως η Αιχμή Σηματοθορυβικού λόγου (Peak Signal to Noise Ratio - PSNR) και το Μέσο Τετραγωνικό Σφάλμα (Mean Square Error- MSE), οι οποίες δεν είναι πάντα συνεπείς με την ανθρώπινη αντίληψη λόγω των υπολογισμών τους ανά εικονοστοιχείο.

Το Μέσο Τετραγωνικό Σφάλμα, έχει κάποια πλεονεκτήματα όπως η ευκολία στη χρήση, ξεκάθαρη φυσική σημασία αφού αναπαριστά την ενέργεια του σήματος σφάλμα (error signal), ικανοποιεί το Θεώρημα του Parseval, έχει την ικανότητα να χρησιμοποιείται σε αλγορίθμους βελτιστοποίησης με λύσεις κλειστής μορφής καθώς και άλλα χαρακτηριστικά [6].

Οι μετρικές αυτές αποτέλεσαν πυλώνα στον τομέα αξιολόγησης ποιότητας εικόνας και χαρακτηρίζονται από την χαμηλή πολυπλοκότητα τους. Ωστόσο όπως συμβαίνει σε όλες τις εκφάνσεις της ανθρώπινης ζωής και κατ' επέκταση και στην επιστήμη, στόχος είναι η εξέλιξη και η βελτίωση. Ιστορικά, οι ερευνητές και στην αξιολόγηση της ποιότητας, δημιούργησαν αυτά τα μέτρα με βάση την κατανόηση του ανθρώπινου οπτικού συστήματος, τα μαθηματικά και τη θεωρία της πληροφορίας.

Ο SSIM (Structural similarity index measure) αποδιδόμενος ως Δείκτης δομικής ομοιότητας [7] αποτέλεσε μια σημαντική ανακάλυψη, λαμβάνοντας υπόψη όχι μόνο την αντίθεση και τη φωτεινότητα αλλά και τη δομική ομοιότητα. Η επιτυχία του οδήγησε σε παραλλαγές όπως ο MS-SSIM [8] και ο FSIM [9], που χρησιμοποιούνται ευρέως στην κοινότητα της επεξεργασίας εικόνας. Η απλότητα της διατύπωσης του SSIM είχε ως αποτέλεσμα υψηλά επίπεδα ερμηνευσιμότητας από τον άνθρωπο

(κάθε συστατικό που αποτελεί την τελική τιμή SSIM μπορεί να αναλυθεί και να αξιολογηθεί ξεχωριστά), καθώς και υψηλή αποτελεσματικότητα.

Ακολουθώντας μια διαφορετική κατεύθυνση από την ανάπτυξη "χειροποίητων" λύσεων ή αλλιώς λύσεων που βασίζονται στη γνώση, έχουν γίνει πολλαπλές προσπάθειες για την ενσωμάτωση της μηχανικής μάθησης (ML- Machine Learning) στη διαδικασία σχεδιασμού της ποιότητας της εικόνας. Αυτό εκτείνεται από την εκμάθηση των ελεύθερων παραμέτρων των χειροποίητων μέτρων έως την αξιοποίηση της βαθιάς μάθησης για την εκπαίδευση ενός νευρωνικού δικτύου ομοιότητας εικόνας. Τέτοιου είδους προσεγγίσεις έχουν ανεβάσει με επιτυχία τον πήχη της αποτελεσματικότητας της ομοιότητας εικόνας. Ωστόσο, ειδικά για τις μεθόδους βαθιάς μάθησης, τα μοντέλα που προκύπτουν τείνουν να είναι ιδιαίτερα προσαρμοσμένα σε ένα συγκεκριμένο σύνολο σχολιασμένων δεδομένων, καθώς ως γνωστών είναι μεθοδολογίες άπληστες ως προς τα δεδομένα. Επιπλέον, είναι ιδιαίτερα δαπανηρές υπολογιστικά, αν και οι επιταχυντές υλικού (hardware accelerators) μετριάζουν το πρόβλημα.

Επιπλέον, εισήχθησαν μέθοδοι βαθιάς μάθησης, αξιοποιώντας προ-εκπαιδευμένα βαθιά CNN (Convolutional Neural Networks) δίκτυα για την εξαγωγή ουσιαστικών χαρακτηριστικών για την αξιολόγηση της αντιληπτικής ποιότητας. Μέθοδοι όπως το DeepSim, το DeepIQA, DIQaM-FR και WaDIQaM-FR [10], [11] χρησιμοποιούν προσεγγίσεις βαθιάς μάθησης, εκπαιδεύοντας δίκτυα για την κατανόηση της ποιότητας της εικόνας μέσω της εξαγωγής χαρακτηριστικών και της βελτιστοποίησης των οπτικών βαρών με βάση τις πληροφορίες της βάσης δεδομένων.

Η ανάπτυξη ενός μοντέλου αξιολόγησης ποιότητας εικόνας πλήρους αναφοράς (FR-IQA) απαιτεί προηγμένες τεχνικές, και σε αυτό το πλαίσιο για παράδειγμα το DeepSim αναφέρεται σε μια νέα μέθοδο αξιολόγησης ποιότητας εικόνας (IQA), που εστιάζει στη μελέτη του πώς η αρχιτεκτονική του βαθύς νευρωνικού δικτύου (DNN) επηρεάζει την απόδοση της αξιολόγησης IQA. Χρησιμοποιεί ένα νέο πλαίσιο πλήρους αναφοράς, όπου πρώτα μετρά τις τοπικές ομοιότητες μεταξύ των χαρακτηριστικών που παράγονται από ένα μοντέλο DNN για τη δοκιμαστική εικόνα (test image) και εκείνων της εικόνας αναφοράς. Στη συνέχεια, οι τοπικοί

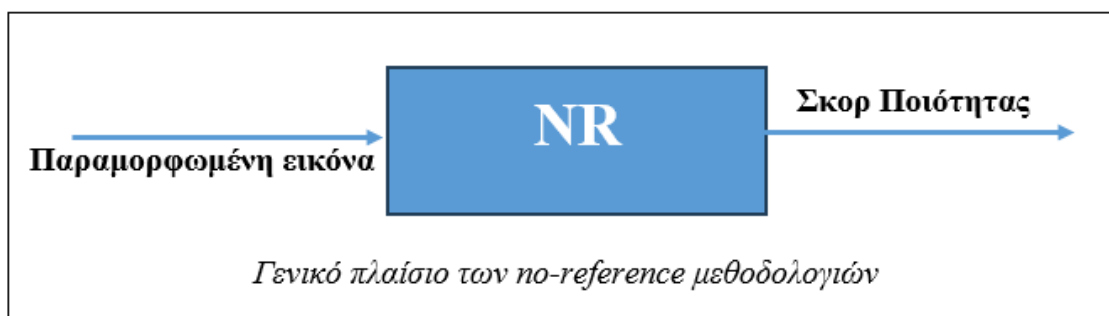
δείκτες ποιότητας συγκεντρώνονται σταδιακά για την εκτίμηση της συνολικής βαθμολογίας ποιότητας.

Ακόμη περισσότερο όμως μπορεί να χρησιμοποιηθεί ακόμα και συνδυασμός δικτύων (π.χ. ένα δίκτυο που εκπαιδεύεται με χρήση διαφορετικών συνελικτικών νευρωνικών δικτύων (CNN)). Η προσέγγιση αυτή καθιστά τα δίκτυα σημαντικά βαθύτερα από τα προηγούμενα μοντέλα IQA, προσφέροντας έτσι μια πιο ακριβή και αξιόπιστη αξιολόγηση της ποιότητας εικόνας. Συνολικά, ο τομέας έχει επεκταθεί σημαντικά, προχωρώντας πέρα από τις βασικές μετρικές προς εξελιγμένα μοντέλα που προσπαθούν να προσομοιώσουν καλύτερα την ανθρώπινη αντίληψη κατά την αξιολόγηση της ποιότητας της εικόνας.

Στη συγκεκριμένη εργασία ασχολούμαστε με μεθοδολογίες όπου έχουμε πλήρη πρόσβαση στην αρχική εικόνα οπότε προσπαθούμε να υπολογίσουμε την τοπική ποιότητα και εν τέλει την συνολική ποιότητα αφού επιλέξουμε μία κατάλληλη στρατηγική η οποία θα παράγει εν τέλει μία τιμή για την προτεινόμενη μετρική. Στη περίπτωση που ενδιαφερόμαστε να χρησιμοποιούμε και την πληροφορία χρώματος, τότε χρησιμοποιούμε, ανάλογα με το χρωματικό μοντέλο που μας ενδιαφέρει και το κατάλληλο κανάλι. Όπως και να έχει, βασισμένοι στη μεθοδολογία SSIM, προτάθηκε ένα αποδοτικό παράδειγμα συνάρτησης ομοιότητας (similarity function).

## **2. NR (No-Reference) IQA Αλγόριθμοι:**

Οι αλγόριθμοι χωρίς αναφορά αξιολογούν την ποιότητα μιας εικόνας χωρίς την ανάγκη ύπαρξης κάποιας μορφής αρχικής εικόνας. Στο πραγματικό κόσμο η ύπαρξη της αρχικής εικόνας δεν είναι πάντα εφικτή. Γι' αυτό είναι απαραίτητες μετρικές κατάλληλες για περιπτώσεις όπου διατίθεται μόνο η παραμορφωμένη εικόνα, όπως για παράδειγμα σε μια απεικόνιση μιας ακτινογραφίας. Οι διάφορες προσεγγίσεις εν γένει χρησιμοποιούν διάφορα μοντέλα και χαρακτηριστικά της εικόνας για να εκτιμήσουν την ποιότητά της.



**Εικόνα 1.2:** Γενικό πλαίσιο των μεθοδολογιών χωρίς αναφορά στην αρχική εικόνα

Πολλές φορές αναφέρονται και ως τυφλή αξιολόγηση ποιότητας εικόνας (BIQA-Blind Image Quality Assessment) [12]). Πρόκειται για ένα πεδίο έρευνας που επικεντρώνεται στην ανάπτυξη αλγορίθμων και μοντέλων που μπορούν να αξιολογούν αυτόματα την ποιότητα εικόνων χωρίς να είναι απαραίτητη η γνώση της εικόνας αναφοράς ή της μη παραμορφωμένης εικόνας. Ο στόχος είναι να μιμηθούν την ανθρώπινη αντίληψη και ανθρώπινη κρίση της ποιότητας της εικόνας.

Στο πλαίσιο της τυφλής αποτίμησης εμπίπτει το KG-IQA (αξιολόγηση ποιότητας εικόνας με βάση τη γνώση - Knowledge), το οποίο αποσκοπεί στην ενίσχυση των μοντέλων BIQA με την ενσωμάτωση γνώσεων συγκεκριμένου τομέα από το **Ανθρώπινο Οπτικό Σύστημα (HVS)** και τη **Στατιστική Φυσικής Σκηνης (NSS)**.

Στα μοντέλα βαθιάς μάθησης η περιορισμένη διαθεσιμότητα σχολιασμένων δειγμάτων εκπαίδευσης, οδηγεί στον κίνδυνο της υπερεκπαίδευσης. Ένα ιδανικό BIQA μοντέλο αναμένεται να εκπαιδευτεί με όσο το δυνατόν λιγότερα δείγματα, έτσι ώστε να μπορεί εύκολα να αναπτυχθεί για εφαρμογές πραγματικής ζωής με χαμηλό κόστος. Ο στόχος λοιπόν αυτών των μοντέλων είναι η μείωση του όγκου των δειγμάτων εκπαίδευσης και η βελτίωση της ικανότητας γενίκευσης.

Το Ανθρώπινο Οπτικό Σύστημα περιλαμβάνει διάφορες αντιληπτικές πτυχές, όπως φωτεινότητα, αντίθεση, υφή, χρώμα, κίνηση, και χωρικοί μηχανισμοί επεξεργασίας, μεταξύ άλλων. Στο πλαίσιο των αλγορίθμων Αξιολόγησης Ποιότητας Εικόνας (IQA), ο στόχος είναι η πρόβλεψη βαθμολογιών ποιότητας που να αντανακλούν στενά την ανθρώπινη αντίληψη. Η ενσωμάτωση των χαρακτηριστικών της ευαισθησίας στην αντίθεση (CSC) και της απλής αισθητής διαφοράς (JND) είναι ζωτικής σημασίας για τα μοντέλα KG-IQA. Για παράδειγμα η CSC συνδέεται με

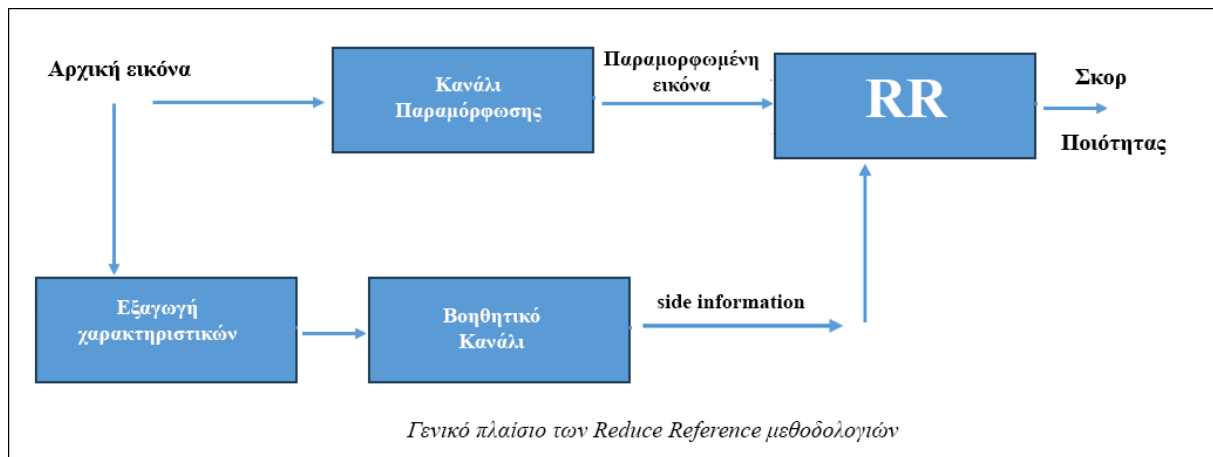
την αναγνώριση της αντίθεσης σε περιοδικά αντικείμενα με βάση τη χωρική συχνότητα, ενώ το JND αναφέρεται στην ελάχιστη αντιληπτή αλλαγή σε ένα ερέθισμα που επηρεάζεται σημαντικά από αλλαγές στη φωτεινότητα.

Οι στατιστικές φυσικής σκηνής (Natural Scene Statistics, NSS) αναφέρονται σε στατιστικές ιδιότητες που συναντώνται σε φυσικές εικόνες. Έχουν αποσπάσει σημαντική προσοχή, λόγω του κρίσιμου ρόλου τους στην αναπαράσταση και την επεξεργασία αυτών των εικόνων. Έχουν εντοπιστεί ορισμένες βασικές ιδιότητες NSS, όπως η φασματική ισχύς που μειώνεται με τη συχνότητα κατά την εφαρμογή του μετασχηματισμού Fourier σε φυσικές εικόνες. Επιπλέον, οι στατιστικές ιδιότητες των εικόνων παραμένουν σταθερές κατά την αλλαγή της κλίμακας των εικόνων. Με άλλα λόγια, κατά την αλλαγή του μεγέθους μιας εικόνας, όπως το να την κάνουμε μικρότερη ή μεγαλύτερη, οι στατιστικές ιδιότητες των δεδομένων αυτών εικόνων παραμένουν παρόμοιες.

Οι ερευνητές αξιοποίησαν αυτές τις ιδιότητες NSS για να αναπτύξουν ιδιότητες NSS με επίγνωση της παραμόρφωσης ειδικά για την αξιολόγηση ποιότητας τυφλών εικόνων (BIQA). Διάφορες μελέτες πρότειναν διάφορες στατιστικές περιγραφές και χαρακτηριστικά αξιολόγησης με βάση τους συντελεστές wavelet, τους συντελεστές MSCN και τους συντελεστές DCT, αντίστοιχα. Τέτοιου είδους χαρακτηριστικά σχεδιάστηκαν για να καταγράφουν και να αναλύουν τις παραμορφώσεις εντός των εικόνων, βοηθώντας σημαντικά στην πρόοδο των μεθόδων BIQA.

Με την ενσωμάτωση της γνώσης της στατιστικής των φυσικών σκηνών στα μοντέλα τους, οι ερευνητές στόχευαν να βελτιώσουν την κατανόηση των φυσικών εικόνων και των παραμορφώσεων, οδηγώντας σε πιο αποτελεσματικά μοντέλα BIQA. Συνολικά, η NSS έχει συμβάλει καθοριστικά στη διαμόρφωση και τη βελτίωση των μεθόδων που χρησιμοποιούνται για την αξιολόγηση της ποιότητας των εικόνων, ιδίως στην κατανόηση και την αντιμετώπιση των παραμορφώσεων.

**3. RR (Reduced-Reference) IQA Αλγόριθμοι:** Οι αλγόριθμοι αυτοί βασίζονται σε περιορισμένα στοιχεία της αρχικής εικόνας, αντί για την πλήρη αναφορά. Χρησιμοποιούν μερικά χαρακτηριστικά για να εκτιμήσουν την ποιότητα. Είναι αναγκαίοι για περιπτώσεις όπου μέρος της αρχικής εικόνας έχει καταστραφεί ή δεν είναι κρίσιμης σημασίας, επομένως περιορίζονται στη σύγκριση τοπικών περιοχών της εικόνας.



**Εικόνα 1.3:** Γενικό πλαίσιο των μεθοδολογιών μειωμένης πρόσβασης σε πληροφορίες της αρχικής εικόνας

### 1.3 MOS (Mean Opinion Score) - DMOS (Differential Mean Opinion Score)

Τα MOS (Mean Opinion Score) και DMOS (Differential Mean Opinion Score) είναι κρίσιμα μέτρα που χρησιμοποιούνται στο πεδίο της αξιολόγησης ποιότητας εικόνας. Αποτελούν το βασικό κριτήριο αξιολόγησης, καθώς αντιπροσωπεύουν τις βαθμολογίες των ανθρώπινων παρατηρητών. Χρησιμοποιούνται ευρέως σε πειράματα αξιολόγησης ποιότητας για να κατανοήσουν την επίδραση των αλλαγών στην αντίληψη των ανθρώπων.

**MOS (Mean Opinion Score):** Η Μέση Βαθμολογία Γνώμης (αποδίδοντας τον όρο στα ελληνικά) είναι ένα μέτρο που υπολογίζει την ποιότητα της εικόνας με βάση την ανθρώπινη αντίληψη. Οι άνθρωποι-παρατηρητές, δηλαδή, βαθμολογούν την ποιότητα της εικόνας με βάση την προσωπική τους κρίση και αίσθηση. Το μέτρο αυτό κυμαίνεται σε κλίμακα από 1 (χαμηλότερη) έως 5 (υψηλότερη). Γενικότερα γνωρίζουμε ότι υψηλότερη τιμή MOS αντιστοιχεί σε υψηλότερη οπτική ποιότητα της εικόνας [13].

**DMOS (Differential Mean Opinion Score):** Το DMOS είναι μια παραλλαγή του MOS που αξιολογεί τη διαφορά στην ποιότητα μεταξύ διαφορετικών εκδόσεων του ίδιου περιεχομένου. Στα ελληνικά θα μπορούσε να αποδοθεί ως Διαφορική Μέση Βαθμολογία Γνώμης. Δηλαδή, αντί να αξιολογεί απευθείας την ποιότητα μιας

εικόνας, το DMOS μετρά τη διαφορά στην ποιότητα που προκύπτει από τη σύγκριση εκδόσεων.

Τόσο το MOS όσο και το DMOS χρησιμοποιούνται για να αξιολογηθεί πως επηρεάζεται η αντίληψη των χρηστών και των παρατηρητών σε τυχόν αλλαγές της ποιότητας και αποτελούν αναπόσπαστο κομμάτι της έρευνας στον τομέα της αξιολόγησης ποιότητας εικόνας. Αναλαμβάνουν τον κρίσιμο ρόλο της διεξαγωγής συμπερασμάτων από πειράματα μετρικών αξιολόγησης ποιότητας εικόνας και συνεπώς, συμβάλλουν στη βελτίωση των αποτελεσμάτων. Συμπερασματικά, αποτελούν το βασικό κριτήριο απόδοσης μιας μετρικής, καθώς αντιπροσωπεύουν τις βαθμολογίες των ανθρώπινων παρατηρητών, τις οποίες και προσπαθούν να προσεγγίσουν, καθώς ο τελικός κριτής είναι ο άνθρωπος.

#### 1.4 Διακριτός Μετασχηματισμός Wavelet (DWT)

Στη συγκεκριμένη ενότητα αναφερόμαστε στον Διακριτό Μετασχηματισμό Wavelet (DWT), που διαμορφώθηκε στα τέλη της δεκαετίας του 1980 από τους Daubechies (1988) και Mallat (1989) και στην εξέλιξή του, έγινε ένα πολυδιάστατο εργαλείο επεξεργασίας σημάτων μετά την πρόταση του Mallat για την πολυανάλυση σημάτων βασισμένη στην αποσύνθεση wavelet. Τα wavelets επιτρέπουν ταυτόχρονα την ανάλυση σε χρόνο και συχνότητα, καθώς η ενέργεια τους είναι συγκεντρωμένη στον χρόνο και έχουν κυματικά (περιοδικά) χαρακτηριστικά. Τα πραγματικά δεδομένα/σήματα είναι περίπλοκα. Οι εικόνες έχουν ομαλές περιοχές που διακόπτονται από ακμές ή απότομες αλλαγές στην αντίθεση. Τα wavelets είναι γρήγορα φθίνουσες κυματοειδείς ταλαντώσεις με μηδενικό μέσο όρο που υπάρχουν για μια πεπερασμένη διάρκεια και είναι ιδανικές για την αναπαράσταση τέτοιων δεδομένων. Ο γενικός τύπος για wavelets είναι:

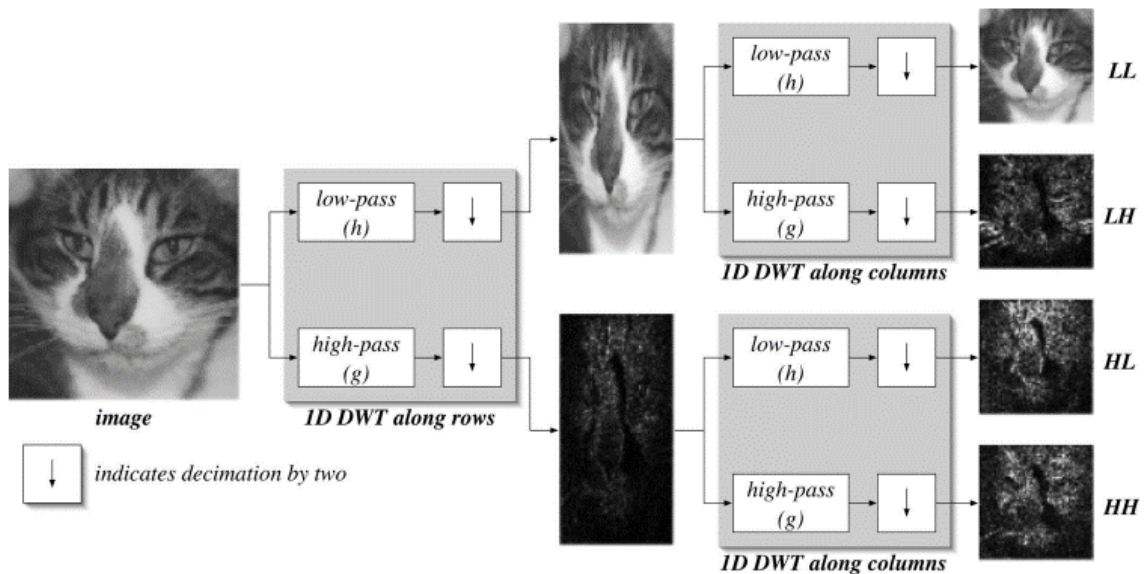
$$W_{\psi}(a, b) = \int_{-\infty}^{+\infty} x(t) \psi \left( \frac{t-b}{a} \right) dt \quad (1.1)$$

όπου  $x$  είναι το πραγματικό σήμα,  $\psi$  είναι μια αυθαίρετη μητρική wavelet,  $a$  είναι η κλίμακα και  $b$  είναι η μετάθεση.

Ο Διακριτός Μετασχηματισμός Wavelet (DWT) είναι ένα ισχυρό εργαλείο που χρησιμοποιείται στην επεξεργασία σημάτων και εικόνων όπως η αποθορυβοποίηση



και η συμπίεση. Ο DWT αποσυνθέτει ένα σήμα - το αρχικό σήμα - σε διαφορετικές συχνότητες (και χρόνους, στην περίπτωση της χρονικής ανάλυσης) χωρίς επικάλυψη ή απώλεια πληροφοριών. Κάθε wavelet αναπαριστά ένα συγκεκριμένο μέρος του σήματος, επιτρέποντας λεπτομερή ανάλυση και επεξεργασία. Η αποσύνθεση του σήματος σε wavelets γίνεται μέσω μιας σειράς από φίλτρα υψηλής (HPF - High-Passe Filter: Διατηρεί τις υψηλές συχνότητες, οι οποίες περιέχουν τις ακμές και τις απότομες αλλαγές της εικόνας) και χαμηλής διέλευσης (LPF - Low Pass Filter: Διατηρεί τις χαμηλές συχνότητες, οι οποίες συνήθως περιέχουν τις βασικές δομές και λεπτομέρειες της εικόνας). Ο DWT πραγματοποιεί πολλαπλά επίπεδα αποσύνθεσης, κάθε φορά μειώνοντας την ανάλυση κατά το ήμισυ, όπως φαίνεται στην εικόνα που ακολουθεί:



**Εικόνα 1.4:** Ένα επίπεδο, διδιάστατου μετασχηματισμού κυματιδίων. Αρχικά, ένας μονοδιάστατος DWT εφαρμόζεται κατά μήκος των γραμμών. Στη συνέχεια, ένας μονοδιάστατος DWT εφαρμόζεται κατά μήκος των στηλών του αποτελέσματος του πρώτου σταδίου, δημιουργώντας τέσσερις περιοχές υποζώνης στον μετασχηματισμένο χώρο: LL, LH, HL και HH.

Όπως φαίνεται και στην εικόνα, αλλά και στα επόμενα, το L σημαίνει φιλτράρισμα χαμηλής διέλευσης και το H σημαίνει υπερβατό φιλτράρισμα. Η ζώνη LL αντιστοιχεί χονδρικά σε μια υποδειγματοληπτική έκδοση της αρχικής εικόνας. Η ζώνη LH τείνει να διατηρεί εντοπισμένα οριζόντια χαρακτηριστικά, ενώ

η ζώνη HL τείνει να διατηρεί εντοπισμένα κάθετα χαρακτηριστικά στην αρχική εικόνα. Τέλος, η ζώνη HH τείνει να απομονώνει εντοπισμένα χαρακτηριστικά σημείων υψηλής συχνότητας στην εικόνα. Συμπερασματικά, στο επίπεδο ένα η εικόνα διασπάται σε τέσσερις περιοχές:

1) **LL (Low-Low)**: Περιέχει τις χαμηλές συχνότητες σε οριζόντιες και κάθετες κατευθύνσεις.

2) **LH (Low-High)**: Περιέχει τις οριζόντιες λεπτομέρειες

3) **HL (High-Low)**: Περιέχει τις κατακόρυφες λεπτομέρειες.

4) **HH (High-High)**: Περιέχει τις διαγώνιες λεπτομέρειες.

Ενώ από το επίπεδο δύο και σε μεγαλύτερο βάθος η διαδικασία αποσύνθεσης επαναλαμβάνεται στις περιοχές LL από το προηγούμενο επίπεδο, επιτρέποντας την περαιτέρω λεπτομερή ανάλυση [29].

Σε αυτό το κεφάλαιο επεκτείνουμε με την σειρά μας την προτεινόμενη μεθοδολογία VSSC με στόχο την δημιουργία μιας νέας μετρικής αξιολόγησης ποιότητας εικόνας, η οποία θα προβλέπει με μεγαλύτερη ακρίβεια την υποκειμενική αξιολόγηση των ανθρώπων.

## 1.5 Οπτική Ευκρίνεια (Visual Saliency)

Έχει εκτιμηθεί ότι τα οπτικά δεδομένα ταξιδεύοντας από τη σκηνή στο ανθρώπινο μάτι έχουν έναν ρυθμό  $10^8$  έως  $10^9$  bits ανά δευτερόλεπτο. Μία τέτοια ροή δεδομένων είναι μία ιδιαίτερα βαριά αποστολή για το ανθρώπινο οπτικό σύστημα. Ευτυχώς, το ΑΟΣ επιλέγει ένα τμήμα των δεδομένων που "βλέπει" οπότε και το επεξεργάζεται περαιτέρω. Αυτός ο μηχανισμός επιλογής που έχει ονομάζεται *Οπτική Προσοχή (Visual Attention)*.

Μία τέτοια επιλεκτική συμπεριφορά θεωρείται ότι καθοδηγείται από 2 είδη μηχανισμών όπου,

- ο πρώτος είναι ο καθοδηγούμενος από το ερέθισμα (stimulus-driven) ή αλλιώς από κάτω προς τα επάνω (bottom-up) μηχανισμός. Σε αυτή την περίπτωση τα χαρακτηριστικά που λειτουργούν με αυτό το μηχανισμό

είναι χαρακτηριστικά των οπτικών σκηνών συμπεριλαμβάνοντας κατεύθυνση, αντίθεση, χρώμα, κίνηση ή κάτι άλλο.

- ο δεύτερος από τον καθοδηγούμενο από την προσδοκία (expectation - driven) ή αλλιώς από επάνω προς τα κάτω (top-down) μηχανισμό. Σε αυτή τη περίπτωση εμπλέκονται θέματα γνωστικής όπως εμπειρία, μνήμη καθώς και πολιτισμικό υπόβαθρο.

Στη περίπτωση της υπολογιστικής όρασης η οπτική προσοχή βασικά σχετίζεται με έναν μηχανισμό προσοχής με κριτήριο την απλότητά του και συνήθως αναφέρεται ως οπτική ευκρίνεια. Η **οπτική ευκρίνεια (visual saliency)** μιας εικόνας ορίζεται σύμφωνα με την αντίληψη του ανθρώπινου οπτικού συστήματος για το πόσο ευδιάκριτη είναι μια τοπική περιοχή. Με άλλα λόγια, υπάρχουν αντικείμενα ή περιοχές που έλκουν, σε μεγαλύτερο βαθμό, την προσοχή του θεατή, σε σύγκριση με τα υπόλοιπα. Έτσι η έννοια της οπτικής ευκρίνειας διαδραματίζει καθοριστικό ρόλο στο τομέα της αξιολόγησης ποιότητας εικόνας, καθώς μας δίνει τη σημαντική πληροφορία όπως αναφέραμε και προηγουμένως.

Η οπτική ευκρίνεια και η προσοχή συνδέονται στενά. Η προσοχή μας φέρνει στο προσκήνιο οπτικά εντυπωσιακά ερεθίσματα, βοηθώντας τα βιολογικά συστήματα να ανιχνεύουν γρήγορα σημαντικά στοιχεία ακόμα και σε ένα κατεστραμμένο οπτικό κόσμο. Μηχανισμοί οπτικής προσοχής επιτρέπουν την επεξεργασία της οπτικής σκηνής μεταφορτώνοντας την προσοχή σε μια περιοχή ή λίγα αντικείμενα κάθε φορά. Επιπλέον, η οπτική ευκρίνεια βοηθά τον εγκέφαλό μας να επιλέξει αποτελεσματικά τον στόχο της προσοχής, με βάση την κυριότητα ορισμένων ερεθισμάτων στο οπτικό πεδίο. Το από κάτω-προς-τα-επάνω, ερεθισμού-καθοδηγούμενο σήμα ανακοινώνει τη σημασία μιας τοποθεσίας στο οπτικό πεδίο, ενώ η επιρροή του χρήστη μπορεί να επηρεάσει την οπτική προσοχή, αντιστρέφοντας ή ενισχύοντας την ευκρίνεια οπτικών τοποθεσιών. Τέλος, η οπτική ευκρίνεια είναι αποτέλεσμα της αλληλεπίδρασης των ερεθισμάτων με το οπτικό σύστημα, παρέχοντας μια ξεχωριστή αντίληψη ανάλογα με τον παρατηρητή και τις συνθήκες [28]. Τα υπολογιστικά μοντέλα της οπτικής ευκρίνειας (δηλαδή από κάτω προς τα πάνω έννοιες της οπτικής προσοχής) σκοπό έχουν να προβλέψουν που εστιάζουν οι άνθρωποι σε μία εικόνα. Τα μοντέλα οπτικής ευκρίνειας έχουν αποδειχθεί χρήσιμα και αρκετά διαφορετικά μοντέλα έχουν προταθεί στη

βιβλιογραφία. Μία σημαντική κατηγορία αυτών είναι οι πολύ-επίπεδες και πολύ-καναλικές μεθοδολογίες οι οποίες μιμούνται το Ανθρώπινο Οπτικό Σύστημα. Αφού χρησιμοποιηθούν, τότε δίνεται η δυνατότητα να δημιουργηθούν σύνολα από χάρτες χαρακτηριστικών (feature maps) οι οποίες κανονικοποιούνται με σκοπό να δημιουργηθούν οι τελικές οπτικές περιοχές.

## 1.6 Δομή της Διατριβής

Η διατριβή περιέχει 4 κεφάλαια. Στο πρώτο κεφάλαιο γίνεται εισαγωγή στο βασικό πρόβλημα που διερευνάται σε αυτή την εργασία ενώ επιπλέον εξετάζονται διάφορες κατηγορίες αλγορίθμων αποτίμησης ποιότητας εικόνων (IQA). Αναφέρεται η έννοια της εξέχουσας οπτικής ευκρίνειας (visual salience). Η κατηγοριοποίηση των αλγορίθμων IQA βασίζεται στον τρόπο με τον οποίο λαμβάνουν υπόψη τους την αρχική εικόνα κατά την αξιολόγηση. Συγκεκριμένα, χωρίζονται σε τρεις κύριες κατηγορίες: τους αλγορίθμους Full-Reference (FR), No-Reference (NR) και Reduced-Reference (RR). Κάθε κατηγορία προσφέρει μοναδικές προσεγγίσεις για την επίλυση του προβλήματος της αξιολόγησης ποιότητας εικόνων, λαμβάνοντας υπόψη διαφορετικές παραμέτρους και περιορισμούς. Επιπλέον, σε αυτό το κεφάλαιο αναλύονται οι έννοιες του MOS (Mean Opinion Score) και DMOS (Differential Mean Opinion Score). Τα MOS και DMOS αποτελούν κρίσιμα μέτρα στον κόσμο της αξιολόγησης ποιότητας εικόνας, καθώς αντιπροσωπεύουν τις βαθμολογίες των ανθρώπινων παρατηρητών.

Στο δεύτερο κεφάλαιο εξετάζονται μερικές από τις πιο γνωστές βάσεις δεδομένων, συνοδευόμενες από ιστογράμματα πληροφορίας MOS ή DMOS. Επιπρόσθετα, παρουσιάζονται μετρικές ποιότητας εικόνας που είναι ευρέως χρησιμοποιούμενες στην έρευνα και περιλαμβάνουν: MSE, PSNR, MDSI, SSIM, MS-SSIM, GMSD, VSI, VCGS, ESSIM, SG\_ESSIM. Στη συνέχεια, αναλύονται οι μετασχηματισμοί Fourier (FT) και Wavelet, οι οποίοι αποτελούν θεμελιώδη εργαλεία στην ανάλυση σημάτων καθώς και εισάγεται η έννοια της οπτικής ευκρίνειας (visual saliency).

Το τρίτο κεφάλαιο αναφέρεται με βάση τη μεθοδολογία πλήρους αναφοράς VCGS αλλά και τροποποιήσεις αυτής (VSSC) προτείνονται επιπλέον τροποποιήσεις λαμβάνοντας υπόψη διάφορα χαρακτηριστικά π.χ. υφής (texture). Επιπλέον, προτείνονται νέες οπτικά εξέχουσες περιοχές με την εφαρμογή μετασχηματισμού κυματιδίων (wavelet) με ανάλυση πολλών επιπέδων, χρήση συναρτήσεων τροποποίησης των συντελεστών, προσαρμογή στατιστικών κατανομών. Πραγματοποιούνται διάφορες ομάδες πειραμάτων οι οποίες εστιάζουν στην επέκταση της μεθόδου VSSC με στόχο τη δημιουργία μιας νέας μετρικής ποιότητας εικόνας.

Στο τέταρτο κεφάλαιο επαναλαμβάνονται τα πειράματα για την αξιολόγηση της ποιότητας των SCI (screen content images), οι οποίες αποτελούν πολύπλοκα περιεχόμενα οθόνης που συνήθως περιλαμβάνουν επαναλαμβανόμενα μοτίβα και λεπτές γραμμές.

## ΚΕΦΑΛΑΙΟ 2

### ΚΟΙΝΑ ΣΥΝΟΛΑ ΔΕΔΟΜΕΝΩΝ

### ΚΑΙ ΜΕΤΡΙΚΕΣ

### ΑΞΙΟΛΟΓΗΣΗΣ

#### 2.1 Πειράματα και Βάσεις Δεδομένων Εικόνων

Κατά την αξιολόγηση της ποιότητας των εικόνων είναι σημαντικό να γίνουν πειράματα κατάλληλα να επιβεβαιώσουν την επίδοση των προτεινόμενων μεθόδων. Υπάρχουν διάφορα πειράματα που μπορεί να γίνουν. Ένα είδος πειραμάτων είναι τα πειράματα συνέπειας (consistency experiments), όπου προσπαθούμε να επιβεβαιώσουμε ότι η αντικειμενική α-αποτίμηση αντιστοιχεί στην ανθρώπινη αποτίμηση. Άλλου είδους πειράματα είναι τα πειράματα επεκτασιμότητας (extensibility experiments), όπου μας ενδιαφέρει να αποδείξουμε εάν η προτεινόμενη μέθοδος εφαρμόζεται σε πολλές εικόνες και παραμορφώσεις χωρίς να χρειάζεται να έχουμε ακόμη περισσότερες εικόνες ή ακόμη περισσότερη εκπαίδευση ανάλογα με την κατηγορία της προτεινόμενης μεθόδου. Άλλου είδους πειράματα σχετίζονται με την δικαιολόγηση της προτεινόμενης μεθοδολογίας και είναι γνωστά ως πειράματα ορθολογικότητας (rationality experiments). Υπάρχουν τα πειράματα ευαισθησίας (sensitivity experiments) όπου έχουν σκοπό να δείξουν την ευαισθησία της μεθοδολογίας, ενώ ιδιαίτερα σημαντικά είναι και τα πειράματα πολυπλοκότητας (complexity experiments) τα οποία ελέγχουν την υπολογιστική επίδοση μίας προτεινόμενης μεθοδολογίας.

Στην έρευνα αξιολόγησης της ποιότητας εικόνας χρησιμοποιούνται διάφορες βάσεις δεδομένων που περιέχουν τιμές MOS (Mean Opinion Score) ή DMOS

(Differential Mean Opinion Score). Αυτές οι τιμές χρησιμοποιούνται για να αξιολογήσουν την ποιότητα της εικόνας ή του βίντεο και να συγκριθούν με τα τελικά αποτελέσματα των μετρικών ποιότητας. Στα επόμενα παρουσιάζονται μερικές από τις πιο γνωστές βάσεις δεδομένων που χρησιμοποιούνται για τον σκοπό αυτόν. Επιπλέον παρουσιάζονται τα ιστογράμματα πληροφορίας των μέτρων MOS ή DMOS για κάθε βάση.

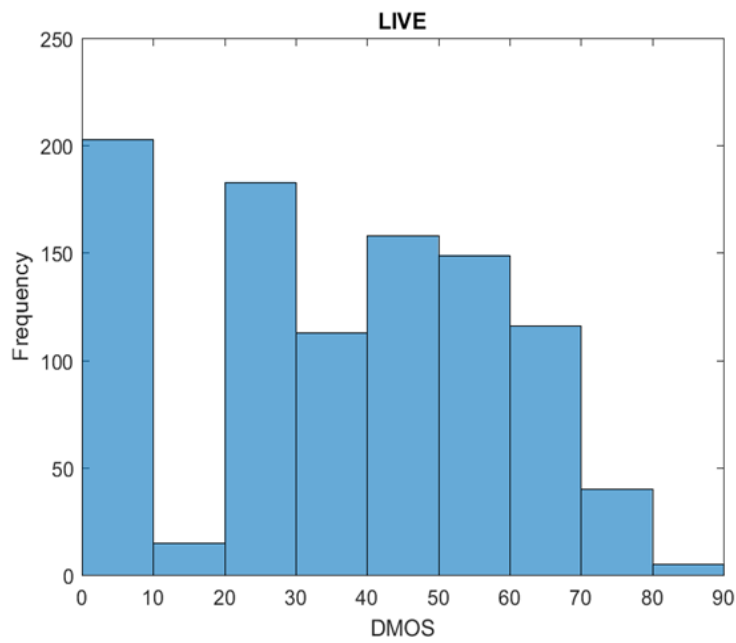
**Πίνακας 2.1:** Συνοπτική πληροφορία των Βάσεων Δεδομένων εικόνων για λόγους αξιολόγησης της ποιότητας των εικόνων

Βάση Δεδομένων	(αριθμός αρχικών εικόνων (reference images))	(αριθμός παραμορφωμένων εικόνων (distorted images))	(αριθμός τύπων παραμόρφωσης (distortion types))	Χρόνος	Αναφορά
LIVE	29	779	5	200 5	[4]
LIVE in the WILD	-	1.162	-	201 5	[5]
KADID-10k	81	10.125	25	201 9	[6], [7]
TID2013	25	3000	24	200 9	[8]

### 1. LIVE Image Quality Assessment Database

29 εικόνες αναφοράς χρησιμοποιήθηκαν με σκοπό να δημιουργηθεί η βάση δεδομένων LIVE Image Quality Assessment Database Release 2 [14]. Περιέχει 779 παραμορφωμένες εικόνες, οι οποίες καλύπτουν πέντε τύπους παραμόρφωσης: JPEG2000 συμπίεση (JP2K), συμπίεση JPEG (JPEG), λευκός θόρυβος (WN), Gaussian blurring (Gblur) και fast fading Rayleigh (FF). Συνολικά περιέχει 982 εικόνες, εκ των οποίων οι 203 είναι αναφοράς. Περίπου 20-29 άνθρωποι

παρατηρητές βαθμολόγησαν κάθε εικόνα. Όλες οι βαθμολογίες μετατράπηκαν σε βαθμολογίες διαφοράς (μεταξύ της παραμορφωμένης και της αναφοράς) και στη συνέχεια σε βαθμολογίες Z όπου κλιμακώθηκαν και μετατοπίστηκαν σε εύρος από 1 έως 100. Έτσι προέκυψε μια τιμή DMOS (Difference Mean Opinion Score) για κάθε παραμορφωμένη εικόνα.



**ΕΙΚΟΝΑ 2.1:** Οπτική αναπαράσταση της συχνότητας εμφάνισης των τιμών DMOS της βάσης δεδομένων LIVE.



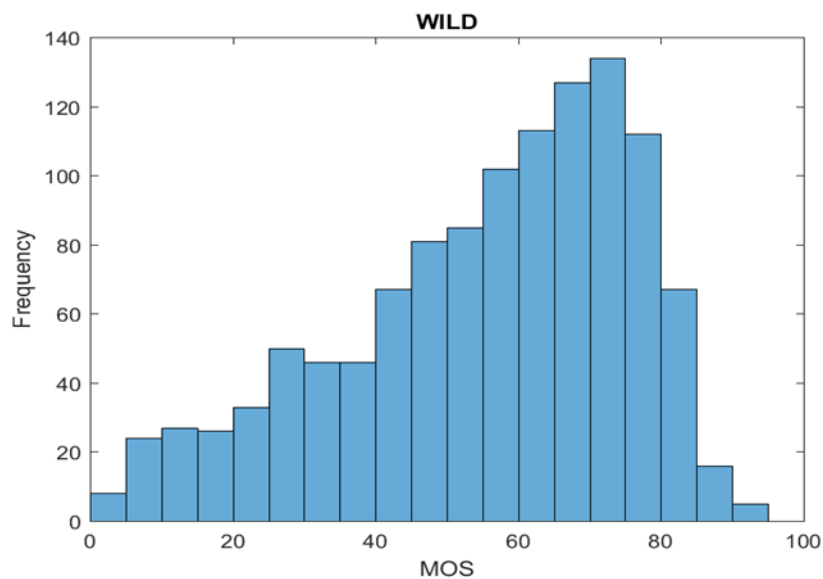
**ΕΙΚΟΝΑ 2.2:** Δείγμα εικόνων της βάσης δεδομένων LIVE.

## 2. LIVE In the Wild Image Quality Challenge Database

Η βάση δεδομένων LIVE In the Wild Image Quality Challenge Database [15] περιλαμβάνει 1.162 "αυθεντικά" παραμορφωμένες εικόνες που συναντώνται στον πραγματικό κόσμο και συμβαίνουν κατά τη λήψη και την αποθήκευση μιας εικόνας χωρίς πρόσθετες τεχνητές παραμορφώσεις. Πολλοί παράγοντες κατά τη λήψη μίας



εικόνας, όπως η ποιότητα της φωτογραφικής μηχανής ή του κινητού, ο φωτισμός, οι καιρικές συνθήκες μπορούν να επηρεάσουν την ποιότητα της. Επιπρόσθετα, κατά την αποθήκευση μιας εικόνας μπορεί να προκύψουν παραμορφώσεις στα χαρακτηριστικά της εικόνας, όπως είναι αλλαγές στον χρωματισμό, την ευκρίνεια, την αντίθεση. Σε αυτή τη βάση δεδομένων, 8100 παρατηρητές προχώρησαν σε πάνω από 350.000 αξιολογήσεις από τις οποίες προέκυψαν και οι τιμές MOS για τις 1.162 εικόνες.



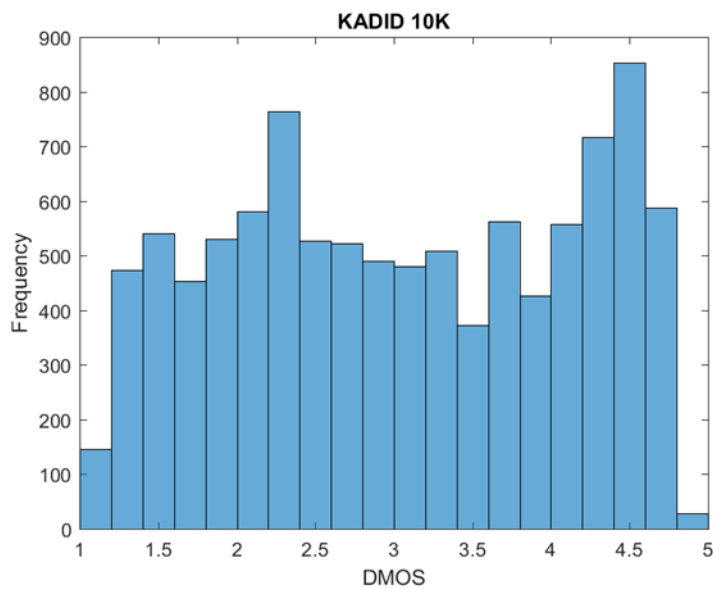
**ΕΙΚΟΝΑ 2.3:** οπτική αναπαράσταση της συχνότητας εμφάνισης των τιμών MOS της βάσης δεδομένων LIVE In the Wild.



**ΕΙΚΟΝΑ 2.4:** Δείγμα εικόνων της βάσης δεδομένων LIVE In the Wild.

### 3. KADID-10k

Οι προηγούμενες βάσεις δεδομένων ήταν μικρές και το περιεχόμενό τους ήταν ιδιαίτερα περιορισμένο. Όσο περισσότερα είναι τα δεδομένα, τόσο πιο αξιόπιστα μπορούν να θεωρηθούν τα αποτελέσματα της έρευνας. Γι' αυτό δημιουργήθηκε η kadid-10k [16], [17] η οποία περιλαμβάνει συνολικά 10.206 εικόνες με τις τιμές DMOS, από τις οποίες οι 81 είναι αναφοράς, και οι υπόλοιπες 10.125 είναι παραμορφωμένες. Οι παραμορφωμένες εικόνες δημιουργήθηκαν με τη χρήση 25 διαφορετικών τύπων παραμορφώσεων διαφορετικών 5 επιπέδων παραμορφώσεων για κάθε από τις 81 αρχικές εικόνες.



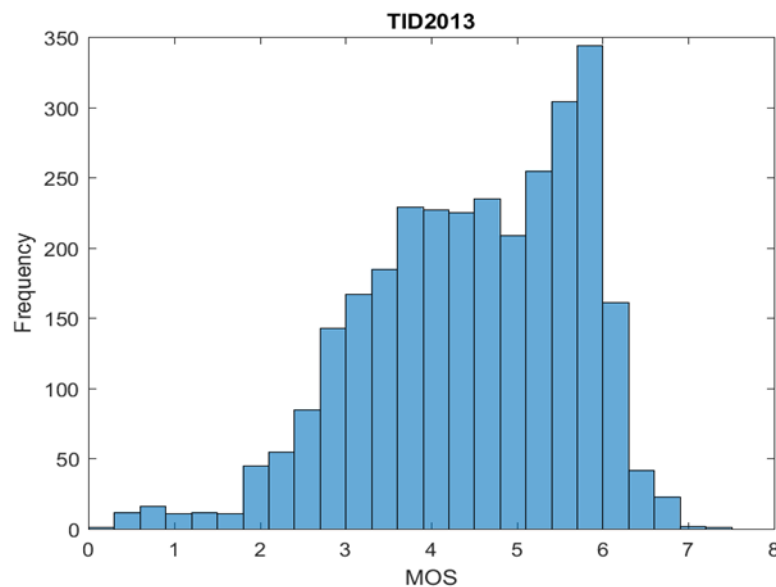
**ΕΙΚΟΝΑ 2.5:** οπτική αναπαράσταση της συχνότητας εμφάνισης των τιμών DMOS της βάσης δεδομένων KADID-10K.



**ΕΙΚΟΝΑ 2.6:** Δείγμα εικόνων της βάσης δεδομένων KADID-10K.

#### 4. TID2013

Η βάση δεδομένων TID2013 [18] περιλαμβάνει 25 εικόνες αναφοράς, σε συνδυασμό με 3.000 παραμορφωμένες εικόνες. Οι παραμορφωμένες εικόνες προκύπτουν από τις αρχικές 25 με τη χρήση 24 διαφορετικών τύπων παραμορφώσεων και 5 επιπέδων παραμορφώσεων. Οι τιμές MOS, όπου 0 αντιστοιχεί στην ελάχιστη και 9 αντιστοιχεί στην μέγιστη, προέκυψαν από τα αποτελέσματα 971 πειραμάτων που διεξήχθησαν σε πέντε διαφορετικές χώρες. Οι 971 ανθρώπινοι παρατηρητές έχουν προβεί σε 524340 συγκρίσεις της οπτικής ποιότητας παραμορφωμένων εικόνων, δηλαδή σε 1048680 αξιολογήσεις ζευγών εικόνων.



**ΕΙΚΟΝΑ 2.7:** οπτική αναπαράσταση της συχνότητας εμφάνισης των τιμών MOS της βάσης δεδομένων TID2013.



**ΕΙΚΟΝΑ 2.8:** Δείγμα εικόνων της βάσης δεδομένων TID2013.

## 2.2 Μετρικές

Οι μέθοδοι IQA μπορούν γενικά να χωριστούν με βάση τις εξής δύο κατηγορίες: υποκειμενική αξιολόγηση με βάση τον άνθρωπο και αντικειμενική αξιολόγηση βάσει αλγορίθμων. Όπως αναφέρθηκε και προηγουμένως, η υποκειμενική αξιολόγηση δεν είναι πάντα εφικτή, επομένως οι ερευνητές εστιάζουν στη δημιουργία μετρικών που προβλέπουν όσο το δυνατόν καλύτερα την ποιότητα μιας εικόνας, όπως θα το έκανε ο άνθρωπος.

Εκτός των προαναφερόμενων, στη συνέχεια, περιγράφονται ορισμένες μετρικές αξιολόγησης ποιότητας εικόνας που χρησιμοποιούνται στις περισσότερες έρευνες, είτε λόγω της ταχύτητάς τους, είτε λόγω της καινοτομίας τους. Στον τομέα της αξιολόγησης ποιότητας εικόνας, οι καινοτομίες μπορεί να αναφέρονται σε νέες μεθόδους ή τεχνικές μέτρησης που προσφέρουν βελτιωμένη ακρίβεια ή να εστιάζουν σε ιδιαίτερες πτυχές της ποιότητας εικόνας που παλαιότερα δεν είχαν εξεταστεί.

Στο τέλος παρουσιάζεται εν συντομία ένας πίνακας με όλες τις μετρικές που έχουν αναφερθεί στη παρούσα διπλωματική, σύμφωνα με την κατηγορία στην οποία εμπίπτει η καθεμία. Για λόγους ευχέρειας στην ανάγνωση και συνέπειας, θεωρούμε τα εξής. Έστω  $\mathbf{X} = \{x_i | 1, 2, \dots, N\}$  και  $\mathbf{Y} = \{y_i | 1, 2, \dots, N\}$  αναπαριστούν τις εικόνες αναφοράς και τις εικόνες παραμόρφωσης αντίστοιχα, όπου  $x_i$ ,  $y_i$  αναπαριστούν τις εντάσεις των  $i$ -οστών δειγμάτων στις εικόνες  $\mathbf{X}$  και  $\mathbf{Y}$  αντίστοιχα, ενώ  $N$  είναι ο αριθμός δειγμάτων (samples ή pixels) της των εικόνων. Τότε έχουμε τους ορισμούς που ακολουθούν.

### 2.2.1 MSE (Mean Square Error)

Το MSE υπολογίζει το μέσο όρο των τετραγωνικών διαφορών μεταξύ των αντίστοιχων εικονοστοιχείων δύο εικόνων. Η MSE επίσης είναι το πρώτο βήμα για τον υπολογισμό του PSNR (Peak Signal-to-Noise Ratio) όπως θα αναφερθεί και στην αμέσως επόμενη μετρική.

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2 \quad (2.1)$$

- $x_i$  είναι η τιμή του εικονοστοιχείου της αρχικής εικόνας στη θέση  $i$ ,

- $y_i$  είναι η τιμή του εικονοστοιχείου της παραμορφωμένης εικόνας στη θέση  $i$ ,
- $N$  είναι ο συνολικός αριθμός των εικονοστοιχείων στην εικόνα.

### 2.2.2 PSNR (Peak Signal to Noise Ratio)

Ο λόγος μέγιστου σήματος προς θόρυβο (PSNR) είναι ένας τεχνικός όρος για τον λόγο μεταξύ της μέγιστης δυνατής ισχύος ενός σήματος και της ισχύος του θορύβου που επηρεάζει την πιστότητα της αναπαράστασής του. Επειδή πολλά σήματα έχουν πολύ μεγάλο δυναμικό εύρος, το PSNR εκφράζεται συνήθως ως λογαριθμικό μέγεθος με χρήση της κλίμακας των ντεσιμπέλ.

$$PSNR = 10 \log_{10} \frac{L^2}{MSE} \quad (2.2)$$

όπου ως  $L$  δηλώνεται το δυναμικό εύρος των εντάσεων των pixels των εικόνων.

Οι συμβατικές μετρικές που χρησιμοποιούνται ευρέως όπως το μέσο τετραγωνικό σφάλμα (MSE) και ο μέγιστος λόγος σήματος προς θόρυβο (PSNR), δεν συσχετίζονται καλά με τις ανθρώπινες υποκειμενικές αξιολογήσεις αλλά και με το περιεχόμενο της εικόνας.

Ο λόγος που προκύπτουν αυτά τα συμπεράσματα αφορά τις προϋποθέσεις όσον αφορά την ποιότητα των εικόνων που αντιλαμβανόμαστε, δηλαδή:

- Τα μέτρα αυτού του τύπου είναι ανεξάρτητα οποιασδήποτε χωρικής ή χρονικής συσχέτισης μεταξύ των δειγμάτων.
- Είναι ανεξάρτητα των σχέσεων μεταξύ σημάτων εικόνων (ή βίντεο) και σημάτων σφάλματος (error signals).
- Προσδιορίζονται από το μέτρο του σήματος σφάλμα μόνο, αγνοώντας οποιαδήποτε επίδραση του πρόσημού των σφαλμάτων.

Δυστυχώς, καμία από τις προαναφερόμενες προϋποθέσεις δεν έχει ισχύ στο πλαίσιο της ποιότητας της αντίληψης των εικόνων (ή βίντεο). Έχειδειχθεί ότι, ακόμα και αν υπάρχουν εικόνες που έχουν ίδια τιμή MSE μπορεί να έχουν δραστικά διαφορετική ποιότητα αντίληψης.

### 2.2.3 MDSI

Το μοντέλο MDSI (Mean Deviation Similarity Index) [19] αξιοποιεί την ομοιότητα κλίσης (GS), την ομοιότητα χρωματικότητας (CS) και τη συγκέντρωση αποκλίσεων (DP). Στόχος όλων των μετρικών είναι η αξιόπιστη διεξαγωγή προβλέψεων, οι οποίες συνάδουν με την ανθρώπινη αντίληψη, μέσω της προσέγγισης του ανθρώπινου οπτικού συστήματος (HVS).

Για τη βελτίωση αυτού παρουσιάζεται μια έκφραση για τον υπολογισμό του κοινού χάρτη ομοιότητας δύο χρωματικών καναλιών με στόχο τον υπολογισμό των χρωματικών μεταβολών. Αυτή η εργασία είναι κρίσιμης σημασίας καθώς οι μετρικές, έως τώρα, παρουσίαζαν ιδιαίτερες αδυναμίες όσον αφορά τους χάρτες ομοιότητας. Το μοντέλο παρέχει χαμηλή πολυπλοκότητα και καλύτερες προβλέψεις.

$$MDSI = \left[ \frac{1}{N} \sum_{i=1}^N |GCS_i^{1/4} - \left( \frac{1}{N} \sum_{i=1}^N GCS_i^{1/4} \right) | \right]^{1/4} \quad (2.3)$$

Το GCS αναφέρεται στον κοινό χάρτη ομοιότητας που προκύπτει από το συνδυασμό των χαρτών ομοιότητας κλίσης (GS ή cGS) με τον κοινό χάρτη χρωματικής ομοιότητας (CS), μέσω ενός συστήματος άθροισης που χρησιμοποιεί έναν σταθμισμένο μέσο όρο. Έτσι, το GCS αποτελεί έναν συνολικό δείκτη ομοιότητας μεταξύ των δύο εικόνων, λαμβάνοντας υπόψη τόσο τις χρωματικές όσο και τις δομικές πληροφορίες τους. Το πιθανό διάστημα τιμών του GCS είναι  $[0-\delta_1, 1+\delta_2]$ , όπου  $\delta_1 < 1$  και  $\delta_2 < 1$ .

Οι χάρτες ομοιότητας κλίσης (GS) και συγκεντρώνουν πληροφορίες σχετικά με τη δομή και την απότομη μετάβαση των εικόνων. Αυτοί οι χάρτες μπορούν να χρησιμοποιηθούν για να αναγνωρίσουν τις περιοχές με ομοιότητα δομής μεταξύ δύο εικόνων. Από την άλλη πλευρά, η προτεινόμενη παραλλαγή του χάρτη ομοιότητας κλίσης (cGS) επικεντρώνεται στον υπολογισμό της ομοιότητας μεταξύ δύο εικόνων με βάση τις αλλαγές στη δομή που προκύπτουν από την προσθήκη ή την αφαίρεση περιοχών. Αυτός ο χάρτης μπορεί να είναι πιο ευαίσθητος σε αλλαγές που επηρεάζουν την δομή της εικόνας.

## 2.2.4 ESSIM

Το μοντέλο ESSIM (edge-strength-similarity-based image quality metric) [20] εμβαθύνει στις ακμές κάθε pixel καθώς υποδεικνύουν κρίσιμες πληροφορίες για την εικόνα. Γενικότερα επικρατεί η αντίληψη ότι μέσω των ακμών επιτυγχάνεται οπτική διαχώριση των αντικειμένων μιας εικόνας. Αυτό όμως δεν συμβαίνει, παρότι θα ήταν ιδιαίτερα χρήσιμο για την έρευνα. Στη πραγματικότητα οι ακμές υποδηλώνουν είτε απότομες μετατοπίσεις, όπως για παράδειγμα στο χρώμα, είτε κάποιου είδους ασυνέχεια.

Ο άνθρωπος αξιολογεί τη ποιότητα εικόνας με κριτήριο την αλλοίωση των σημαντικών πληροφοριών. Η σημαντική πληροφορία για παράδειγμα σε ένα πορτρέτο είναι το πρόσωπο του απεικονιζόμενου. Η συνέπεια των μετρικών IQA τα τελευταία χρόνια με την υποκειμενική αξιολόγηση βελτιώθηκε με την εισαγωγή του δομικού δείκτη ομοιότητας (SSIM). Το μοντέλο SSIM είναι ιδιαίτερα συμβατό με το HVS καθώς και τα δυο συνδέονται με την εξαγωγή δομικών πληροφοριών από μια εικόνα. Πολλές έρευνες και μετρικές βασίστηκαν στον SSIM και στοχεύουν στην εξέλιξή του.

Παρόλα αυτά είναι σημαντικό να τονιστεί το πόσο δύσκολο είναι μια μετρική να εντοπίσει τη σημαντική πληροφορία όπως θα την έκρινε ο φυσικός παρατηρητής. Στο μοντέλο γίνεται η υπόθεση ότι αυτή η πληροφορία στις εικόνες αναπαρίσταται κυρίως από την αντοχή των ακμών κάθε pixel. Η αντοχή των ακμών χρησιμοποιείται για την ποσοτικοποίηση της πιθανότητας ένα εικονοστοιχείο να αποτελεί μέρος μιας ακμής που οριοθετεί ένα αντικείμενο σημαντικής πληροφορίας. Η μετρική ποιότητας εικόνας που προτείνεται βασίζεται στην ομοιότητα της δύναμης των ακμών και λαμβάνει υπόψη τρεις διάφορες πτυχές των χαρακτηριστικών των ακμών για μια πιο ολοκληρωμένη αξιολόγηση (κανονικότητα, ακανόνιστη μορφή, ανισοτροπικές δομές).

1. Την τοπική κανονικότητα κατά μήκος συγκεκριμένης κατεύθυνσης.
2. Την τοπική ανομοιομορφία κατά μήκος της ορθογώνιας κατεύθυνσης.
3. Την δημιουργία ακμών που δεν ακολουθούν συγκεκριμένο προσανατολισμό, προερχόμενες από ανισότροπες δομές.

Η ανισοτροπικότητα αναφέρεται σε μια ιδιότητα που μεταβάλλεται με βάση την κατεύθυνση. Η ακανόνιστη σημαίνει έλλειψη ομοιομορφίας ή προβλεψιμότητας.

Η μετρική αποσκοπεί στην κατανόηση και τη μέτρηση των ακμών στις εικόνες λαμβάνοντας υπόψη και τις αυτές πτυχές - πώς η δύναμη των ακμών ποικίλλει σε διαφορετικές κατευθύνσεις (ανιστροπία) και πόσο ακανόνιστες ή απρόβλεπτες μπορεί να είναι αυτές οι ακμές.

Οι ακμές στις εικόνες δεν είναι πάντα απλές γραμμές- μπορεί να έχουν διαφορετική ισχύ, προσανατολισμό και ακανόνιστα μοτίβα. Λαμβάνοντας υπόψη τόσο την ανιστροπική κανονικότητα (διακυμάνσεις σε διαφορετικές κατευθύνσεις) όσο και την ακανόνιστη (έλλειψη ομοιομορφίας), η μετρική επιχειρεί να συλλάβει αυτή την πολυπλοκότητα των ακμών.

Συνοπτικά η μετρική αξιολογεί την οπτική ευκρίνεια μεταξύ δύο εικόνων συγκρίνοντας τους χάρτες αντοχής ακμών τους, στους οποίους έχει γίνει ενσωμάτωση της κανονικότητας και της μη κανονικότητας της μορφής των ακμών.

Συμβολίζουμε την εικόνα αναφοράς ως  $x = [x_1, \dots, x_i, \dots, x_N] \in R^n$ , Όπου  $i$  δείχνει τα εικονοστοιχεία και  $N$  δηλώνει το συνολικό αριθμό των εικονοστοιχείων.

Στη συνέχεια, προσδιορίζουμε τη δύναμη των ακμών στις διαγώνιες κατευθύνσεις

$$E_i^{2,4}(x) = |\partial x_i^2 - \partial x_i^4|^p . \quad (2.4)$$

Ομοίως, η αντοχή των άκρων στην κατακόρυφη ή οριζόντια κατεύθυνση

$$E_i^{1,3}(x) = |\partial x_i^1 - \partial x_i^3|^p , \quad (2.5)$$

με  $p$  να εισάγεται για να αναβαθμίσει μη γραμμικά την ένταση των ακμών ώστε να προσαρμοστεί στην ανθρώπινη αντίληψη.

Έτσι καταλήγουμε,

$$Ex_i = \max \{E_i^{2,4}(x), E_i^{1,3}(x)\} \quad (2.6)$$

Συμβολίζουμε την παραμορφωμένη εικόνα της  $f$  ως  $y = [y_1, \dots, y_i, \dots, y_N] \in R^n$

Και έχουμε αντίστοιχα,



$$Ey_i = \begin{cases} E_i^{1,3}(y) & \text{εάν } Ex_i = E_i^{1,3}(x) \\ E_i^{2,4}(y) & \text{εάν } Ex_i = E_i^{2,4}(x) \end{cases} \quad (2.7)$$

Επομένως καταλήγουμε στον τελικό τύπο :

$$ESSIM(x, y) = \frac{1}{N} \sum_{i=1}^N \frac{2Ex_iEy_i+C}{(Ex_i)^2+(Ey_i)^2+C} \quad (2.8)$$

$$C = (B255^P)^2 \quad (2.9)$$

Όπου B μια προκαθορισμένη σταθερά.

### 2.2.5 SG\_ESSIM

Το μοντέλο SG\_ESSIM [21] είναι επέκταση του ESSIM και στοχεύει στη εξέλιξή του. Η οπτική ευκρίνεια (visual saliency) αποτελεί σημαντικό παράγοντα για την διεξαγωγή συμπερασμάτων στις μετρικές εκτίμησης ποιότητας εικόνας. Αυτό το χαρακτηριστικό στοχεύει να ενσωματώσει το SG\_ESSIM μοντέλο, ώστε να το συνδυάσει με τον χάρτη αντοχής των ακμών και να βελτιώσει τον ESSIM.

Στην αναφερόμενη έρευνα, η τοπική υποβάθμιση της ποιότητας της εικόνας προσδιορίστηκε τόσο από την αντικειμενική υποβάθμιση της δύναμης των ακμών όσο και από την συνολική σημασία τους. Όταν η ποιότητα μιας εικόνας μειώνεται, οι ακμές μπορούν να γίνουν λιγότερο ευκρινείς, το οποίο θα έχει ως αποτέλεσμα την απώλεια λεπτομερειών. Ορισμένες ακμές έχουν μεγαλύτερη βαρύτητα για την κατανόηση της εικόνας από τον άνθρωπο, επομένως εάν αυτές υποβαθμιστούν θα επηρεαστεί η αντίληψη για την ποιότητα της εικόνας.

Η μετρική SG\_ESSIM περιλαμβάνει τον προσδιορισμό του χάρτη ισχύος των ακμών της αρχικής εικόνας αλλά και της παραμορφωμένης. Επιπρόσθετα υπολογίζει τον τοπικό χάρτη ποιότητας εικόνας με βάση την οπτική ευκρίνεια. Τέλος, γίνεται ενοποίηση αυτών των χαρτών για την εξαγωγή της συνολικής αντιληπτικής ποιότητας της παραμορφωμένης εικόνας, δίνοντας έμφαση στις περιοχές με ισχυρότερες ακμές ως πιο ευδιάκριτες για την ανθρώπινη οπτική αντίληψη. Συνοπτικά, έγινε ενσωμάτωση της οπτικής ευκρίνειας στην αξιολόγηση της τοπικής ποιότητας της εικόνας, παρέχοντας μια πιο ολοκληρωμένη μέθοδο

αξιολόγησης που λαμβάνει υπόψη τόσο τα χαρακτηριστικά των ακμών όσο και την ανθρώπινη οπτική αντίληψη.

Όμοια με ESSIM συμβολίζουμε την εικόνα αναφοράς ως  $x = [x_1, \dots, x_i, \dots, x_N] \in R^n$

,  
 όπου  $i$  δείχνει τα εικονοστοιχεία και  $N$  δηλώνει το συνολικό αριθμό των εικονοστοιχείων.

Στη συνέχεια, προσδιορίζουμε τη δύναμη των ακμών στις διαγώνιες κατευθύνσεις

$$E_i^{2,4}(x) = |\partial x_i^2 - \partial x_i^4|^p . \quad (2.10)$$

Ομοίως, η αντοχή των άκρων στην κατακόρυφη ή οριζόντια κατεύθυνση

$$E_i^{1,3}(x) = |\partial x_i^1 - \partial x_i^3|^p, \quad (2.11)$$

με  $p$  να εισάγεται για να αναβαθμίσει μη γραμμικά την ένταση των ακμών ώστε να προσαρμοστεί στην ανθρώπινη αντίληψη.

Έτσι καταλήγουμε,

$$Ex_i = \max\{E_i^{2,4}(x), E_i^{1,3}(x)\} \quad (2.12)$$

Συμβολίζουμε την παραμορφωμένη εικόνα της  $f$  ως  $y = [y_1, \dots, y_i, \dots, y_N] \in R^n$

Και έχουμε αντίστοιχα,

$$Ey_i = \begin{cases} E_i^{1,3}(y) & \text{εάν } Ex_i = E_i^{1,3}(x) \\ E_i^{2,4}(y) & \text{εάν } Ex_i = E_i^{2,4}(x) \end{cases} \quad (2.13)$$

Επομένως καταλήγουμε :

$$SM(Ex_i, Ey_i) = \frac{2Ex_iEy_i + H(V_i)}{(Ex_i)^2 + (Ey_i)^2 + H(V_i)} \quad (2.14)$$

$$H(V_i) = Ke^{-V(i)/h} \quad (2.15)$$

$$V_i = \max \{Ex_i, Ey_i\} \quad (2.16)$$

όπου  $e$  είναι ο αριθμός του Euler,  $K$  είναι ένας παράγοντας κλιμάκωσης και  $h$  είναι ένας παράγοντας εξασθένησης.

Με τελικό τύπο :

$$SG - ESSIM = \frac{1}{N} \sum_{i=1}^N SM(Ex_i, Ey_i) \quad (2.17)$$

### 2.2.6 VCGS

Το VCGS (Visual Saliency With Color Appearance and Gradient Similarity) [DC] είναι ένα IQA μοντέλο που συνδυάζει την οπτική ευκρίνεια (Visual saliency) με την ομοιότητα εμφάνισης χρώματος, την ομοιότητα κλίσης και την χρωματική ομοιότητα.

Δύο νέοι δείκτες εμφάνισης χρώματος, η ζωντάνια και το βάθος, λειτουργούν με Φίλτρο Log-Gabor και επιλέγονται για τη δημιουργία του χάρτη ομοιότητας οπτικής ευκρίνειας VC. Ο χάρτης VC είναι παράγοντας κρίσιμης σημασίας για δύο λόγους. Πρώτον, καθώς χρησιμοποιείται ως διακριτικό για τον υπολογισμό της τοπικής ποιότητας της παραμορφωμένης εικόνας και δεύτερον, διότι λειτουργεί ως συνάρτηση βάρους για να αντικατοπτρίζει τη σημασία του τοπικού τομέα.

Δημιουργούνται τρεις χάρτες ομοιότητας. Το χαρακτηριστικό Visual saliency εξάγεται από την εμφάνιση χρώματος στον χρωματικό χώρο CIELAB, ο οποίος είναι πιο συμβατός με την ανθρώπινη διαίσθηση. Οι chrominance similarity maps μπορούν να εκφράσουν τη χρωματική διαφορά μεταξύ δύο εικόνων σε επίπεδο pixel από τα κανάλια χρωμάτων. Η κλίση χρησιμοποιείται για δομικό χαρακτηριστικό για να προκύψει ο τελευταίος χάρτης ομοιότητας.

Αρχικά, η οπτική ευκρίνεια με την εμφάνιση χρώματος (VC) της αρχικής εικόνας  $X$  και της παραμορφωμένης εικόνας  $Y$  μπορεί να υπολογιστεί ως εξής:

$$VC = ((V * g)^2 + (D * g)^2)^{1/2} \quad (2.18)$$

όπου το \* δίνει την πράξη συνέλιξης, το g προκύπτει μόνο κατά προσέγγιση εκτελώντας μια αριθμητική αντίστροφη μετατροπή Fourier, ζωντάνια V και βάθος D.

Στη συνέχεια, η οπτική ομοιότητα εμφάνισης χρώματος  $S_{VC}$  για τις δύο εικόνες αντιπαραβάλλεται ως:

$$S_{VC} = \frac{2VC_1VC_2+K_{VC}}{VC_1^2+VC_2^2+K_{VC}} \quad (2.19)$$

Όπου η παράμετρος  $K_{VC}$  είναι μια σταθερά για τον έλεγχο της αριθμητικής σταθερότητας και  $VC_1 VC_2$  (VC) της αρχικής εικόνας και της παραμορφωμένης εικόνας αντίστοιχα.

Η ομοιότητα κλίσης (SG) υπολογίζεται με την ακόλουθη επαγόμενη SSIM εξίσωση:

$$S_G = \frac{2G_1G_2+K_G}{G_1^2+G_2^2+K_G} \quad (2.20)$$

Όπου η παράμετρος  $K_G$  είναι μια σταθερά για τον έλεγχο της αριθμητικής σταθερότητας,  $G_1$  και  $G_2$  είναι τα μεγέθη των κλίσεων του καναλιού φωτεινότητας ( $L^*$ ) στον χρωματικό χώρο CIELAB για την αρχική και την παραμορφωμένη εικόνα, αντίστοιχα.

Η ομοιότητα των συνιστωσών χρωματισμού  $S_C$  στο CIELAB χρωματικό χώρο, δηλαδή, κόκκινο-πράσινο  $a^*$  και μπλε-κίτρινο  $b^*$  κανάλια, μπορούν απλά να οριστούν ως εξής:

$$S_C = \frac{2a_1a_2+K_C}{a_1^2+a_2^2+K_C} \frac{2b_1b_2+K_C}{b_1^2+b_2^2+K_C} \quad (2.21)$$

Όπου η παράμετρος  $K_C$  είναι μια σταθερά για τον έλεγχο της αριθμητικής σταθερότητας,  $G_1$  και  $G_2$  είναι τα μεγέθη των κλίσεων του καναλιού φωτεινότητας ( $L^*$ ) στον χρωματικό χώρο CIELAB για την αρχική και την παραμορφωμένη εικόνα, αντίστοιχα.

Με τελικό τύπο :

$$VCGS = \frac{\sum_{\Omega} S_{VC}(S_G)^a (S_C)^{\lambda} VC_m}{\sum_{\Omega} VC_m} \quad (2.22)$$

Όπου  $\Omega$  σημαίνει το χωρικό πεδίο και

$$VC_m = \max (VC_1, VC_2) \quad (2.23)$$

### 2.2.7 VSI

Στο μοντέλο προτείνεται ένας ιδιαίτερα αποτελεσματικός δείκτης με βάση την οπτική ευκρίνεια VS (VSI). Η οπτική ευκρίνεια διαδραματίζει καθοριστικό ρόλο για το VSI (visual saliency-based index) [23]. Αποτελεί παράγοντα δημιουργίας τοπικού χάρτη ομοιότητας και εφαρμόζεται ως συνάρτηση βάρους ώστε να κατανοήσουμε τη σημαντικότητα μιας τοπικής περιοχής. Όπως απέδειξαν οι Zhang et al, εάν η ποιότητα μιας εικόνας είναι ορατά μειωμένη, αυτό δεν σημαίνει μόνο μια γενική απώλεια ευκρίνειας, αλλά σημαίνει επίσης ότι συγκεκριμένα οπτικά στοιχεία (χαρακτηριστικά χαμηλού επιπέδου) όπως το χρώμα, η αντίθεση και η υφή μπορεί να υποστούν αξιοσημείωτες αλλαγές που μπορούν να παρατηρηθούν από τους ανθρώπους. Η οπτική ευκρίνεια (Visual Saliency, VS) δεν έχει καλή ευαισθησία στην αντίθεση. Με χρήση του gradient modulus (GM) για αντιστάθμιση της έλλειψης αυτής της ευαισθησίας σε αυτή την εργασία γίνεται προσπάθεια να αντιμετωπιστεί το πρόβλημα.

Ορίζονται οι μερικές παράγωγοι  $G_x$  και  $G_y$  για την εικόνα χρησιμοποιώντας τον τελεστή Scharr. Έπειτα γίνεται ο υπολογισμός Gradient Modulus (GM) ως:

$$GM = ((G_x)^2 + (G_y)^2)^{1/2} \quad (2.24)$$

Οι χρωματικές παραμορφώσεις δεν μπορούν να χαρακτηριστούν καλά από το gradient, οπότε μετατρέπουμε τις εικόνες RGB σε έναν αντίθετο χρωματικό χώρο:

$$\begin{bmatrix} L \\ M \\ N \end{bmatrix} = \begin{bmatrix} 0.06 & 0.63 & 0.27 \\ 0.30 & 0.04 & -0.35 \\ 0.34 & -0.60 & 0.17 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (2.25)$$

Τα κανάλια L, M και N χρησιμοποιούνται για την περιγραφή της ποιότητας της εικόνας με χρωματικές παραμορφώσεις.

Για δύο εικόνες x και y, εξάγουμε τους χάρτες VS (VS1 και VS2), τους χάρτες GM (G1 και G2), και τους χρωματικούς χάρτες (M1, M2 και N1, N2).

Ομοιότητα VS (SVS)

$$S_{VS} = \frac{2VS_1VS_2+C_1}{VS_1^2+VS_2^2+C_1} \quad (2.26)$$

Ομοιότητα GM (SG)

$$S_G = \frac{2G_1G_2+C_2}{G_1^2+G_2^2+C_2} \quad (2.27)$$

Χρωματική Ομοιότητα (SC)

$$S_C = \frac{2M_1M_2+C_3}{M_1^2+M_2^2+C_3} \frac{2N_1N_2+C_3}{N_1^2+N_2^2+C_3} \quad (2.28)$$

Ο συνολικός τοπικός χάρτης ομοιότητας S υπολογίζεται ως:

$$S = S_{VS}S_G^\alpha S_C^\beta \quad (2.29)$$

Όπου α και β είναι παράμετροι που ρυθμίζουν τη σχετική σημασία των χαρακτηριστικών VS, GM και χρωματικότητας.

Έπειτα για τον τελικό τύπο χρησιμοποιείται το

$$VS_m = \max (VS_1, VS_2) \quad (2.30)$$

για να υπολογιστεί η βαρύτητα του S, δηλαδή της τοπικής ομοιότητας, σε σχέση με τη συνολική ομοιότητα.

Επομένως η συνολική ομοιότητα υπολογίζεται ως:

$$VSI = \frac{\sum_{xe\Omega} SVS_m}{\sum_{xe\Omega} VS_m} \quad (2.31)$$

Όπου  $\Omega$  συμβολίζει ολόκληρο το χωρικό πεδίο.

## 2.2.8 SSIM

Η ανθρώπινη οπτική αντίληψη είναι ιδιαίτερα συνυφασμένη με την εξαγωγή δομικών πληροφοριών από μια εικόνα και γενικότερα από ένα τοπίο. Βασιζόμενοι σε αυτό το γεγονός, παρουσιάστηκε ένας δείκτης δομικής ομοιότητας SSIM (Structral SIMilarity) [24]. Ο δείκτης αυτός μετράει πόσο αλλοιώνονται οι δομικές πληροφορίες μεταξύ των εικόνων, με στόχο να ανταποκρίνεται καλύτερα με το ανθρώπινο οπτικό σύστημα. Αντί να εκτιμά τα αντιληπτά σφάλματα που οφείλονται σε υποβαθμίσεις της εικόνας, θεωρεί αυτές τις υποβαθμίσεις ως μεταβολές στις δομικές πληροφορίες. Οι φυσικές εικόνες έχουν ισχυρές εξαρτήσεις μεταξύ των εικονοστοιχείων, ειδικά εκείνων που βρίσκονται χωρικά κοντά. Αυτές οι εξαρτήσεις περιέχουν κρίσιμες πληροφορίες.

Ο δείκτης SSIM αποσκοπεί στην ποσοτικοποίηση της ομοιότητας μεταξύ δύο σημάτων εικόνας με βάση τις δομικές πληροφορίες τους, εστιάζοντας στη φωτεινότητα, την αντίθεση και τις δομικές συγκρίσεις. Η μέθοδος αυτή υπολογίζει τις συγκρίσεις με βάση τα μήκη και τις γωνίες των διανυσμάτων, ευθυγραμμίζοντας τα περιγράμματα με ένα πολικό σύστημα συντεταγμένων. Λαμβάνει υπόψη τις προσαρμογές μεγέθους και σχήματος στα περιγράμματα με βάση τα υποκείμενα χαρακτηριστικά του σήματος, σε αντίθεση με τα παραδοσιακά μοντέλα που μπορεί να βασίζονται αποκλειστικά στο μέγεθος του σήματος. Εάν οι εικόνες είναι πανομοιότυπες, ο SSIM θα δώσει την τιμή 1, υποδεικνύοντας τέλεια ταύτιση. Όσο πιο κοντά είναι η τιμή στο 1, τόσο πιο παρόμοιες θεωρούνται οι εικόνες όσον αφορά τις δομικές πληροφορίες.

Για τη σύγκριση της φωτεινότητας, ορίζουμε:

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \quad (2.32)$$

Η συνάρτηση σύγκρισης αντίθεσης έχει παρόμοια μορφή:

$$c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad (2.33)$$

Επιπρόσθετα, ορίζουμε τη συνάρτηση σύγκρισης δομών ως εξής:

$$s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \quad (2.34)$$

Ο δείκτης SSIM συνδυάζει τα τρία μέτρα ομοιότητας: τη φωτεινότητα (l), την αντίθεση (c) και τη δομή (s) και ορίζεται ως εξής:

$$SSIM(x, y) = l(x, y)^\alpha + c(x, y)^\beta + s(x, y)^\gamma \quad (2.35)$$

όπου οι παράμετροι  $\alpha > 0$ ,  $\beta > 0$  και  $\gamma > 0$  καθορίζουν τη σχετική σημασία των τριών συνιστωσών.

Για να απλοποιηθεί η έκφραση, οι τέθηκαν  $\alpha = \beta = \gamma = 1$ , κάτι που σημαίνει ότι οι τρεις συνιστώσες έχουν ίση σημασία.

Σταθερές για αριθμητική σταθερότητα: Επίσης, τέθηκε η σταθερά  $C_3 = C_2/2$  για να απλοποιηθεί περαιτέρω η έκφραση και να εξασφαλιστεί η αριθμητική σταθερότητα.

Αυτό έχει ως αποτέλεσμα μια ειδική μορφή του δείκτη :

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{x,y} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (2.36)$$

Όπου  $C_1, C_2$  κατάλληλες σταθερές που χρησιμοποιούνται για να αποφευχθεί η αστάθεια όταν οι τιμές των παρονομαστών πλησιάζουν το μηδέν, η πληροφορία φωτεινότητας κάθε εικόνας αντιπροσωπεύεται από το μέσο όρο  $\mu$ , τυπική απόκλιση  $\sigma$ , συνδιακύμανση  $\sigma_{xy}$ .



### 2.2.9 MS\_SSIM

Το μοντέλο MS\_SSIM(multi-scale SSIM) αποτελεί συνέχεια του SSIM και αποσκοπεί στην κάλυψη ορισμένων κενών που είχαν δημιουργηθεί. Ο δείκτης SSIM είναι μια μετρική που χρησιμοποιείται για την αξιολόγηση της ομοιότητας μεταξύ δύο εικόνων. Μετράει πόσο κοντά είναι οι δομικές πληροφορίες μέσα σε αυτές τις εικόνες σε σχέση με αυτό που αντιλαμβάνεται το ανθρώπινο μάτι. Ο περιορισμός είναι ότι ο δείκτης SSIM είναι μια προσέγγιση μιας κλίμακας.

Στην ουσία, η προσέγγιση μίας κλίμακας σημαίνει ότι ο δείκτης SSIM λειτουργεί σε μία μόνο συγκεκριμένη κλίμακα ή επίπεδο λεπτομέρειας κατά τη σύγκριση εικόνων. Δεν λαμβάνει υπόψη τις μεταβολές στις συνθήκες θέασης, όπως οι αλλαγές στην ανάλυση της οθόνης ή οι διαφορές στις αποστάσεις θέασης. Για να είναι ακριβής η αξιολόγηση της ποιότητας της εικόνας σε διαφορετικές συνθήκες προβολής, είναι ζωτικής σημασίας να λαμβάνονται υπόψη πολλαπλές κλίμακες ή επίπεδα λεπτομέρειας. Ως εκ τούτου, μια προσέγγιση πολλαπλών κλιμάκων, όπως το μοντέλο MS\_SSIM, αποσκοπεί στη διόρθωση αυτού του μειονεκτήματος με την ενσωμάτωση και την εξέταση πολλαπλών κλιμάκων για μια πιο ολοκληρωμένη αξιολόγηση της ποιότητας της εικόνας.

Η μέθοδος SSIM πολλαπλών κλιμάκων περιλαμβάνει μια επαναληπτική διαδικασία: ξεκινώντας από την αρχική εικόνα (Κλίμακα 1), εφαρμόζει ένα χαμηλοπερατό φίλτρο και υποβαθμίζει την εικόνα κατά έναν παράγοντα 2. Η διαδικασία αυτή συνεχίζεται επαναληπτικά, δημιουργώντας διαφορετικές κλίμακες μέχρι την Κλίμακα M. Σε κάθε κλίμακα, εκτελούνται υπολογισμοί για τη σύγκριση αντίθεσης και τη σύγκριση δομής. Η σύγκριση φωτεινότητας υπολογίζεται μόνο στην υψηλότερη κλίμακα (κλίμακα M). Ωστόσο, ένα από τα πιο δύσκολα προβλήματα είναι η ρύθμιση των παραμέτρων του μοντέλου γι' αυτό και θεωρείται ότι χρήζει βελτίωσης.

Παρόμοια με παραπάνω, τώρα όμως η συνολική αξιολόγηση SSIM προκύπτει από το συνδυασμό των μετρήσεων σε διαφορετικές κλίμακες χρησιμοποιώντας :

$$SSIM(x, y) = l_M(x, y)^{\alpha M} \prod_{j=1}^M c_j(x, y)^{\beta j} s_j(x, y)^{\gamma j} \quad (2.37)$$

Στην  $j$ -οστή κλίμακα, η σύγκριση αντίθεσης και η σύγκριση δομής υπολογίζονται και συμβολίζονται ως  $c_j(x, y)$  και  $s_j(x, y)$ , αντίστοιχα. Η σύγκριση φωτεινότητας υπολογίζεται μόνο στην κλίμακα  $M$  και συμβολίζεται ως  $l_M(x, y)$ . Οι εκθέτες  $\alpha_M$ ,  $\beta_j$  και  $\gamma_j$  χρησιμοποιούνται για την προσαρμογή της σχετικής σημασίας των διαφόρων συνιστωσών.

### 2.2.10 GMSD

Το μοντέλο GMSD (Gradient Magnitude Similarity Deviation) [24] αξιοποιεί τις κλίσεις των εικόνων, οι οποίες είναι ιδιαίτερα ευαίσθητες στις παραμορφώσεις. Εξετάζει τις διακυμάνσεις στις τοπικές δομές εντός των παραμορφωμένων εικόνων, χρησιμοποιώντας τον υπολογισμό της ομοιότητας του μεγέθους κλίσης (GMS) μεταξύ της αρχικής εικόνας και της παραμορφωμένης. Για τη συνολική εκτίμηση της ποιότητας, χρησιμοποιείται μια στρατηγική συγκέντρωσης βαθμολογίας με βάση την τυπική απόκλιση του χάρτη GMS. Ο αλγόριθμος GMSD επιτρέπει τη γρήγορη και αποτελεσματική πρόβλεψη της ποιότητας της εικόνας. Αυτό μας δείχνει ότι η ποικιλομορφία των τοπικών δομών στην εικόνα συνδέεται με τις μεταβολές στην τοπική ποιότητα, η οποία έχει σημαντική επίδραση στις υποκειμενικές αξιολογήσεις της ποιότητας της εικόνας.

Αρχικά, ορίζονται τα φίλτρα Prewitt κατά μήκος της οριζόντιας ( $x$ ) και της κάθετης ( $y$ ) κατεύθυνσης ως εξής:

$$h_x = \begin{bmatrix} 1/3 & 0 & -1/3 \\ 1/3 & 0 & -1/3 \\ 1/3 & 0 & -1/3 \end{bmatrix} \quad (2.38)$$

$$h_y = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 0 & 0 & 0 \\ -1/3 & -1/3 & -1/3 \end{bmatrix} \quad (2.39)$$

*Συνέλιξη των  $h_x$  και  $h_y$  με τις εικόνες αναφοράς και τις παραμορφωμένες εικόνες*

δίνει τις εικόνες οριζόντιας και κάθετης κλίσης των  $r$  και  $d$ . Τα μεγέθη κλίσης των  $r$  και  $d$  στη θέση  $i$ , που συμβολίζονται με  $m_r(i)$  και  $m_d(i)$ , υπολογίζονται ως εξής:

$$m_r(i) = \sqrt{(r * h_x)^2(i) + (r * h_y)^2(i)} \quad (2.40)$$

$$m_d(i) = \sqrt{(d * h_x)^2(i) + (d * h_y)^2(i)} \quad (2.41)$$

Με τις εικόνες μεγέθους κλίσης  $m_r(i)$  και  $m_d(i)$  διαθέσιμες, ο χάρτης ομοιότητας μεγέθους κλίσης (GMS) υπολογίζεται ως εξής:

$$GMS(i) = \frac{2m_r(i)m_d(i) + c}{m_r^2(i) + m_d^2(i) + c} \quad (2.42)$$

Εφαρμόζοντας μέσο όρο στον χάρτη GMS, ονομάζουμε το μοντέλο IQA ως Μέσος Όρος Ομοιότητας Μεγέθους Κλίσης (GMSM) και ο τύπος υπολογισμού του GMSM είναι ο εξής:

$$GMSM = \frac{1}{N} \sum_{i=1}^N GMS(i) \quad (2.43)$$

Τέλος, υπολογίζουμε την τυπική απόκλιση του χάρτη GMS και την λαμβάνουμε ως τον τελικό δείκτη IQA, δηλαδή έχουμε το Gradient Magnitude Similarity Deviation (GMSD):

$$GMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N (GMS(i) - GMSM)^2} \quad (2.44)$$

## 2.2.11 Πίνακας Συγκέντρωσης Όλων Των Μετρικών

Πίνακας 2.2: Κατηγορίες Αντικειμενικών Τεχνικών Πλήρους Αναφοράς

Κατηγορία	Μετρικές	Θετικά	Αρνητικά
Βάσει των Στατιστικών των Pixel	<ul style="list-style-type: none"> <li>• MSE</li> <li>• PSNR</li> </ul>	Εύκολη εφαρμογή με χαμηλό υπολογιστικό κόστος και χαμηλή πολυπλοκότητα.	Μη συμβατό με το HVS.
Βάσει δομικών Παραμορφώσεων	<ul style="list-style-type: none"> <li>• MDSI</li> <li>• SSIM</li> <li>• MS_SSIM</li> <li>• GMSD</li> </ul>	Συμβατό με HVS.	Υψηλή υπολογιστική πολυπλοκότητα. Λειτουργεί καλά μόνο σε σωστή κλίμακα, συνθήκες θέασης και ανάλυση της οθόνης.
Βάσει της οπτικής ευκρίνειας	<ul style="list-style-type: none"> <li>• VSI</li> <li>• VCGS</li> <li>• ESSIM</li> <li>• SG_ESSIM</li> </ul>	Συμβατό με HVS.	Πολύ υψηλή υπολογιστική πολυπλοκότητα.

## 2.3 Μετασχηματισμοί Fourier και wavelet για την αξιολόγηση ποιότητας εικόνας

Ο μετασχηματισμός Fourier (FT) και ο μετασχηματισμός wavelet χρησιμεύουν ως θεμελιώδη εργαλεία στην ανάλυση σημάτων, ωστόσο προσεγγίζουν το έργο αυτό με διαφορετικές μεθοδολογίες και λειτουργίες.

Ο μετασχηματισμός Fourier (FT) [25] αποτελεί θεμελιώδες εργαλείο σε πολλούς τομείς, όπως η ανάλυση κυκλωμάτων, ο σχεδιασμός φίλτρων, η επεξεργασία σήματος και η ανακατασκευή εικόνας. Ευρέως αναγνωρισμένος για την ικανότητά του να αναλύει τα σήματα στις συνιστώσες συχνότητάς τους, ο FT λειτουργεί κυρίως σε ενεργειακά σήματα με πεπερασμένη ενέργεια. Βρίσκει εφαρμογή μέσω των διακριτών μορφών του - του διακριτού μετασχηματισμού Fourier (DFT) και του γρήγορου μετασχηματισμού Fourier (FFT) - οι οποίες χρησιμεύουν ως βασικά διακριτά εργαλεία για την ανάλυση σημάτων στο πεδίο του χρόνου. Ωστόσο, η διαδικασία διακριτοποίησης σε αυτούς τους μετασχηματισμούς εισάγει ορισμένα ζητήματα, γεγονός που καθιστά αναγκαία τη σαφή κατανόηση των σχετικών προβλημάτων και των προδιαγραφών των παραμέτρων για την αποφυγή μη φυσικών ή μη μαθηματικών αποτελεσμάτων.

Μαθηματικά, ο FT ορίζεται για σήματα συνεχούς χρόνου και απαιτεί την παρατήρηση του χρονικού σήματος σε ένα άπειρο-επεκτεινόμενο χρονικό παράθυρο για την ακριβή μετάβαση στο πεδίο της συχνότητας. Ορίζουμε ως μετασχηματισμό Fourier μίας συνάρτησης  $f(t)$  τη μιγαδική συνάρτηση πραγματικής μεταβλητής  $f(\omega)$ , που δίνεται από τη σχέση:

$$f(\omega) = \int_{-\infty}^{\infty} f(t)e^{-i\omega t} dt \quad (2.45)$$

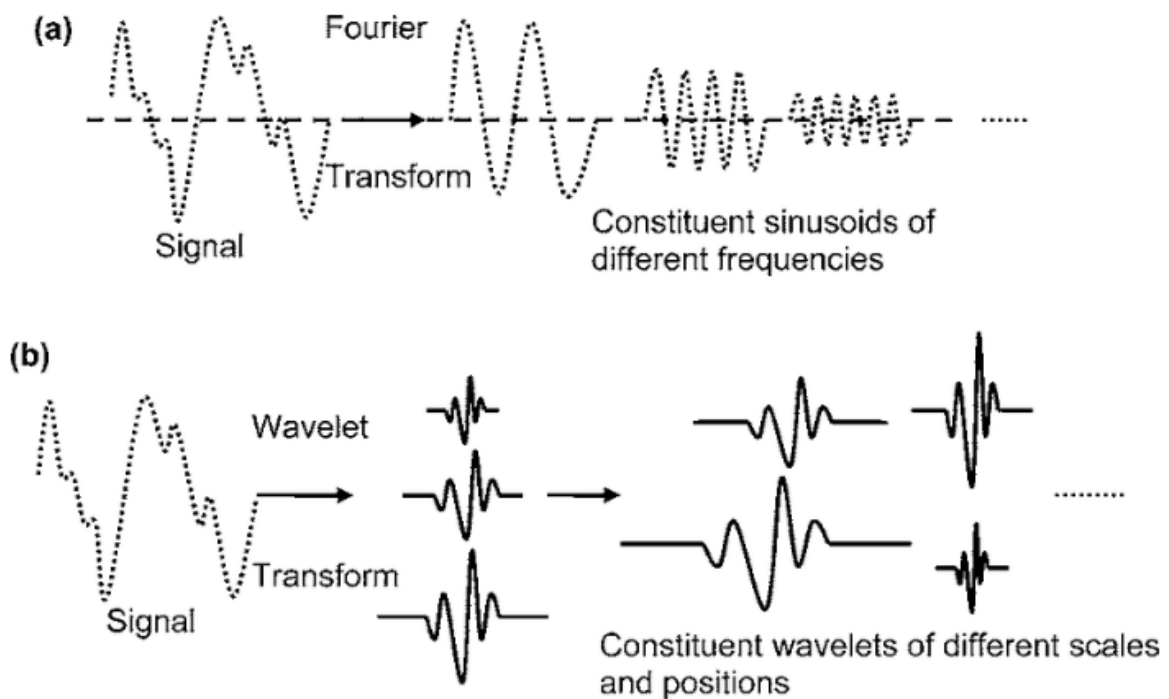
Αντίθετα, ο μετασχηματισμός κυματιδίων [26], συγγενής με τον FT, προβάλλει τα σήματα σε ένα σύνολο συναρτήσεων βάσης για να τα αναλύσει στο πεδίο της συχνότητας. Ωστόσο, αποκλίνει σημαντικά ως προς τη λειτουργικότητά του. Σε αντίθεση με την ομοιόμορφη, σταθερή ανάλυση χρόνου-συχνότητας του FT, ο μετασχηματισμός κυματιδίων παρουσιάζει ένα ιδιαίτερο χαρακτηριστικό - ακρίβεια υψηλής συχνότητας στις χαμηλές συχνότητες και αυξημένη χρονική

ανάλυση στις υψηλές συχνότητες. Για την επίτευξη αυτής της προσαρμοστικής ανάλυσης, ο μετασχηματισμός wavelet χρησιμοποιεί ένα σύνολο ορθογώνιων βάσεων με μεταβλητή ανάλυση. Αυτό επιτυγχάνεται με την επέκταση και τη μετάφραση της συνάρτησης βάσης wavelet, επιτρέποντας μια πιο λεπτομερή και εντοπισμένη αναπαράσταση ή προσέγγιση ενός σήματος.

Ο μετασχηματισμός wavelet, που θεωρείται μια πρωτοποριακή εξέλιξη στη μαθηματική ανάλυση, προσφέρει ευελιξία στην εφαρμογή του σε πολυάριθμους τομείς. Η προσαρμοστικότητά του επιτρέπει την περίπλοκη ανάλυση στην επεξεργασία σήματος, στην επεξεργασία εικόνας, στην αναγνώριση προτύπων, στην ανάλυση ομιλίας και σε διάφορους άλλους τομείς που επωφελούνται από την εντοπισμένη κατανόηση του σήματος.

Στην ουσία, ενώ ο FT παρέχει μια τυποποιημένη οπτική των συχνοτήτων του σήματος, η προσαρμοστικότητα και η εντοπισμένη ανάλυση του μετασχηματισμού κυματιδίων δίνουν τη δυνατότητα μιας ολοκληρωμένης κατανόησης των σημάτων σε διάφορες εφαρμογές, χαρακτηρίζοντάς τον ως ένα ισχυρό εργαλείο στη μαθηματική ανάλυση και την επεξεργασία σήματος.

Συνοψίζοντας, εάν το σήμα που μας ενδιαφέρει είναι σταθερό και οι συνιστώσες συχνότητάς του είναι σταθερές καθ' όλη τη διάρκεια του χρόνου, η FT μπορεί να είναι καταλληλότερη λόγω της ομοιόμορφης ανάλυσης. Ωστόσο, για σήματα που δεν είναι στάσιμα, μεταβατικά ή παρουσιάζουν αλλαγές στη συχνότητα με την πάροδο του χρόνου, ο μετασχηματισμός Wavelet τείνει να είναι πιο χρήσιμος, καθώς παρέχει μια πιο λεπτομερή και εντοπισμένη αναπαράσταση τέτοιων σημάτων. Επομένως, η επιλογή μεταξύ FT και μετασχηματισμού Wavelet εξαρτάται από την ειδική φύση και τις απαιτήσεις του σήματος που αναλύεται και τους στόχους της ανάλυσης. Και οι δύο μετασχηματισμοί έχουν τα δυνατά τους σημεία και αποτελούν πολύτιμα εργαλεία σε διαφορετικά σενάρια.



**EIKONA 2.9:** (α) Μετασχηματισμός Fourier ενός σήματος.(β) Μετασχηματισμός wavelet ενός σήματος.

Όπως φαίνεται στην EIKONA στην ανάλυση Fourier, τα σήματα αναλύονται σε ημιτονοειδή διαφορετικών συχνοτήτων, γεγονός που επιτρέπει την ανάλυση στο πεδίο της συχνότητας, αλλά χάνει πληροφορίες σχετικά με το πότε συνέβησαν τα γεγονότα στο πεδίο του χρόνου [27].

Ωστόσο, πολλά σήματα παρουσιάζουν χαρακτηριστικά όπως ολίσθηση, τάσεις, απότομες αλλαγές και γεγονότα που αρχίζουν και τελειώνουν, τα οποία είναι ζωτικής σημασίας για την ανάλυση αλλά δεν μπορούν να αποτυπωθούν καλά μόνο με την ανάλυση Fourier. Η ανάλυση κυματιδίων αντιμετωπίζει αυτόν τον περιορισμό με τη χρήση μιας τεχνικής παραθύρων με περιοχές μεταβλητού μεγέθους, γνωστές ως κυματίδια. Αυτά τα wavelets επιτρέπουν την ανάλυση σημάτων σε διαφορετικές χρονικές κλίμακες, επιτρέποντας την ανίχνευση πληροφοριών τόσο χαμηλής όσο και υψηλής συχνότητας με ακρίβεια.

Στην ουσία, ενώ η ανάλυση Fourier είναι αποτελεσματική για στάσιμα σήματα, η ανάλυση wavelet είναι πιο κατάλληλη για σήματα με μη στάσιμα ή μεταβατικά χαρακτηριστικά, παρέχοντας μια πιο ολοκληρωμένη κατανόηση της συμπεριφοράς του σήματος με την πάροδο του χρόνου.

Στο πλαίσιο της αξιολόγησης της ποιότητας της εικόνας η ικανότητα του μετασχηματισμού Wavelet να συλλαμβάνει τοπικές λεπτομέρειες και παραλλαγές μπορεί να είναι πιο χρήσιμη. Η προσαρμοστικότητα του στην αξιολόγηση διαφορετικών αναλύσεων συχνότητας εντός της εικόνας μπορεί να είναι ιδιαίτερα επωφελής για την αξιολόγηση συγκεκριμένων πτυχών όπως η υφή, η ευκρίνεια ή η δομή -παράγοντες κρίσιμοι για την αξιολόγηση της ποιότητας της εικόνας. Ο μετασχηματισμός Wavelet προτιμάται συχνά λόγω της ικανότητάς του να παρέχει μια πιο λεπτομερή και εντοπισμένη ανάλυση, προσφέροντας πληροφορίες για διάφορες πτυχές της ποιότητας μιας εικόνας.

## 2.4 Κριτήρια αξιολόγησης (Αποτίμησης)

Προκειμένου να αξιολογηθεί κατά πόσο ένα μοντέλο είναι σε θέση να προβλέψει την αντίληψη των ανθρώπινων παρατηρητών, οι συγκρίσεις γίνονται μεταξύ των βαθμολογιών που υπολογίζονται με τη χρήση του προτεινόμενου μοντέλου και των τιμών που βαθμολογούνται από τους παρατηρητές. Συνήθως χρησιμοποιούνται τέσσερα κριτήρια αξιολόγησης για τα μοντέλα IQA:

- ✓ ο συντελεστής κατάταξης Spearman (SROCC),
- ✓ ο συντελεστής γραμμικής συσχέτισης Pearson (PLCC),
- ✓ ο συντελεστής κατάταξης Kendall (KROCC) και
- ✓ το μέσο τετραγωνικό σφάλμα (RMSE).

Ο **SROCC** μετρά τη μονοτονία της πρόβλεψης, δηλαδή τον βαθμό στον οποίο το μοντέλο συμφωνεί με την κατάταξη των υποκειμενικών αξιολογήσεων, και ορίζεται ως εξής:

$$SROCC = 1 - \frac{6 \sum_{i=1}^n [R(p_i) - R(s_i)]^2}{n(n^2 - 1)} \quad (2.46)$$

όπου το  $R(p_i)$  και το  $R(s_i)$  αντιπροσωπεύουν τις βαθμολογίες των προβλέψεων  $p_i$  και των υποκειμενικών βαθμολογιών  $s_i$ , αντίστοιχα, και  $n$  είναι ο αριθμός των



συνόλων βαθμολογιών. Ο **KROCC** είναι επίσης ένας συντελεστής συσχέτισης που μετρά τη μονοτονία της πρόβλεψης, και δίνεται από :

$$KROCC = \frac{n_c - n_d}{\frac{n(n-1)}{2}} \quad (2.47)$$

όπου το  $n_c$  και το  $n_d$  εκφράζουν τον αριθμό των σύμφωνων και των αντίθετων ζευγών στις βαθμολογίες, αντίστοιχα. Τόσο ο SROCC όσο και ο KROCC λαμβάνουν υπόψη την κατάταξη της βαθμολογίας και παραβλέπουν την σχετική απόσταση μεταξύ των βαθμολογιών, κάτι που διαφέρει από τον ακόλουθο PLCC. Και οι τρεις αυτοί συντελεστές κυμαίνονται από -1 έως 1, και η απόλυτη τιμή καθενός κοντά στο 1 σημαίνει ότι η πιστότητα ενός αντικειμενικού μοντέλου θεωρείται υψηλή. Για να υπολογιστούν οι δείκτες **PLCC** και **RMSE**, υιοθετείται μια λογιστική παλινδρόμηση για να ληφθούν οι ίδιες κλίμακες τιμών με τις υποκειμενικές κρίσεις χρησιμοποιώντας:

$$p(x) = \beta_1 \left[ \frac{1}{2} - \frac{1}{1 + \exp(\beta_2(x - \beta_3))} \right] + \beta_4 x + \beta_5 \quad (2.48)$$

όπου τα  $\beta_1, \dots, \beta_5$  είναι οι παράμετροι που πρέπει να προσαρμοστούν, το  $x$  αντιπροσωπεύει τις αρχικές βαθμολογίες IQA, και το  $p(x)$  είναι η βαθμολογία IQA μετά την παλινδρόμηση. Ο PLCC είναι ένας συντελεστής συσχέτισης που χρησιμοποιείται για τη μέτρηση της ακρίβειας πρόβλεψης ενός μοντέλου, δηλαδή της ικανότητας πρόβλεψης των υποκειμενικών αξιολογήσεων με χαμηλό σφάλμα. Για η ζεύγη μοντέλων και υποκειμενικών βαθμολογιών ( $p_i, s_i$ ), ο PLCC υπολογίζεται ως εξής:

$$PLCC = \frac{\sum_{i=1}^n (p_i - \hat{p})(s_i - \hat{s})}{\sqrt{\sum_{i=1}^n (p_i - \hat{p})^2} \sqrt{\sum_{i=1}^n (s_i - \hat{s})^2}} \quad (2.48)$$

όπου τα  $\hat{p}$  και  $\hat{s}$  δηλώνουν τη μέση τιμή των προβλεπόμενων βαθμολογιών του μοντέλου και των υποκειμενικών βαθμολογιών, αντίστοιχα. Για το RMSE μια μικρότερη τιμή υποδηλώνει καλύτερη απόδοση.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (s_i - p_i)^2} \quad (2.49)$$

# ΚΕΦΑΛΑΙΟ 3

## ΜΕΤΡΙΚΕΣ ΑΞΙΟΛΟΓΗΣΗΣ

### ΠΛΗΡΟΥΣ ΑΝΑΦΟΡΑΣ

#### 3.1 Μετρικές αξιολόγησης πλήρους αναφοράς

Όπως αναφέρθηκε και στο πρώτο κεφάλαιο η ενασχόλησή μας στη συγκεκριμένη εργασία αφορά τη δημιουργία νέων μοντέλων ή μετρικών αξιολόγησης της ποιότητας των εικόνων. Κάτι τέτοιο μπορεί να το δει κανείς ως μία υπολογιστική προσπάθεια μίμησης των μέσων προβλέψεων που θα έκαναν ανθρώπινοι παρατηρητές. Στη παρούσα εργασία αναφερόμαστε σε μετρικές πλήρους αναφοράς (full reference) στην αρχική εικόνα.

Ουσιαστικά αξιολογούμε την αντιληπτή ποιότητα μιας παραμορφωμένης εικόνας σε σχέση με την εικόνα αναφοράς της. Αυτή η πρόβλεψη ποιότητας είναι εύκολη υπόθεση για το ανθρώπινο οπτικό σύστημα (HVS) και το αποτέλεσμα της αξιολόγησης είναι αξιόπιστο. Η αυτόματη όμως αξιολόγηση της ποιότητας, υπό το βλέμμα μιας αντικειμενικής αξιολόγησης, δεν είναι εύκολη υπόθεση, διότι οι εικόνες μπορεί να έχουν υποστεί διάφορους τύπους και βαθμούς παραμορφώσεων. Οπότε, τι μπορεί να χρησιμοποιήσει κανείς ως χαρακτηριστικά ή πληροφορία για να εξαχθεί μία μετρική ποιότητας; Συνήθως οι μετρικές αυτού του είδους μετρούν δομικές πληροφορίες, πληροφορίες φωτεινότητας και πληροφορίες αντίθεσης στους τομείς του χώρου και της συχνότητας. Τα πιο επιτυχημένα μοντέλα IQA στη βιβλιογραφία ακολουθούν μια στρατηγική από πάνω προς τα κάτω όπως φαίνεται στο σχήμα που ακολουθεί[25].



**Εικόνα 3.1:** Σκορ ποιότητας μεταξύ της αρχικής και της παραμορφωμένης εικόνας.

Όπως μπορεί να παρατηρήσει κανείς, έχοντας στη διάθεσή μας την εικόνα αναφοράς αλλά και την παραμορφωμένη εικόνα ουσιαστικά υπολογίζεται ένας χάρτης ομοιότητας μεταξύ των δύο. Στη συνέχεια χρησιμοποιείται μία στρατηγική συγκέντρωσης των τιμών (pooling) που μετατρέπει τις τιμές αυτού του χάρτη ομοιότητας σε έναν ενιαίο δείκτη ποιότητας.

Όταν χρησιμοποιούμε χαρακτηριστικά και σκοπός μας είναι να προκύψουν συγκεκριμένοι χάρτες, τότε πολλές μετρικές χρησιμοποιούν μία απλοποιημένη μετρική ομοιότητας για βαθμωτές (scalar) τιμές της μορφής:

$$S(a, b, C) = \frac{2ab + C}{a^2 + b^2 + C} \quad (3.1)$$

όπου  $C > 0$

Ας θεωρήσουμε ως παράδειγμα τη γνωστή μετρική SSIM, οπότε αντιλαμβάνεται κανείς ότι η φωτεινότητα, η αντίθεση και οι δομικές πληροφορίες συνιστούν έναν χάρτη ομοιότητας για το συγκεκριμένο δείκτη [2] όπως αναφέρθηκε στο πρώτο κεφάλαιο. Στη συνέχεια, η μετρική SSIM χρησιμοποιεί τη μέση τιμή ως συγκέντρωση για τον υπολογισμό της τελικής βαθμολογίας ομοιότητας.

Ακολουθώντας αντίστοιχη φιλοσοφία, οι διάφορες μετρικές μπορεί να χρησιμοποιούν χαρακτηριστικά κλίσης όπως το μέγεθός της, χαρακτηριστικά από τον τομέα των κυματιδίων (wavelets) ή βασισμένα σε κάποιο άλλο πεδίο μετασχηματισμού. Σε αυτό το σημείο ενισχύεται η διαδικασία της ομοιότητας με

το να χρησιμοποιούνται διάφορα χρωματικά κανάλια – ανάλογα με το χρωματικό μοντέλο – ή κάποιου άλλου είδους χαρακτηριστικό που μπορεί να μας δώσει ένα χάρτη. Όπως και να έχει, η χρήση των χαρακτηριστικών σχετίζεται με τη δυνατότητά τους να συλλαμβάνουν τις διάφορες παραμορφώσεις.

Όπως αναφέραμε προηγουμένως, ο υπολογισμός του χάρτη ομοιότητας ακολουθείται από μια στρατηγική συγκέντρωσης. Οι σύγχρονες στρατηγικές συγκέντρωσης για την αξιολόγηση της αντιληπτής ποιότητας εικόνας (IQA) βασίζονται συνήθως στη μέση τιμή και στη σταθμισμένη μέση τιμή. Πρόκειται για ισχυρές στρατηγικές συγκέντρωσης που συνήθως παρέχουν μέτρια έως υψηλή απόδοση για διαφορετικές IQA.

Με τη μέση συγκέντρωση όλα τα συστατικά στο χάρτη ομοιότητας λαμβάνονται υπόψιν υπολογίζοντας το μέσο όρο οπότε και προσδιορίζεται η συνολική πρόβλεψη ποιότητας. Όμως οι εικόνες έχουν διαφορετικές τοπικές περιοχές που θα πρέπει με κάποιο τρόπο να συνυπολογίζονται αφού συνεισφέρουν στη συνολική ποιότητα. Επομένως, υιοθετούνται σταθμισμένες στρατηγικές βελτιώνοντας την ακρίβεια της ποιότητας έχοντας όμως το κόστος υπολογισμού των βαρών.

Οι μετρικές όμως δεν εξαντλούνται στη χρήση αυτών των συγκεντρώσεων, αφού στη βιβλιογραφία μπορεί κανείς να δει τη συγκέντρωση Minkowski [29], τη συγκέντρωση εκατοστημορίων [31] καθώς και άλλες.

Μία άλλη μορφή συγκέντρωσης τιμών είναι η τιμή της τυπικής απόκλισης όπως αυτή χρησιμοποιήθηκε με επιτυχία από την μέθοδο GMSD [19]. Τέτοιου είδους όμως μετρικές βασισμένες στην τυπική απόκλιση επιδεικνύουν καλή επίδοση μόνο όταν έχουμε ένα χαρακτηριστικό.

Σε γενικές γραμμές, τα χαρακτηριστικά τα ευαίσθητα σε παραμορφώσεις, που μας παρέχουν τον χάρτη ομοιότητας και η στρατηγική συγκέντρωσης είναι πολύ σημαντικοί παράγοντες στο σχεδιασμό μοντέλων IQA υψηλής απόδοσης. Επιπλέον στοιχεία που ελέγχει κανείς σε τέτοιες μεθοδολογίες είναι να είναι αποτελεσματικές και αξιόπιστες.

Ένα επιπλέον στοιχείο που δίνουμε σημασία είναι η προσπάθεια γενικά των διαφόρων μετρικών να συμβιβάσουν έννοιες όπως η ακρίβεια πρόβλεψης και η πολυπλοκότητα ενός μοντέλου.

### **3.2 Προσεγγίσεις βασισμένες σε φυσικές εικόνες (natural images) - NSS - Στατιστική φυσικών σκηνών**

Η προσέγγιση που βασίζεται στις στατιστικές ιδιότητες των φυσικών εικόνων (NSS- Natural Scene Statistics) έχει εφαρμοστεί σε διάφορες εργασίες αποτίμησης της ποιότητας των εικόνων π.χ. σε διάφορες μετρικές χωρίς αναφορά (no-reference). Τέτοιου είδους μεθοδολογίες χρησιμοποιούν την αρχή ότι οι φυσικές εικόνες κατέχουν κάποιες κανονικές (regular) στατιστικές ιδιότητες οι οποίες τροποποιούνται με μετρήσιμο τρόπο λόγω της ύπαρξης κάποιων παραμορφώσεων. Έτσι έχουν προκύψει γνωστές μεθοδολογίες που μοντελοποιούν τις εικόνες είτε στο πεδίο των κυματιδίων (DIVINE) [29], είτε στο πεδίο του Διακριτού Μετασχηματισμού Συνημιτόνων (DCT) [12].

Είναι σημαντικό να τονιστεί ότι οι φυσικές εικόνες είναι ιδιαίτερα δομημένες και το ανθρώπινο οπτικό σύστημα ιδιαίτερα προσαρμοζόμενο στην εξαγωγή της δομημένης πληροφορίας.

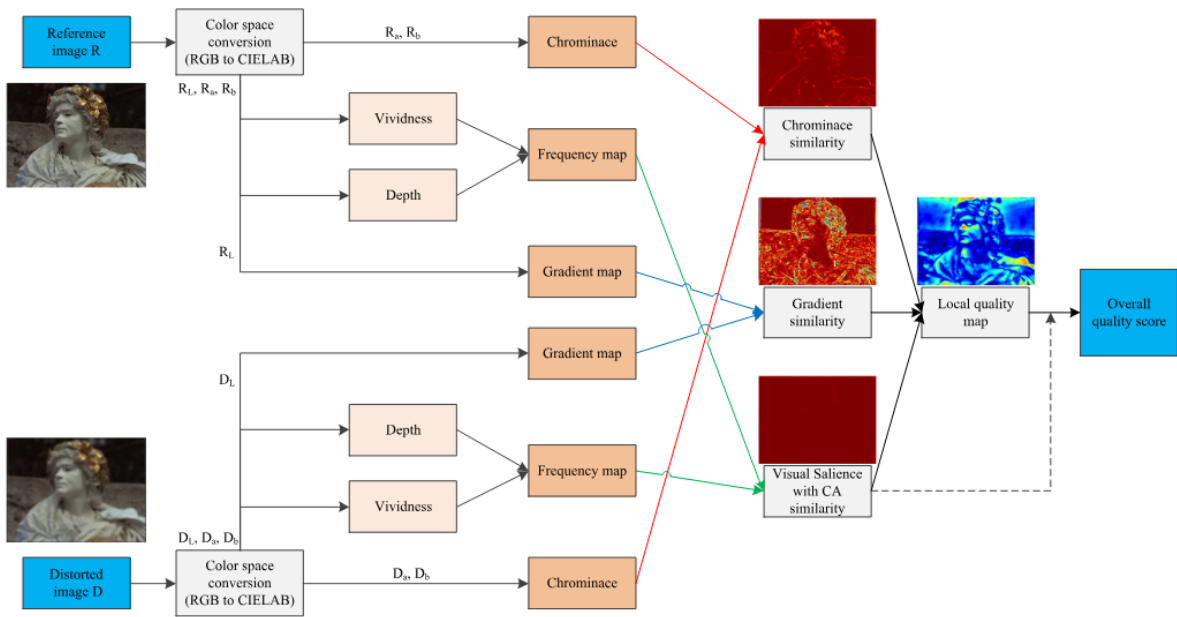
### **3.3 Μετρική βασισμένη στην αντιληπτή οπτική ποιότητα έγχρωμων εικόνων**

Οι Chenyang Shi και Yandan Lin στην εργασία τους [22] πρότειναν τη μεθοδολογία VCGS, η οποία αναλύθηκε λεπτομερώς σε προηγούμενο κεφάλαιο.

Το προτεινόμενο μοντέλο της συγκεκριμένης εργασίας είναι γενικού σκοπού και στην πραγματικότητα αποδίδει σταθερά καλά σε σχέση με τις κοινές χρωματικές και αχρωματικές παραμορφώσεις. Τρεις χάρτες ομοιότητας χαρακτηριστικών

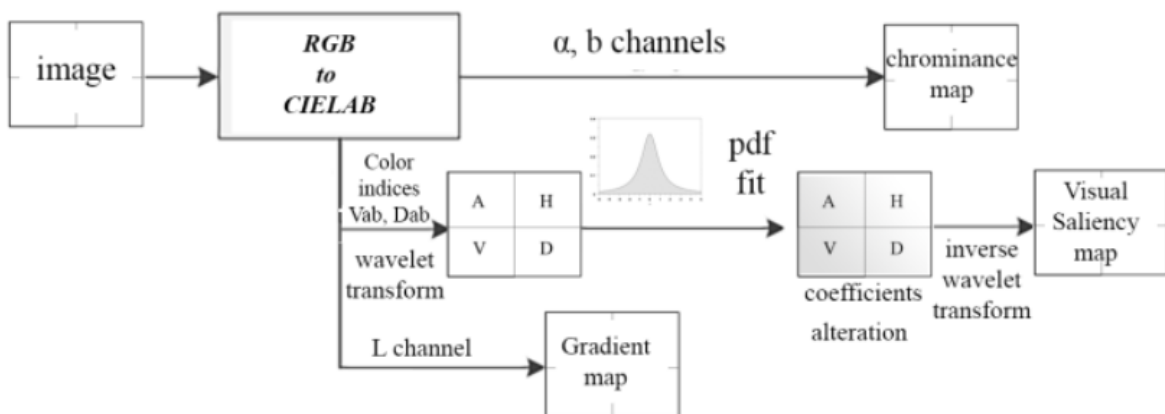
περιλαμβάνονται στο προτεινόμενο μοντέλο. Το χαρακτηριστικό οπτικής υπεροχής εξάγεται από την εμφάνιση χρώματος στον χρωματικό χώρο CIELAB, η οποία θεωρείται πιο συμβατή με την ανθρώπινη διαίσθηση [30]. Οι χάρτες ομοιότητας χρωματισμού μπορούν να εκφράσουν τη διαφορά χρώματος μεταξύ δύο εικόνων σε επίπεδο pixel από τα κανάλια χρωματισμού. Για το χαρακτηριστικό της δομής, η κλίση χρησιμοποιείται για την εξαγωγή ενός άλλου χάρτη ομοιότητας λόγω της ανωτερότητάς του οπότε αυτοί οι τρεις χάρτες ομοιότητας συνδέονται και συγκεντρώνονται.

Εμπνευσμένο από τα [23] και [25], σε αυτή την έρευνα, ένα ζωνοπερατό (band-pass) φίλτρο χρησιμοποιείται για την ανίχνευση της οπτικής υπεροχής. Λόγω της καλύτερης απόδοσης του φίλτρου Log-Gabor, επιλέγεται επίσης ως τελεστής για την απόκτηση οπτικής σημασίας ενώ οι δείκτες εμφάνισης χρώματος είχαν χρησιμοποιηθεί άμεσα στο μοντέλο IQA. Για να χαρακτηριστεί καλύτερα η χρωματική εμφάνιση των εικόνων που γίνονται αντιληπτές από το ανθρώπινο οπτικό σύστημα, οι αρχικές εικόνες RGB μετατρέπονται σε χρωματικό χώρο που είναι πιο συμβατός με τη διαίσθηση του ανθρώπου. Μεταξύ διαφόρων αντιληπτικών χρωματικών μοντέλων, το CIELAB χρησιμοποιείται σε αυτή την έρευνα, οπότε δύο νέες μεταβλητές CIELAB, *η ζωντάνια (vividness) και το βάθος (depth)* [30], εισήχθησαν για να επεκτείνουν τη χρησιμότητα του CIELAB ως μεταβλητή εμφάνισης χρώματος.



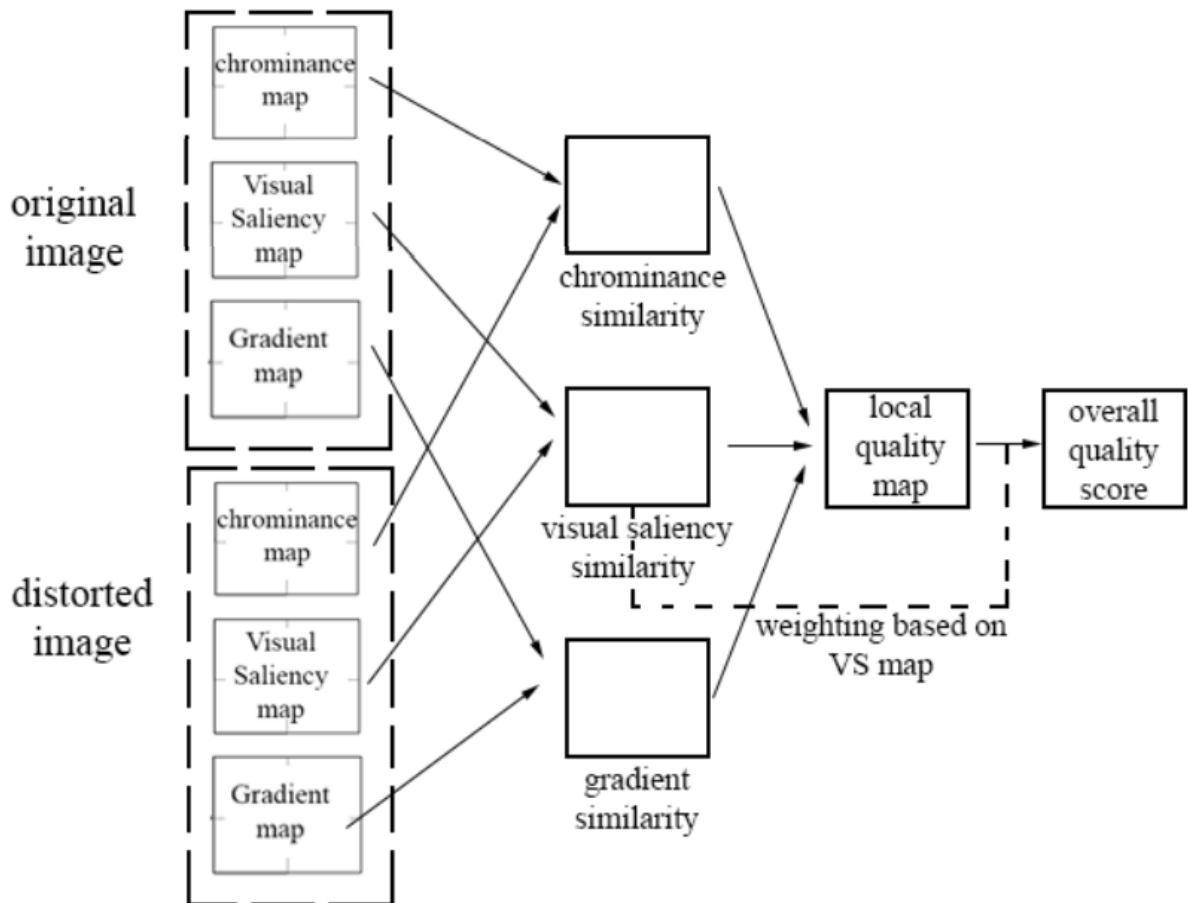
Εικόνα 3.2: Η μεθοδολογία VCGS.

Μια μετεξέλιξη της συγκεκριμένης μεθοδολογίας είναι η VSSC. Η VSSC συνιστά μία περαιτέρω προσπάθεια για την ανάπτυξη και βελτίωση των μεθόδων αξιολόγησης της ποιότητας των εικόνων, συμβάλλοντας στη συνεχή πρόοδο του ερευνητικού πεδίου. Σε αντίθεση με την αρχική μορφή της VCGS, όπου χρησιμοποιούνται ζωνοπερατά (band pass) φίλτρα με τη βοήθεια του Log-Gabor για την ανίχνευση του saliency, η VSSC υιοθετεί μια διαφορετική προσέγγιση.



Εικόνα 3.3: Ροή της VSSC για τη δημιουργία χαρτών χαρακτηριστικών της οπτικής περιοχής, της κλίσης και της χρωματικότητας.





**Εικόνα 3.4:** Το διάγραμμα ροής για τη δημιουργία των χαρτών ομοιότητας της οπτικής ευκρίνειας, της κλίσης και της χρωματικότητας, του τελικού τοπικού χάρτη ποιότητας και της σταθμισμένης συγκέντρωσης με βάση την οπτική ευκρίνεια για την εξαγωγή της συνολικής βαθμολογίας ποιότητας.

Εφαρμόζεται wavelet μετασχηματισμός και εξετάζονται οι παραγόμενοι συντελεστές. Αξιοποιώντας τις λεπτομέρειες κατεύθυνσης που παρέχονται από τον μετασχηματισμό, αντικαθιστά αυτήν την πληροφορία με παραμέτρους που χρησιμοποιούνται στην εκτίμηση των παραμέτρων μίας προσαρμοσμένης κατανομής Student's-t στους συντελεστές που εξετάζουμε. Στη συνέχεια, μέσω αντίστροφου μετασχηματισμού, επιστρέφει στο χωρικό πεδίο και υπολογίζεται το visual saliency. Σε προηγούμενο κεφάλαιο αναλύθηκε η χρησιμότητα του μετασχηματισμού wavelet, καθώς και οι διαφορές του με τον μετασχηματισμό Fourier.

### 3.4 Στατιστική περιγραφή συντελεστών κυματιδίων

Οι συντελεστές κυματιδίων κάθε επιπέδου μπορεί να περιγραφούν από κάποια στατιστική κατανομή, ενώ εμείς έχουμε προσαρμόσει (fit) με την Student\_t κατανομή [30]. Οι συντελεστές αυτοί μπορεί να είναι κάθε φορά διαφορετικοί, ανάλογα με το τι θέλουμε να εξετάσουμε ή ποιο επίπεδο του μετασχηματισμού έχουμε επιλέξει. Μπορούμε να μηδενίζουμε κάποιους, κρατώντας τις τιμές κάποιων, ή να κάνουμε αντικατάσταση με τις αντίστοιχες πληροφορίες που προκύπτουν από κατάλληλη παραμετροποίηση και εκτίμηση παραμέτρων σε κάθε θέση.

Ο ορισμός της μονοδιάστατης στατιστικής κατανομής της Student's-t που χρησιμοποιείται στη παρούσα διπλωματική εργασία είναι ο εξής:

$$St_1(x_1|0, \lambda, \nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left(\frac{\lambda}{\pi\nu}\right)^{\frac{1}{2}} \left(1 + \frac{\lambda}{\nu} x_i\right)^{\frac{\nu+1}{2}} \quad (3.2)$$

όπου  $i=1, \dots, N$  είναι το πλήθος των δεδομένων – παρατηρήσεων ενδιαφέροντος,  $\nu$  είναι το πλήθος των βαθμών ελευθερίας, ενώ οι τιμές που λαμβάνουν τα δεδομένα μας μπορεί να ανήκουν σε ολόκληρη τη γραμμή των πραγματικών αριθμών, ήτοι  $-\infty < x < +\infty$ . Επιπλέον, ως  $\Gamma$  ορίζεται η γνωστή μας Γάμμα συνάρτηση, όπου:

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx \quad (3.3)$$

Η Student's t κατανομή είναι συμμετρική κοντά στο 0, που συμβαδίζει με το γεγονός ότι η μέση τιμή της είναι μηδέν και η τιμή ασυμμετρίας είναι, επίσης, μηδέν.

Η Student-t έχει αποδειχτεί κατάλληλη για την μοντελοποίηση δεδομένων, που έχουν παχιές (βαριές) ουρές (heavy tails). Το πεδίο μετασχηματισμού των κυματιδίων (wavelet domain) είναι βασικό παράδειγμα, καθώς υπάρχει μεγάλο πλήθος μικρών συντελεστών (ως προς την τιμή) και μικρό πλήθος πολύ μεγάλων συντελεστών.

Συνεπώς, η Student's t [30] κατανομή, ανάλογα με την επιλογή των παραμέτρων, μπορεί να έχει βαριές ουρές. Όσο οι βαθμοί ελευθερίας αυξάνονται,

η συνάρτηση πυκνότητας πιθανότητας προσεγγίζει την κανονική κατανομή, ενώ, όσο πλησιάζουν το μηδέν, δεν έχει κάποια τιμή. Η κατανομή είναι συμμετρική, όσον αφορά την μέση τιμή της, και η διακύμανση είναι:

$$\frac{v}{(v-2)} \text{ με } v > 2 \quad (3.4)$$

διαφορετικά δεν ορίζεται. Γενικά, μόνο οι ροπές που είναι μικρότερες από το  $v$  ορίζονται.

Η μορφή του προτεινόμενου παραμετρικού μοντέλου καθορίζει, εν τέλει, τη λεπτομερή συμπεριφορά του μοντέλου αξιολόγησης και επιτρέπει έναν συγκεκριμένο βαθμό προσαρμοστικότητας του μοντέλου σε διαφορετικές πηγές δεδομένων (σε διαφορετικές μορφές των εικόνων που αξιολογούμε). Η Student's  $t$  είναι γνωστό ότι επιτρέπει μεγάλη ευελιξία στην μορφή της κατανομής της και, επιπλέον, μπορούμε να υπολογίσουμε την αβεβαιότητα, ως προς την φύση της ουράς της εμπειρικής κατανομής που μελετάμε. Κάτι τέτοιο καθιστά τη προτεινόμενη κατανομή ένα ευέλικτο μοντέλο, που μπορεί να καταγράψει το σχήμα της υποκείμενης συνάρτησης πυκνότητας πιθανότητας των δεδομένων που μας ενδιαφέρουν.

Για την περίπτωση μας, είναι σημαντικό να μπορούμε να έχουμε καλές εκτιμήσεις των παραμέτρων που μας ενδιαφέρουν. Με στόχο την εύρεση των απαραίτητων παραμέτρων του προτεινόμενου μοντέλου, που πρόκειται να χρησιμοποιηθούν ως χαρακτηριστικά για τον αλγόριθμο αξιολόγησης, κάνουμε χρήση του επαναληπτικού αλγορίθμου EM (Expectation Maximization). Γενικά, πολλοί επαναληπτικοί αλγόριθμοι βασισμένοι στον EM έχουν αναπτυχθεί για να υπολογίσουν τις εκτιμήσεις της μέγιστης πιθανοφάνειας (ML-Maximum Likelihood) για τις παραμέτρους της  $t$ -κατανομής. Παρόλο που υπεισέρχεται μία μορφή υπολογιστικής πολυπλοκότητας, έχει αποδειχτεί ότι η ML εκτίμηση έχει πολύ μεγάλη ακρίβεια και, γι' αυτό το λόγο, έχει επιλεγεί και στη συγκεκριμένη μεθοδολογία.

Στην εργασία μας, μεταχειριζόμαστε τους συντελεστές ενδιαφέροντος ως τυχαίες μεταβλητές και υποθέτουμε ότι τα  $x_i$  για  $i = 1, \dots, N$  είναι ανεξάρτητα και ομοιόμορφα κατανεμημένα δείγματα από την ίδια Student- $t$  κατανομή (αλλά με

διαφορετικό τελεστή  $k$ , ανάλογα με το επίπεδο wavelet ανάλυσης που χρησιμοποιείται) με μέσο μηδέν και παραμέτρους  $\lambda_k$  και  $\nu_k$ :

$$x_i^k \sim St(x_i^k; 0, \lambda_k, \nu_k), \forall i, \forall k, \quad (2.7)$$

$$St(x; 0, \lambda_k, \nu_k) = \frac{\Gamma(\frac{\nu_k+1}{2})}{\Gamma(\frac{\nu_k}{2})} \left(\frac{\lambda_k}{\pi\nu_k}\right)^{\frac{1}{2}} \left(1 + \frac{\lambda_k}{\nu_k} x^2\right)^{-\frac{\nu_k+1}{2}} \quad (3.5)$$

Η υπόθεση της Student-t κατανομής συνεπάγεται μία διαδικασία δημιουργίας δεδομένων δύο επιπέδων [30]. Για να το αναλύσουμε περισσότερο, ας σκεφτεί κανείς ένα είδος ιεραρχικού μοντέλου δημιουργίας των συντελεστών, όπου αρχικά οι παράμετροι  $a_i^k$  δημιουργούνται, με ανεξάρτητο τρόπο η κάθε μία, από μία Gamma κατανομή. Σε ένα δεύτερο επίπεδο, κάθε συντελεστής  $x_i^k$  δημιουργείται από μία Κανονική κατανομή με μέσο μηδέν και ακρίβεια (precision)  $\lambda_k a_i^k$ .

Κατ' αρχάς, πρέπει να σημειωθεί ότι η συνάρτηση πυκνότητας πιθανότητας της εξίσωσης (2.7) μπορεί να γραφεί σαν το ολοκλήρωμα:

$$p(x_i^k) = St(x_i^k; 0, \lambda_k, \nu_k) = \int_0^\infty p(x_i^k | a_i^k) p(a_i^k) da_i^k \quad (3.6)$$

όπου ισχύει:

$$p(x_i^k | a_i^k) = N\left(0, (\lambda_k a_i^k)^{-1}\right), p(a_i^k) = Gamma\left(\frac{\nu_k}{2}, \frac{\nu_k}{2}\right) \quad (3.7)$$

Για να ερμηνεύσουμε τους πιο πάνω τύπους πιο διαισθητικά, θα πρέπει να δούμε τι γίνεται με τη χωρική μεταβλητότητα/προσαρμοστικότητα του συγκεκριμένου μοντέλου. Σκεπτόμενοι τη χρησιμοποιούμενη μονοδιάστατη έκδοση των πιο πάνω τύπων, τότε σε κάθε θέση -  $i$  - παρατηρούμε ότι οι συντελεστές  $x_i^k$  μοντελοποιούνται ως μία Γκαουσιανή κατανομή με μέσο μηδέν και διακύμανση  $(a_i^k)^{-1}$ . Οι μεταβλητές  $x_i^k$  ονομάζονται «κρυφές», λόγω του γεγονότος ότι δεν εμφανίζονται στην εξίσωση (εξάγονται λόγω της ολοκλήρωσης), αλλά παίζουν κρίσιμο ρόλο στο μοντέλο που δημιουργεί τους συντελεστές μας.

### 3.5 Πρώτη ομάδα πειραμάτων - χρήση νέων χαρακτηριστικών υφής

Στην πρώτη ομάδα πειραμάτων, επεκτείνουμε την VSSC προσθέτοντας δύο νέα χαρακτηριστικά υφής:

- τη *διαφορά μπλοκ των αντίστροφων πιθανοτήτων (BDIP)* και
- τη *διακύμανση μπλοκ των τοπικών συντελεστών συσχέτισης (BVLC)*

[31]

Το χαρακτηριστικό BDIP χρησιμοποιεί τοπικές πιθανότητες σε μπλοκ της εικόνας για τη μέτρηση των τοπικών μεταβολών στη φωτεινότητα μιας εικόνας.

Από την άλλη πλευρά, το χαρακτηριστικό BVLC χρησιμοποιεί τις μεταβολές των τοπικών συντελεστών συσχέτισης σε μπλοκ της εικόνας για να μετρήσει την τοπική ομαλότητα της υφής. Παρακάτω έχουμε ένα παράδειγμα όπου φαίνεται η αρχική εικόνα και τα αποτελεσμάτων των εικόνων μετά την εφαρμογή των BDIP και BVLC χαρακτηριστικών.

Το BDIP ορίζεται ως η διαφορά μεταξύ του αριθμού των εικονοστοιχείων σε ένα μπλοκ και του λόγου του αθροίσματος των εντάσεων των εικονοστοιχείων στο μπλοκ προς τη μέγιστη ένταση στο μπλοκ. Δηλαδή

$$BDIP = M^2 - \frac{\sum_{(i,j) \in B} I(i,j)}{\max_{(i,j) \in B} I(i,j)} \quad (3.8)$$

Όπου  $I(i,j)$  είναι η ένταση ενός εικονοστοιχείου και  $B$  ένα μπλοκ μεγέθους  $M \times M$ . Όσο μεγαλύτερη είναι η διακύμανση των εντάσεων σε ένα μπλοκ, τόσο υψηλότερη είναι η τιμή του BDIP.

Το δεύτερο χαρακτηριστικό υφής είναι το BVLC, όπου κάθε τοπικός συντελεστής συσχέτισης ορίζεται ως τοπική συνδιακύμανση κανονικοποιημένη με την τοπική διακύμανση.

$$p(k,l) = \frac{\frac{1}{M^2} \sum_{(i,j) \in B} I(i,j)I(i+k,j+l) - \mu_{0,0} \mu_{k,l}}{\sigma_{0,0} \sigma_{k,l}} \quad (3.9)$$

όπου  $B$  συμβολίζει ένα μπλοκ μεγέθους  $M \times M$  και  $\mu_{o,o}$  και  $\sigma_{o,o}$  συμβολίζουν τον τοπικό μέσο όρο και την τυπική απόκλιση του μπλοκ  $B$ , αντίστοιχα. Ο συμβολισμός  $(k,l)$  υποδηλώνει ένα ζεύγος οριζόντιας και κάθετης μετατόπισης που σχετίζονται με τους τέσσερις προσανατολισμούς  $(-90^\circ, 0^\circ, -45^\circ, 45^\circ)$ . Κατά συνέπεια, τα  $\mu_{k,l}$  και  $\sigma_{k,l}$  αντιπροσωπεύουν τη μέση τιμή και την τυπική απόκλιση του μπλοκ που μετατοπίζεται κατά  $(k,l)$ , αντίστοιχα.

Ως αποτέλεσμα, η τιμή του BVLC εκφράζεται ως εξής:

$$BVLC = \max_{(k,l) \in O^4} [p(k,l)] - \min_{(k,l) \in O^4} [p(k,l)] \quad (3.10)$$

$$O^4 = \{(0,1), (1,0), (1,1), (1,-1)\} \quad (3.11)$$

οι προσανατολισμοί αυτοί αντιστοιχούν σε συγκεκριμένες γωνίες στην εικόνα:

- Κάθετος  $(0,1)$ :  $90^\circ$
- Οριζόντιος  $(1,0)$ :  $0^\circ$
- Διαγώνιος  $(1,1)$ :  $45^\circ$
- Αντίθετος διαγώνιος  $(1,-1)$ :  $-45^\circ$

Επομένως όσο μεγαλύτερη είναι η τραχύτητα, τόσο υψηλότερη θα είναι η τιμή του BVLC. Να σημειωθεί ότι η τραχύτητα αναφέρεται στη μεταβλητότητα ή την ανισότητα στην υφή του μπλοκ. Ένα μπλοκ με υψηλή τραχύτητα έχει μεγάλη διακύμανση στις εντάσεις των εικονοστοιχείων του, που μπορεί να περιλαμβάνει πολλές λεπτομέρειες ή ανωμαλίες.)



**Εικόνα 3.5 :** Αρχική εικόνα, εικόνα BDIP και BVLC εικόνα.

Αυτά τα χαρακτηριστικά προστίθενται στον αρχικό πίνακα ομοιότητας με σκοπό την βελτίωση της αξιολόγησης της ποιότητας της εικόνας. Προσαρμόζουμε τη μεθοδολογία, εφαρμόζοντας το χαρακτηριστικό BDIP στην αρχή, στη συνέχεια το χαρακτηριστικό BVLC, στη συνέχεια τον συνδυασμό των BDIP και BVLC και τέλος το χαρακτηριστικό BDIP χωρίς τη χρήση της πληροφορίας κλίσης, η οποία είναι βασική για τη μέθοδο VSSC. Οι αξιολογήσεις περιλαμβάνουν τις μετρικές SROCC, KROCC, PLCC, και RMSE. Στο παρακάτω πίνακα παρουσιάζονται τα αποτελέσματα των πειραμάτων.

**Πίνακας 3.1:** Επίδοση μεθοδολογίας με προσθήκη χαρακτηριστικών υφής στη μεθοδολογία της VSSC

	VCGS	VSSC	BDIP	BVLC	BDIP+BVLC	BDIP χωρίς πληροφορία κλίσης (gradient)
<b>SROCC</b>	0.8926	0.8303	0.8304	0.6975	0.6975	0.6740
<b>KROCC</b>	0.7166	0.6422	0.6423	0.5140	0.5140	0.4847
<b>PLCC</b>	0.9000	0.8623	0.8624	0.7476	0.7475	0.7105
<b>RMSE</b>	0.5404	0.6277	0.6276	0.8234	0.8234	0.8723

Συμπερασματικά, σε σύγκριση με την μέθοδο VSSC, προκύπτουν οι παρακάτω παρατηρήσεις :

Η προσθήκη του χαρακτηριστικού BDIP στην υπάρχουσα μεθοδολογία (VSSC) φαίνεται να οδηγεί σε βελτίωση σε όλες τις μετρικές απόδοσης (SROCC, KROCC, PLCC) σε σχέση με τη μέθοδο VSSC από μόνη της. Το RMSE παραμένει σε λογικά χαμηλά επίπεδα, υποδεικνύοντας καλή προσαρμογή του μοντέλου.

Αντίθετα, η προσθήκη του χαρακτηριστικού BVLC δεν επιφέρει βελτίωση στα αποτελέσματα σε σχέση με τη μέθοδο VSSC. Οι τιμές των μετρικών μειώνονται, ενώ το RMSE αυξάνεται, υποδεικνύοντας μεγαλύτερη αβεβαιότητα στις προβλέψεις. Όμοια και για την συνδυαστική χρήση των χαρακτηριστικών BDIP και BVLC. Η προσθήκη του χαρακτηριστικού BDIP χωρίς τη χρήση της `gradientSimMatrix` οδηγεί σε χειρότερα αποτελέσματα.

Συνολικά, η προσθήκη του χαρακτηριστικού BDIP βελτιώνει την πρόβλεψη της ποιότητας της εικόνας, ενώ η προσθήκη του χαρακτηριστικού BVLC δεν φαίνεται να είναι επωφελής.

### 3.6 Δεύτερη ομάδα πειραμάτων

Στη δεύτερη ομάδα πειραμάτων, εφαρμόζεται μια τροποποίηση στην VSSC με σκοπό την εξέταση της πιθανότητας βελτίωσης των αποτελεσμάτων. Αρχικά, η VSSC υπολογίζει τις τιμές  $V_{ab}$  και  $D_{ab}$  από τα κανάλια L, a και b της εικόνας Lab.

Όταν μιλάμε για τον ορισμό της ζωντάνιας ( $V^{*ab}$ ) και του βάθους ( $D^{*ab}$ ) στον χρωματικό χώρο CIELAB, αναφερόμαστε σε αποστάσεις από ένα σημείο αναφοράς στον χρωματικό χώρο. Συγκεκριμένα:

**Vividness ( $V^{*ab}$ ) - "Ζωντάνια":** Αντιπροσωπεύει την απόσταση από ένα χρώμα που ορίζεται από μια συγκεκριμένη φωτεινότητα ( $L^*$ ) και χρωματικότητα ( $C_{ab}$ ) σε ένα σημείο αναφοράς, όπου όταν η χρωματικότητα ( $C_{ab}$ ) είναι 0 και η φωτεινότητα ( $L^*$ ) βρίσκεται στο πιο σκοτεινό σημείο της ( $L^* = 0$ ). Ουσιαστικά, μετράει πόσο πολύ αποκλίνει ένα χρώμα από μια εντελώς αποχρωματισμένη (γκρίζα) εκδοχή του εαυτού του στην πιο σκούρα δυνατή τιμή του, δηλαδή αποτυπώνει πόσο ζωντανό ή έντονο εμφανίζεται ένα χρώμα.

**Depth ( $D_{ab}$ )- Βάθος:** Αντιπροσωπεύει την απόσταση από ένα χρώμα που ορίζεται από μια συγκεκριμένη φωτεινότητα ( $L^*$ ) και χρωματικότητα ( $C_{ab}$ ) σε ένα σημείο αναφοράς, όπου όταν η χρωματικότητα ( $C_{ab}$ ) είναι 0 και η φωτεινότητα ( $L^*$ ) βρίσκεται στο πιο φωτεινό σημείο ( $L^* = 100$ ). Μετράει πόσο ένα χρώμα αποκλίνει από μια εντελώς αποχρωματισμένη έκδοση του εαυτού του στην ελαφρύτερη δυνατή τιμή του, δηλαδή καταγράφει πόσο βαθύ ή σκοτεινό εμφανίζεται ένα χρώμα. [32]

Το κανάλι a αντιπροσωπεύει την πληροφορία για τις αποχρώσεις του χρώματος κόκκινου-πράσινου, ενώ το κανάλι b αντιστοιχεί στην πληροφορία για τις αποχρώσεις του μπλε-κίτρινου. Επιπρόσθετα, το L κανάλι αντιπροσωπεύει τη



φωτεινότητα και κινείται από το μαύρο στο άσπρο, με το 0 να αντιπροσωπεύει το μαύρο και το 100 να αντιπροσωπεύει το άσπρο.

Στη συνέχεια, εφαρμόζεται μετασχηματισμός wavelet και υπολογίζεται ο SF χάρτης. Πιο συγκεκριμένα, εκτελείται ένας 2D διακριτός μετασχηματισμός κυματιδίων (DWT) σε μια εικόνα εισόδου χρησιμοποιώντας το κυματίδιο Daubechies 4 (db4). Εκτιμώνται οι παράμετροί για τους οριζόντιους, κάθετους και διαγώνιους συντελεστές λεπτομέρειας της μετασχηματισμένης εικόνας. Μετά την αναδιαμόρφωση και την επεξεργασία των εκτιμώμενων παραμέτρων, μηδενίζονται οι κάθετοι και διαγώνιοι συντελεστές λεπτομέρειας. Τέλος, ανακατασκευάζεται η εικόνα χρησιμοποιώντας μόνο τους οριζόντιους συντελεστές λεπτομέρειας και τους συντελεστές προσέγγισης. Ο SF χάρτης κατ' επέκταση προκύπτει από την εφαρμογή μετασχηματισμού wavelet στις τιμές  $V_{ab}$  και  $D_{ab}$  (οι οποίες αντιπροσωπεύουν την αντίστοιχη ζωτικότητα και βάθος του χρώματος στην εικόνα) και τον υπολογισμό της Ευκλείδειας απόστασης μεταξύ τους. Τέλος, ανακατασκευάζεται ο χάρτης VS στο μέγεθος της αρχικής εικόνας και κανονικοποιούνται οι τιμές του στο διάστημα [0,1].

Η τροποποίηση περιλαμβάνει δύο διαφορετικές παραλλαγές της VSSC. Στην πρώτη περίπτωση, χρησιμοποιείται μόνο το κανάλι A για τον υπολογισμό του Visual Saliency, ενώ στη δεύτερη περίπτωση χρησιμοποιείται το κανάλι B, ενώ στην αρχική έκδοση της VSSC όπως αναφέρθηκε προηγουμένως χρησιμοποιούνται συνδυαστικά και τα τρία κανάλια a, b και L για αυτόν τον υπολογισμό. Επιπλέον, για κάθε περίπτωση, χρησιμοποιείται ο λογάριθμος της κρυμμένης πληροφορίας των οριζόντιων συντελεστών του μετασχηματισμού wavelet, ενώ οι συντελεστές προσέγγισης cA χρησιμοποιούνται είτε αυτούσιοι είτε μηδενίζονται, ανάλογα με την παραλλαγή.

Αυτό οδηγεί στην πραγματοποίηση συνολικά τεσσάρων διαφορετικών πειραμάτων για την αξιολόγηση της απόδοσης του συστήματος σε διάφορες βάσεις δεδομένων. Αυτά τα πειράματα συγκρίνονται με την αρχική VSSC προκειμένου να διαπιστωθεί εάν οι τροποποιήσεις οδηγούν σε βελτιωμένα αποτελέσματα.

Στον παρακάτω πίνακα παρουσιάζονται οι συγκεντρωτικοί βαθμοί για τις βάσεις δεδομένων TID2013 και KADID-10K, οι οποίοι χρησιμοποιήθηκαν για τη σύγκριση της απόδοσης της τροποποιημένης VSSC με την αρχική έκδοση.

Πίνακας 3.2. Επίδοση μεθοδολογίας με τροποποίηση συντελεστών κυματιδίων και επιλογή συγκεκριμένου καναλιού του χρωματικού μοντέλου CIELAB

Αλγόριθμος	SROCC	KROCC	PLCC	RMSE	Μετρική
VCGS	0.8926	0.7166	0.9000	0.5404	TID2013
VSSC	0.8303	0.6422	0.8623	0.6277	TID2013
Διατήρηση cA (κανάλι a)	0.8462	0.6537	0.8526	0.6478	TID2013
Μηδενισμός cA (κανάλι a)	0.8553	0.6677	0.8679	0.6157	TID2013
Διατήρηση cA (κανάλι b)	0.8521	0.6618	0.8574	0.6380	TID2013
Μηδενισμός cA (κανάλι b)	<b>0.8570</b>	<b>0.6697</b>	<b>0.8688</b>	<b>0.6138</b>	TID2013
VCGS	0.8705	0.6828	0.8680	0.5375	KADID10K
VSSC	0.8443	0.6564	0.8425	0.5832	KADID10K
Διατήρηση cA (κανάλι a)	0.8515	0.6600	0.8498	0.5706	KADID10K
Μηδενισμός cA (κανάλι a)	0.8318	0.6385	0.8297	0.6044	KADID10K
Διατήρηση cA (κανάλι b)	<b>0.8654</b>	<b>0.6763</b>	<b>0.8631</b>	<b>0.5468</b>	KADID10K
Μηδενισμός cA (κανάλι b)	0.8323	0.6392	0.8301	0.6036	KADID10K

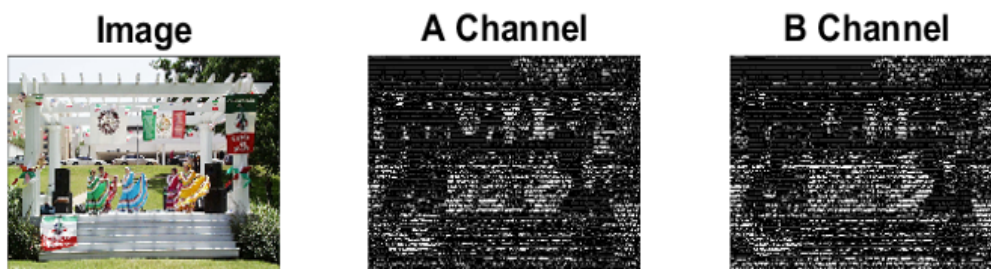
Συνοψίζοντας τα αποτελέσματα, παρατηρούμε ότι για τη βάση δεδομένων TID2013, η προσέγγιση που μηδένισε του συντελεστές προσέγγισης cA (στο κανάλι b) μας έδωσε την καλύτερη απόδοση. Στη βάση δεδομένων KADID10K, η προσέγγιση που διατήρησε το cA (στο κανάλι b) ήταν η πιο αποτελεσματική.

Αυτό υποδεικνύει ότι οι διάφορες βάσεις δεδομένων μπορεί να επηρεάζονται διαφορετικά από την παρουσία ή την απουσία των συντελεστών προσέγγισης (cA), και συνεπώς δεν μπορούμε να καταλήξουμε σε ένα γενικό συμπέρασμα χωρίς να ληφθούν υπόψη όλοι οι παράγοντες και οι μετρικές που χρησιμοποιούνται.

Παρακάτω παρουσιάζεται η επίδραση των καναλιών a και b σε εικόνες από διαφορετικές βάσεις δεδομένων, με την παρουσία ή την απουσία του cA.



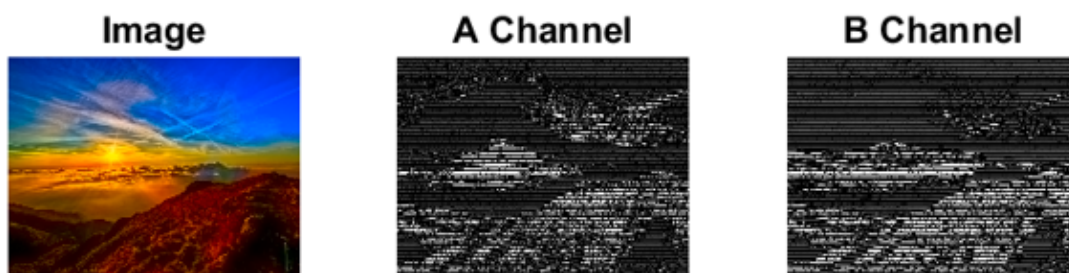
**Εικόνα 3.6:** Αρχική εικόνα από τη βάση δεδομένων TID2013, ανακατασκευασμένες εικόνες για το κανάλι a και το κανάλι b, έχοντας κρατήσει τους συντελεστές προσέγγισης (cA) χωρίς τροποποίηση.



**Εικόνα 3.7:** Αρχική εικόνα από τη βάση δεδομένων TID2013, ανακατασκευασμένες εικόνες για το κανάλι a και το κανάλι b, έχοντας μηδενίσει τους συντελεστές προσέγγισης cA.



**Εικόνα 3.8 :** Αρχική εικόνα από τη βάση δεδομένων KADID-10K, ανακατασκευασμένες εικόνες για το κανάλι a και το κανάλι b, έχοντας κρατήσει τους συντελεστές λεπτομέρειας ( $c_A$ ).



**Εικόνα 3.9:** Αρχική εικόνα από τη βάση δεδομένων KADID-10K, ανακατασκευασμένες εικόνες για το κανάλι a και το κανάλι b, έχοντας μηδενίσει το  $c_A$ .

### 3.7 Τρίτη ομάδα πειραμάτων

Στα πλαίσια της τρίτης ομάδας πειραμάτων, αναπτύχθηκε μια νέα μετρική αξιολόγησης ποιότητας εικόνων. Η μετρική αυτή χρησιμοποιεί τους χάρτες κλίσης για να υπολογίσει ένα μέτρο αξιολόγησης ποιότητας μεταξύ μιας αρχικής εικόνας και μιας παραμορφωμένης εικόνας.

Πιο συγκεκριμένα, αναπτύσσεται ένα μοντέλο όπου αρχικά, οι εικόνες εισόδου μετατρέπονται σε διπλή ακρίβεια. Στη συνέχεια, υπολογίζονται τέσσερις χάρτες κλίσης για τις εικόνες, όπου ο πρώτος χάρτης κλίσης περιέχει πληροφορίες για τις κάθετες μεταξύ τους πρώτες διαφορές στις 0 και 90 μοίρες της αρχικής εικόνας, ενώ ο δεύτερος περιέχει πληροφορίες για τις κάθετες μεταξύ τους πρώτες διαφορές στις 0 και 90 μοίρες της παραμορφωμένης εικόνας. Όμοια υπολογίζονται ακόμη δυο χάρτες για τις κάθετες μεταξύ τους διαφορές στις 45 και 135 μοίρες.

Έπειτα, χρησιμοποιώντας αυτούς τους χάρτες, για να επιτευχθεί η αξιολόγηση ποιότητας για τις διαφορές μεταξύ των δύο εικόνων, υπολογίζεται ο εξής τύπος :

$$quality\_map = \frac{2 \times gradientMap1 \times gradientMap2 + T}{gradientMap1^2 + gradientMap2^2 + T} \quad (3.12)$$

Για την παράμετρο T, χρησιμοποιήθηκαν οι ακόλουθες τιμές: 0.1, 1, 175, 1000 και 10000. Τέλος, υπολογίζεται η τυπική απόκλιση αυτού του μέτρου ποιότητας και επιστρέφεται ως αποτέλεσμα.

Πραγματοποιήθηκαν τρία διαφορετικά πειράματα για την αξιολόγηση της νέας μετρικής. Σε κάθε πείραμα, η μετρική εφαρμόστηκε σε δύο εικόνες: μια αρχική εικόνα και μια παραμορφωμένη εικόνα. Τα πειράματα διαφοροποιούνταν στον τρόπο υπολογισμού των χαρτών κλίσης και στις παραμέτρους που χρησιμοποιούνταν στον υπολογισμό του τελικού σκορ.

**Πείραμα 1:** Στο πρώτο πείραμα, οι χάρτες κλίσης υπολογίστηκαν με βάση τις πρώτες διαφορές στις 0 και 90 μοίρες. Αυτή η προσέγγιση χρησιμοποιήθηκε για τον υπολογισμό του τελικού σκορ.

**Πείραμα 2:** Στο δεύτερο πείραμα, οι χάρτες κλίσης παρήχθησαν με τις πρώτες διαφορές στις 0 και 90 μοίρες, αλλά με μειωμένη διάσταση στην εικόνα. Αυτή η προσέγγιση χρησιμοποιήθηκε για τον υπολογισμό του τελικού σκορ.

**Πείραμα 3:** Στο τρίτο πείραμα, οι χάρτες κλίσης παρήχθησαν με τις πρώτες διαφορές στις 45 και 135 μοίρες, αλλά με μειωμένη διάσταση στην εικόνα. Αυτή η προσέγγιση χρησιμοποιήθηκε για τον υπολογισμό του τελικού σκορ.

Παρακάτω παρουσιάζονται τα τελικά αποτελέσματα των πειραμάτων για την αξιολόγηση της νέας μετρικής για τη βάση δεδομένων TID2013. Ο πίνακας περιέχει τις μετρικές SROCC, KROCC, PLCC και RMSE για διάφορες τιμές της παραμέτρου T.

**Πίνακας 3.3.** Επίδοση μεθοδολογίας βασισμένης στις πρώτες διαφορές σε διάφορες κατευθύνσεις. Χρήση διαφορετικών τιμών κατωφλίσωσης  $T=\{0.1, 1, 175, 1000, 10000\}$

Πείραμα 1	T=0.1	T=1	T=175	T=1000	T=10000
SROCC	0.4548	0.4548	0.4573	0.4645	0.4895
KROCC	0.3034	0.3034	0.3050	0.3098	0.3267
PLCC	0.4665	0.4666	0.4686	0.4746	0.4973
RMSE	1.0965	1.0965	1.0952	1.0912	1.0755

Πείραμα 2	T=0.1	T=1	T=175	T=1000	T=10000
SROCC	0.5165	0.5171	0.5640	0.6170	0.6524
KROCC	0.3527	0.3531	0.3878	0.4284	0.4557
PLCC	0.5058	0.5063	0.5519	0.6065	0.6465
RMSE	1.0694	1.0690	1.0338	0.9856	0.9457

Πείραμα 3	T=0.1	T=1	T=175	T=1000	T=10000
SROCC	0.5570	0.5572	0.5819	0.6175	0.6617
KROCC	0.3825	0.3827	0.4018	0.4298	0.4651
PLCC	0.5466	0.5469	0.5708	0.6074	0.6544
RMSE	1.0381	1.0379	1.0179	0.9848	0.9374

Συμπερασματικά, οι καλύτερες αποδόσεις παρατηρήθηκαν ακόμα και μετά τη μείωση των διαστάσεων στα πειράματα, καθώς φαίνεται να παρέχουν εξίσου ακριβείς μετρήσεις με μειωμένο υπολογιστικό φόρτο. Κάτι τέτοιο γενικά είναι αναμενόμενο σε εικόνες του φυσικού κόσμου (φυσικές εικόνες). Η μείωση των διαστάσεων συνέβαλε στη βελτίωση της απόδοσης της μετρικής, ενώ ταυτόχρονα επιτρέπει την ταχύτερη εκτέλεση του αλγορίθμου.

Για την επέκταση του πειράματος, πραγματοποιήθηκε μια επιπλέον επεξεργασία στα δεδομένα μας. Αφαιρέσαμε τη μέση τιμή από τις εικόνες πριν υπολογίσουμε τους χάρτες κλίσης.. Οι τιμές που εξετάστηκαν για το  $T$  ήταν 1, 175 και 1000.

**Πίνακας 3.4.** Επίδοση μεθοδολογίας βασισμένης στις πρώτες διαφορές σε διάφορες κατευθύνσεις αφαιρώντας τη μέση τιμή. Χρήση διαφορετικών τιμών κατωφλίωσης  $T=\{0.1, 1, 175, 1000, 10000\}$

<b>Πείραμα 1</b>	<b>T=1</b>	<b>T=175</b>	<b>T=1000</b>
<b>-mean</b>			
<b>SROCC</b>	0.4837	0.5031	0.5073
<b>KROCC</b>	0.3324	0.3420	0.3424
<b>PLCC</b>	0.4788	0.4961	0.5025
<b>RMSE</b>	1.0883	1.0764	1.0718

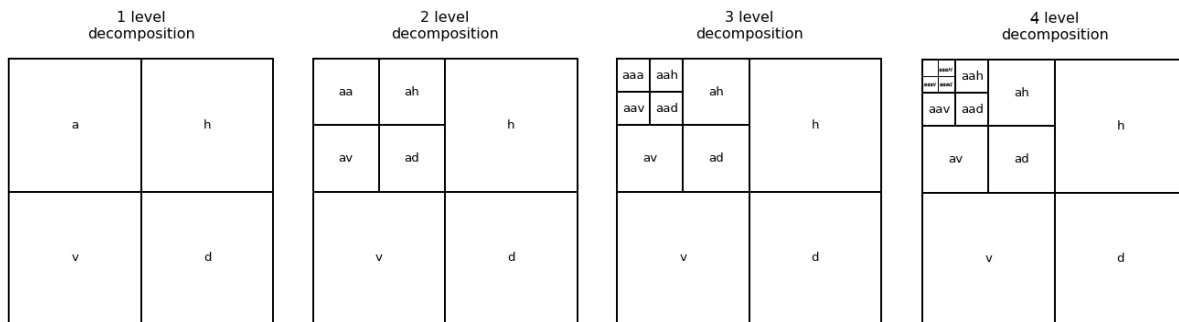
<b>Πείραμα 2</b>	<b>T=1</b>	<b>T=175</b>	<b>T=1000</b>
<b>-mean</b>			
<b>SROCC</b>	0.5987	0.6503	0.6905
<b>KROCC</b>	0.4225	0.4610	0.4905
<b>PLCC</b>	0.6291	0.6605	0.6890
<b>RMSE</b>	0.9636	0.9308	0.8985

<b>Πείραμα 3</b>	<b>T=1</b>	<b>T=175</b>	<b>T=1000</b>
<b>-mean</b>			
<b>SROCC</b>	0.6221	0.6758	0.7183
<b>KROCC</b>	0.4423	0.4833	0.5159
<b>PLCC</b>	0.6499	0.6857	0.7172
<b>RMSE</b>	0.9422	0.9024	0.8639

Όπως μπορούμε να διαπιστώσουμε, αυτή η επεξεργασία βελτίωσε την ακρίβεια της μετρικής μας, καθώς εστίασε στις πραγματικές διαφορές μεταξύ των εικόνων και μείωσε την επίδραση τυχόν μη ουσιαστικών διακυμάνσεων στα δεδομένα.

### 3.8 Τέταρτη ομάδα πειραμάτων - Υπολογισμός visual saliency βασισμένος σε μη-χαμηλοπερατή πληροφορία με πολλά επίπεδα μετασχηματισμού

Στη συγκεκριμένη ομάδα πειραμάτων ακολουθούμε έναν εναλλακτικό τρόπο υπολογισμού της οπτικά εξέχουσας περιοχής εφαρμόζοντας επαναληπτικά τον διδιάστατο μετασχηματισμό κυματιδίων (μέχρι 4 επίπεδα ανάλυσης) και τροποποιώντας ανάλογα τους συντελεστές μετασχηματισμού. Η τροποποίηση των συντελεστών καθορίζει εν τέλει τη νέα οπτική περιοχή η οποία και χρησιμοποιείται ως στάθμιση με σκοπό να προκύψει το τελικό σκορ. Στην εικόνα που ακολουθεί βλέπουμε τη "δομή" ενός μετασχηματισμού 4 επιπέδων.



Εικόνα 3.10. Εφαρμογή διδιάστατου μετασχηματισμού κυματιδίων 4 επιπέδων

**1ο ΒΗΜΑ:** Αρχικά, μετατρέπουμε την εικόνα εισόδου στο CIELab χρωματικό χώρο λόγω της αντιληπτικής ομοιομορφίας που παρουσιάζει.

**2ο ΒΗΜΑ:** Για να αποφευχθεί ο θόρυβος, τα πρότυπα υψής ή τα τεχνουργήματα λόγω κωδικοποίησης (coding artifacts), αρχικά εφαρμόζουμε ένα χαμηλοπερατό φίλτρο στην αρχική εικόνα:

$$X_n = X * G_{m \times m}, \tag{3.13}$$

όπου  $X$  είναι η αρχική/παραμορφωμένη εικόνα,  $X_n$  είναι η εικόνα χωρίς θόρυβο και  $G_{m \times m}$ , είναι ένα διδιάστατο Gaussian φίλτρο. Επιλέγεται ένα μικρό φίλτρο με τιμή  $m=3$ , με σκοπό να φιλτράρει τα συστατικά υψηλότερης συχνότητας.

**3ο ΒΗΜΑ:** Στη συνέχεια εφαρμόζουμε ένα διακριτό μετασχηματισμό κυματιδίων (DWT) με σκοπό να λάβουμε τις υπο-μπάντες της εικόνας εισόδου,



χρησιμοποιώντας ένα Db4 (Daubechies 4) κυματίδιο, με σκοπό να είμαστε όσο το δυνατόν προσαρμοστικοί ως προς τη γειτονιά, τον υπολογιστικό χρόνο και εν τέλει το συνολικό αποτέλεσμα. Έτσι έχουμε:

$$[A^c, H^c, V^c, D^c] = DWT2_N(X_n) \quad (3.14)$$

όπου  $N$  είναι ο μέγιστος αριθμός των επιπέδων αποσύνθεσης,  $s$  είναι ο δείκτης ανάλυσης όπου  $s \in \{1, \dots, N\}$  και  $c$  τα κανάλια επιλογής μας δηλαδή  $c \in \{L, a, b\}$ . Επιπλέον έχουμε  $A_N^c$  που είναι οι συντελεστές προσέγγισης (που περιέχουν την πιο χοντροκομμένη πληροφορία) στο τελευταίο επίπεδο ανάλυσης ή αποσύνθεσης. Επίσης έχουμε σε κάθε επίπεδο, ανάλογα με τα  $s, c$  τα  $H_N^c, V_N^c, D_N^c$ , δηλαδή τις λεπτομερείς οριζόντιες, κάθετες και διαγώνιες υπομπάντες.

Επομένως τι είναι αυτό που συμβαίνει; Οι συντελεστές κυματιδίων στα διάφορα επίπεδα χρησιμοποιούνται με την εφαρμογή του αντίστροφου μετασχηματισμού με σκοπό να δημιουργήσουν τους αντίστοιχους χάρτες. Όμως όπως γνωρίζουμε και όπως έχουμε δει από τα διάφορα πειράματα μέχρι τώρα τα συστατικά χαμηλών συχνοτήτων σε μία εικόνα, συνήθως αναπαριστούν τις περιοχές των αντικειμένων, ενώ τα συστατικά υψηλών συχνοτήτων σε μία εικόνα συνήθως αναπαριστούν τις λεπτομέρειες.

**4ο ΒΗΜΑ:** Προσεγγίζοντας με έναν γενικό τρόπο την έννοια των συστατικών χαμηλής συχνότητας, τότε πρέπει να πούμε ότι στην ουσία είναι το πρώτο ερέθισμα στο ανθρώπινο οπτικό σύστημα. Γι' αυτό το λόγο όταν βλέπουμε μία εικόνα, πρώτα βλέπουμε την περιοχή που αφορά ένα αντικείμενο και στη συνέχεια αρχίζουμε να δίνουμε ενδιαφέρον στις λεπτομέρειες του αντικειμένου. Επομένως εδώ ερχόμαστε σε δύο εκδοχές του πειράματος που προτείνουμε όσον αφορά τους συντελεστές κυματιδίων. Αρχικά εφαρμόζουμε επαναληπτικά τον μετασχηματισμό και εν συνεχεία μηδενίζουμε στο τελευταίο επίπεδο τις πληροφορίες προσέγγισης (χαμηλής συχνότητας) έχοντας κατά νου την εξής λογική. Δεν κρατάμε τις πληροφορίες χαμηλής συχνότητας γιατί στην ουσία θα μας δώσουν εν τέλει – εφαρμόζοντας τον αντίστροφο μετασχηματισμό – μια εκδοχή της αρχικής εικόνας, πιθανόν κάπως θολωμένη. Τι κάνουμε όμως; Αρχίζουμε και τροποποιούμε τους

συντελεστές λεπτομέρειας μέσω διαφόρων τεχνικών προσπαθώντας να δώσουμε έμφαση στις οπτικά εξέχουσες περιοχές.

### **3.8.1 Χρήση πληροφορίας λεπτομέρειας - μηδενισμός πληροφορίας προσέγγισης**

Σύμφωνα με τα προηγούμενα βήματα επιστρέφουμε κρατώντας τις πληροφορίες λεπτομέρειας όπως έχουν. Σε ό, τι ακολουθεί στους δύο επόμενους πίνακες παρουσιάζονται τα αποτελέσματα της σύγκρισης των αλγορίθμων που προτείνουμε και τους ονομάζουμε VCSW και με το ίδιο όνομα VCSW\_name\_ ανάλογα με τις εκδοχές τους που υλοποιούνται σε δύο βάσεις δεδομένων: TID2013, KADID10K.

Στην περίπτωση που θέλουμε να δούμε τι γίνεται με την οπτικά εξέχουσα περιοχή που παράγεται με τη συγκεκριμένη μεθοδολογία στην εικόνα που ακολουθεί, βλέπουμε την εξέχουσα οπτική περιοχή όπου κατά τον αντίστροφο μετασχηματισμό μηδενίζονται οι συντελεστές «χοντροκομμένης» πληροφορίας της εικόνας βάθους Dab (depth) και διατηρούνται οι συντελεστές λεπτομέρειας χωρίς κάποια τροποποίηση.



(α)



(β)

**Εικόνα 3.11:** Visual saliency της προτεινόμενης μεθοδολογίας με χρήση του μετασχηματισμού κυματιδίων στην εικόνα βάθους Dab. (α) αρχική εικόνα (β) visual saliency

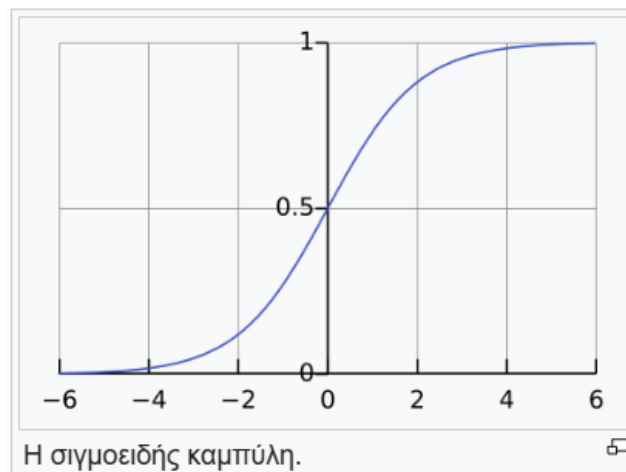
**Πίνακας 3.5.** Επίδοση μεθοδολογίας βασισμένης στην εφαρμογή 4 επιπέδων ανάλυσης μέσω εφαρμογής του δισδιάστατου μετασχηματισμού κυματιδίων, μηδενισμό των συντελεστών προσέγγισης και διατήρηση των συντελεστών λεπτομέρειας

Αλγόριθμος	SROCC	KROCC	PLCC	RMSE	Βάση δεδομένων
VCGS	<b>0.8926</b>	<b>0.7166</b>	<b>0.9000</b>	<b>0.5404</b>	TID2013
VCSW -Vab	0.8769	0.6910	0.8809	0.5868	TID2013
VCSW -Dab	0.8719	0.6862	0.8749	0.6004	TID2013
VCSW - LChannel	0.8753	0.6894	0.8815	0.5853	TID2013
VCSW - AChannel	0.8527	0.6604	0.8615	0.6295	TID2013
VCSW - BChannel	0.8501	0.6578	0.8641	0.6240	TID2013
VCGS	<b>0.8705</b>	<b>0.6828</b>	<b>0.8680</b>	<b>0.5375</b>	KADID
VCSW -Vab	0.8531	0.6638	0.8511	0.5683	KADID
VCSW -Dab	0.8567	0.6684	0.8548	0.5619	KADID
VCSW - LChannel	0.8555	0.6663	0.8531	0.5649	KADID
VCSW - AChannel	0.8401	0.6476	0.8382	0.5904	KADID
VCSW - BChannel	0.8503	0.6595	0.8485	0.5729	KADID

### 3.8.2 Χρήση πληροφορίας λεπτομέρειας με στάθμιση - μηδενισμός πληροφορίας προσέγγισης - χρήση σιγμοειδούς συνάρτησης

Στη συγκεκριμένη ομάδα πειραμάτων αλλάζουμε μεν τις πληροφορίες λεπτομέρειας, δίνοντας όμως μεγαλύτερη βαρύτητα σε εκείνους τους συντελεστές που έχουν συστατικά «σχετικά» χαμηλών συχνοτήτων. Ο σκοπός της τροποποίησης έχει σαν αποτέλεσμα να χρησιμοποιηθεί μία συνάρτηση στάθμισης.

Στη περίπτωση αυτή εφαρμόζουμε μία σιγμοειδή συνάρτηση στους συντελεστές λεπτομέρειας κατά τον αντίστροφο μετασχηματισμό αφού επιστρέψουμε από το 4<sup>ο</sup> επίπεδο μέχρι το 1<sup>ο</sup> επίπεδο. Στο 1<sup>ο</sup> επίπεδο η σιγμοειδής συνάρτηση εφαρμόζεται στις τιμές των συντελεστών λεπτομέρειας.



Εικόνα 3.12. Σιγμοειδής συνάρτηση

Που όμως εφαρμόζονται όλα αυτά; Αρχικά εφαρμόζονται σε κάθε κανάλι L, a, b ξεχωριστά ενώ σε μία άλλη εκδοχή του πειράματος εφαρμόζονται στη ζωντάνια ( $V^*ab$ ) και το βάθος ( $D^*ab$ ) στον χρωματικό χώρο CIELAB. Στον επόμενο πίνακα εφαρμόζουμε τη σιγμοειδή συνάρτηση στους συντελεστές ενδιαφέροντος.

Πίνακας 3.6. Επίδοση μεθοδολογίας βασισμένης στην εφαρμογή 4 επιπέδων ανάλυσης μέσω εφαρμογής του δισδιάστατου μετασχηματισμού κυματιδίων, μηδενισμό των συντελεστών προσέγγισης και διατήρηση των συντελεστών λεπτομέρειας. Στο τελευταίο στάδιο εφαρμόζεται σιγμοειδής συνάρτηση.

Αλγόριθμος	SROCC	KROCC	PLCC	RMSE	Μετρική
VCGS	<b>0.8926</b>	<b>0.7166</b>	<b>0.9000</b>	<b>0.5404</b>	TID2013
VCSW_sigmoid -Vab	0.8753	0.6896	0.8808	0.5869	TID2013
VCSW_sigmoid -Dab	0.8754	0.6893	0.8805	0.5877	TID2013
VCSW_sigmoid -LChannel	0.8769	0.6914	0.8823	0.5835	TID2013
VCSW_sigmoid -AChannel	0.8671	0.6774	0.8744	0.6014	TID2013
VCSW_sigmoid -BChannel	0.8674	0.6776	0.8763	0.5973	TID2013
VCGS	<b>0.8705</b>	<b>0.6828</b>	<b>0.8680</b>	<b>0.5375</b>	KADID
VCSW_sigmoid -Vab	0.8558	0.6662	0.8532	0.5646	KADID
VCSW_sigmoid -Dab	0.8559	0.6661	0.8534	0.5643	KADID
VCSW_sigmoid -LChannel	0.8533	0.6632	0.8507	0.5691	KADID
VCSW_sigmoid -AChannel	0.8435	0.6511	0.8414	0.5851	KADID
VCSW_sigmoid -BChannel	0.8532	0.6623	0.8513	0.5680	KADID

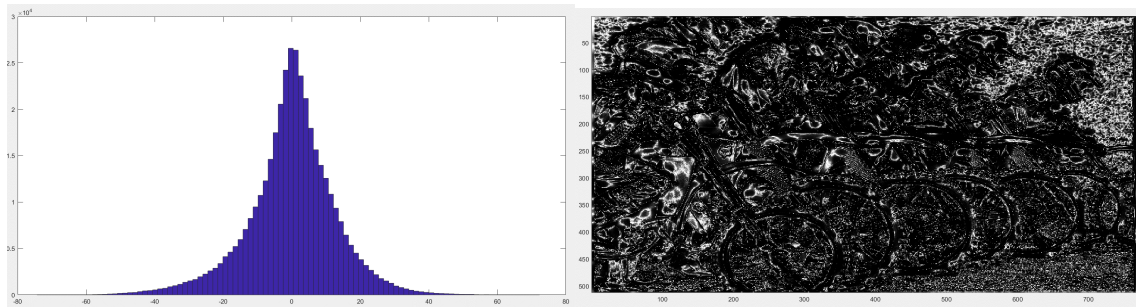
Στη περίπτωση εφαρμογής της σιγμοειδούς συνάρτησης η οπτικά εξέχουσα περιοχή είναι της μορφής:



**Εικόνα 3.13.** Visual saliency της προτεινόμενης μεθοδολογίας με εφαρμογή της σιγμοειδούς συνάρτησης στους συντελεστές ενδιαφέροντος με χρήση της εικόνας βάθους Dab

### 3.8.3 Χρήση πληροφορίας λεπτομέρειας - μηδενισμός πληροφορίας προσέγγισης - προσαρμογή με Student't-t κατανομή

Στη συγκεκριμένη ομάδα πειραμάτων, εξακολουθούμε να ακολουθούμε τη φιλοσοφία των δύο προηγούμενων ομάδων πειραμάτων -όπου γίνεται χρήση 4 επιπέδων ανάλυσης με τη βοήθεια του μετασχηματισμού wavelet. Η διαφορά έγκειται στο ότι κατά την εφαρμογή του αντίστροφου μετασχηματισμού όταν λαμβάνουμε εν τέλει την 'αρχική' εικόνα τότε αυτή από τη μία υποδηλώνει ένα είδος εξέχουσας οπτικής περιοχής από την άλλη όμως συνιστά και ένα σύνολο δεδομένων που η μορφή του ιστογράμματος έχει μορφή που μπορεί να προσαρμοστεί από την στατιστική κατανομή Student's -t. Για να γίνει πιο κατανοητό κάτι τέτοιο, εμφανίζουμε στις εικόνες που ακολουθούν τη μορφή των δεδομένων αφού έχει εφαρμοστεί ο αντίστροφος μετασχηματισμός κυματιδίων μέχρι να ανακατασκευαστεί η 'αρχική' εικόνα αλλά και η προτεινόμενη εξέχουσα περιοχή. Πρέπει να σημειωθεί ότι οι 'κρυφές' μεταβλητές κατά την εκτίμηση των παραμέτρων της Student's t όπως αυτή ορίστηκε στα προηγούμενα βάσει του EM αλγορίθμου μας δίνουν τον χάρτη των εξέχουσών οπτικών περιοχών (visual salient areas).



**Εικόνα 3.14.** Visual saliency της προτεινόμενης μεθοδολογίας με εφαρμογή της προσαρμοσμένης Student's-t στατιστικής κατανομής και τη δημιουργία visual saliency μέσω των 'hidden' μεταβλητών κατά την εκτίμηση των παραμέτρων



**Πίνακας 3.7.** Επίδοση μεθοδολογίας βασισμένης στην εφαρμογή 4 επιπέδων ανάλυσης μέσω εφαρμογής του δισδιάστατου μετασχηματισμού κυματιδίων, μηδενισμό των συντελεστών προσέγγισης και διατήρηση των συντελεστών λεπτομέρειας. Στο τελευταίο στάδιο εφαρμόζεται προσαρμογή (fit) της Student's-t και εν συνεχεία προτείνονται οι visual salient περιοχές.

Αλγόριθμος	SROCC	KROCC	PLCC	RMSE	Μετρική
VCGS	<b>0.8926</b>	<b>0.7166</b>	<b>0.9000</b>	<b>0.5404</b>	TID2013
VCSW t_maxall Vab	0.8557	0.6594	0.8488	0.6555	TID2013
VCSW t_maxall Dab	0.8611	0.6661	0.8558	0.6412	TID2013
VCSW t_maxall LChannel	0.8570	0.6606	0.8495	0.6541	TID2013
VCSW t_maxall AChannel	0.8673	0.6781	0.8694	0.6125	TID2013
VCSW t_maxall BChannel	0.8697	0.6813	0.8714	0.6082	TID2013

## ΚΕΦΑΛΑΙΟ 4

### ΕΙΚΟΝΕΣ ΜΕ ΠΕΡΙΕΧΟΜΕΝΟ ΟΘΟΝΩΝ (SCREEN CONTENT IMAGES - SCI)

#### 4.1 Screen Content Images

Οι SCI (screen content images) εικόνες αποτελούν ένα είδος εικόνας που αναδύεται στο διαδίκτυο και αντιπροσωπεύουν πολύπλοκα περιεχόμενα οθόνης. Αυτές οι εικόνες είναι συνήθως μια ανάμειξη εικονογραφημάτων και γραφικών με κείμενο που παράγεται από υπολογιστή. Το κείμενο αυτό μπορεί να είναι πληκτρολογημένο ή παραγόμενο από προγράμματα, ενώ τα γραφικά μπορεί να περιλαμβάνουν σχήματα, διαγράμματα, ή άλλα είδη εικονογραφήματος.

Οι SCI έχουν ευρεία χρήση σε διάφορες εφαρμογές πολυμέσων, όπως συστήματα κοινής χρήσης πληροφοριών μεταξύ υπολογιστών και έξυπνων συσκευών, συστήματα υπολογιστικού νέφους, τηλεδιάσκεψης, και διαφήμισης προϊόντων. Η αυξημένη δημοτικότητα των έξυπνων τηλεφώνων οδήγησε σε αυξημένη ζήτηση για τη δημιουργία και κοινοποίηση πληροφοριών σε μορφή SCI. Από τη στιγμή που η απομακρυσμένη πρόσβαση, ο διαμοιρασμός της οθόνης ή ακόμη και το gaming εμπεριέχουν έννοιες μετάδοσης, λήψης ή συμπίεσης των SCI εικόνων, είναι προφανές ότι η διατήρηση της ποιότητάς τους ενισχύει την ποιότητα της εμπειρίας που αποκομίζει ο τελικός χρήστης.



**EIKONA 4.1 :** Παραδείγματα εικόνων SCI (Screen Content Images)

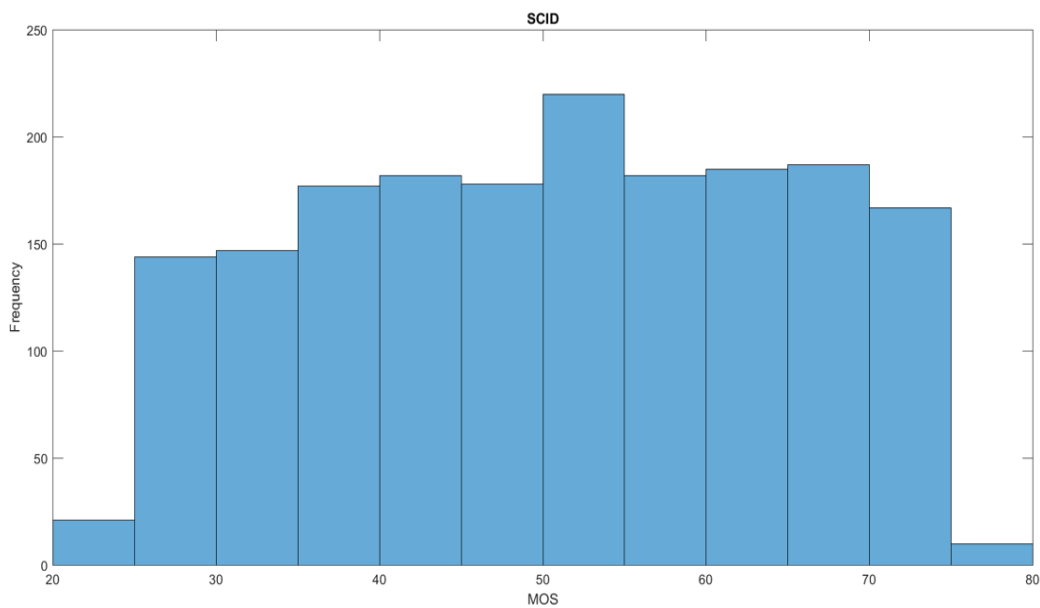
Οι SCI διαφέρουν από τις φυσικές εικόνες, καθώς συχνά περιέχουν επαναλαμβανόμενα μοτίβα, λεπτές γραμμές και κοφτές γωνίες, ενώ οι φυσικές εικόνες συνήθως περιλαμβάνουν λείες γραμμές, πυκνές γραμμές και περισσότερες πληροφορίες χρώματος. Ως εκ τούτου, οι υφιστάμενες μετρικές αξιολόγησης ποιότητας εικόνας που σχεδιάστηκαν για φυσικές εικόνες ενδέχεται να μην είναι αποτελεσματικές για την αξιολόγηση της ποιότητας των SCI. Για να αντιμετωπιστεί αυτό το πρόβλημα, έχουμε οδηγηθεί στην ανάπτυξη νέων μεθόδων αξιολόγησης ποιότητας που λαμβάνουν υπόψη τις ιδιαιτερότητες των SCI, όπως οι επαναλαμβανόμενες δομές και οι λεπτές γραμμές. [33]

Επιπρόσθετα, σε διάφορα σενάρια εφαρμογής, όπως συστήματα απομακρυσμένης επεξεργασίας και εικονικής επιφάνειας εργασίας, η οθόνη ανανεώνεται, συμπιέζεται και μεταδίδεται στον ανθρώπινο παρατηρητή. Η ανάγκη για αντικειμενικές μεθόδους αξιολόγησης ποιότητας που να είναι ικανές να εκτιμήσουν την αντιληπτή ποιότητα των SCI είναι σημαντική, καθώς μπορούν να χρησιμοποιηθούν τόσο για την παρακολούθηση της ποιότητας της οθόνης σε απομακρυσμένα συστήματα υπολογιστών όσο και για την ανάπτυξη και βελτιστοποίηση αλγορίθμων επεξεργασίας εικόνας/βίντεο.

Συνεπώς είναι απαραίτητο ένα μοντέλο που θα λαμβάνει υπόψη, όχι μόνο τις επαναλαμβανόμενες δομές και τις λεπτές γραμμές των SCI όπως αναφέρθηκε παραπάνω, αλλά και την πρωτεύουσα οπτική πληροφορία και την αβεβαιότητα. [34]

## 4.2 Βάση Δεδομένων SCID

Η βάση δεδομένων SCID περιλαμβάνει 40 εικόνες αναφοράς, σε συνδυασμό με 1.800 παραμορφωμένες εικόνες. Οι παραμορφωμένες εικόνες προκύπτουν από τις αρχικές 40 με τη χρήση 9 διαφορετικών τύπων παραμορφώσεων και 5 επιπέδων παραμορφώσεων. Οι SCIs (εικόνες περιεχομένου οθόνης) αποτελούνται από έναν συνδυασμό κειμένων, διαγραμμάτων, λογοτύπων, γραφικών κ.λπ. Παρουσιάζεται το ιστόγραμμα πληροφορίας του μέτρου MOS.



**EIKONA 4.2 :** οπτική αναπαράσταση της συχνότητας εμφάνισης των τιμών MOS της βάσης δεδομένων SCID.



**EIKONA 4.3:** Δείγμα εικόνων της βάσης δεδομένων SCID.

### 4.3 Πειράματα για SCI

Παρακάτω παρουσιάζονται τα αποτελέσματα όλων των πειραμάτων για τις γνωστές μετρικές MDSI, ESSIM, SG\_ESSIM, VCGS και VSI, για την βάση δεδομένων SCI.

**Πίνακας 4.1** Επίδοση γνωστών μεθοδολογιών στην σχετική με εικόνες που εμφανίζουν περιεχόμενο οθονών (SCI)

Μετρική	MDSI	ESSIM	SG_ESSIM	VCGS	VSI
SROCC	0.7806	0.6993	0.7641	<b>0.8256</b>	0.7621
KROCC	0.5822	0.4959	0.5651	<b>0.6296</b>	0.5649
PLCC	0.7883	0.7074	0.7829	<b>0.8277</b>	0.7698
RMSE	8.7135	10.0101	8.8109	<b>7.9483</b>	9.0403

#### 4.3.1 Τροποποίηση της μεθοδολογίας VSSC με πρόσθεση χαρακτηριστικών υφής

Έπειτα, επαναλαμβάνονται όλα τα πειράματα για την SCID, τα οποία διεξήχθησαν για τις βάσεις δεδομένων TID2013 και KADID-10K, όπως παρουσιάστηκαν στα προηγούμενα κεφάλαια. Αυτά τα πειράματα δεν περιλαμβάνουν τις τυπικές μετρικές που χρησιμοποιούνται ευρέως, αλλά αντίθετα προσθέτουν τροποποιημένες μετρικές που αναπτύχθηκαν σε αυτήν τη μελέτη.

Πίνακας 4.2 Επίδοση μετρικών της εργασίας στην σχετική με εικόνες που εμφανίζουν περιεχόμενο οθονών (SCI)

<i>Αλγόριθμος</i>	<b>SROCC</b>	<b>KROCC</b>	<b>PLCC</b>	<b>RMSE</b>
<i>VCGS</i>	0.8256	0.6296	0.8277	7.9483
<i>VSSC</i>	0.8102	0.6092	0.8078	8.3474
<i>VSSC+bdip</i>	0.8097	0.6085	0.8073	8.3584
<i>VSSC</i> <i>Διατήρηση cA (κανάλι a)</i>	0.7878	0.5880	0.7901	8.6813
<i>VSSC</i> <i>Μηδενισμός cA (κανάλι a)</i>	<b>0.8113</b>	<b>0.6096</b>	<b>0.8119</b>	<b>8.2677</b>
<i>VSSC</i> <i>Διατήρηση cA (κανάλι b)</i>	0.8005	0.6000	0.8040	8.4203
<i>VSSC</i> <i>Μηδενισμός cA (κανάλι b)</i>	0.8108	0.6088	0.8115	8.2764

Με βάση τον παραπάνω πίνακα παρατηρούμε τα εξής: Αρχικά, ενώ η SCID παρουσιάζει σημαντικά υψηλότερες τιμές RMSE σε σύγκριση με άλλες βάσεις δεδομένων, δεύτερον, το χαρακτηριστικό bdip φαίνεται να έχει ελάχιστη επίδραση στη βελτίωση της VSSC σε άλλες βάσεις δεδομένων, αλλά στην περίπτωση της SCID προκαλεί μια σημαντική μείωση. Τρίτον, ενώ η προσέγγιση στο κανάλι b οδήγησε σε βελτιωμένες αποδόσεις για άλλες βάσεις δεδομένων, αυτό δεν συμβαίνει στην περίπτωση της SCID.

### 4.3.2 Μεθοδολογία με χρήση πρώτων διαφορών - πληροφορία κλίσης

Στην προσπάθεια επέκτασης των πειραμάτων στη βάση δεδομένων SCI, εφαρμόσαμε παρόμοια προσέγγιση με προηγούμενο κεφάλαιο. Κάθε πείραμα περιλάμβανε την αξιολόγηση δύο εικόνων: μιας αρχικής και μιας παραμορφωμένης εικόνας.

**Πείραμα 1:** Οι χάρτες κλίσης υπολογίστηκαν βάσει των πρώτων διαφορών στις 0 και 90 μοίρες, όπως και στην προηγούμενη βάση δεδομένων.

**Πείραμα 2:** Οι χάρτες κλίσης παράχθηκαν με τις πρώτες διαφορές στις 0 και 90 μοίρες, αλλά με μειωμένη διάσταση στις εικόνες.

**Πείραμα 3:** Οι χάρτες κλίσης παρήχθησαν με τις πρώτες διαφορές στις 45 και 135 μοίρες, επίσης με μειωμένη διάσταση στις εικόνες.

Επίσης αφαιρέσαμε τη μέση τιμή από τις εικόνες πριν υπολογίσουμε τους χάρτες κλίσης. Οι παράμετροι που εξετάστηκαν για το  $T$  ήταν 1, 175 και 1000.

**Πίνακας 4.3** Επίδοση της μεθοδολογίας που χρησιμοποιεί πληροφορία κλίσης χρησιμοποιώντας πρώτες διαφορές στην σχετική με εικόνες που εμφανίζουν περιεχόμενο οθονών

<b>Πείραμα 1</b>	<b>T=1</b>	<b>T=175</b>	<b>T=1000</b>
SROCC	0.6119	0.6297	0.6500
KROCC	0.4371	0.4513	0.4674
PLCC	0.6443	0.6599	0.6769
RMSE	10.8311	10.6409	10.4251
<b>Πείραμα 1 - mean</b>	<b>T=1</b>	<b>T=175</b>	<b>T=1000</b>
SROCC	0.5791	0.6400	0.6762
KROCC	0.4092	0.4597	0.4903
PLCC	0.6121	0.6638	0.6938
RMSE	11.1987	10.5923	10.1993
<b>Πείραμα 2</b>	<b>T=1</b>	<b>T=175</b>	<b>T=1000</b>
SROCC	0.6683	0.7188	0.7705
KROCC	0.4845	0.5275	0.5744
PLCC	0.6860	0.7290	0.7750
RMSE	10.3047	9.6950	8.9502
<b>Πείραμα 2 - mean</b>	<b>T=1</b>	<b>T=175</b>	<b>T=1000</b>
SROCC	0.7155	0.8076	0.8456
KROCC	0.5336	0.6110	0.6497
PLCC	0.7202	0.8054	0.8444
RMSE	9.8248	8.3952	7.5858
<b>Πείραμα 3</b>	<b>T=1</b>	<b>T=175</b>	<b>T=1000</b>
SROCC	0.6909	0.7204	0.7601
KROCC	0.5051	0.5298	0.5653
PLCC	0.7066	0.7316	0.7660
RMSE	10.0218	9.6545	9.1037
<b>Πείραμα 3 - mean</b>	<b>T=1</b>	<b>T=175</b>	<b>T=1000</b>



<b>SROCC</b>	0.7113	0.8079	<b>0.8500</b>
<b>KROCC</b>	0.5282	0.6106	<b>0.6543</b>
<b>PLCC</b>	0.7156	0.8063	<b>0.8489</b>
<b>RMSE</b>	9.8921	8.3770	<b>7.4865</b>

Με την αφαίρεση του μέσου, παρατηρήσαμε σημαντική βελτίωση στα αποτελέσματα. Ιδιαίτερα στην περίπτωση της παραμέτρου  $T=1000$ , σημειώθηκε όχι μόνο βελτίωση αλλά και υπέρβαση της καλύτερης μέχρι τώρα μετρικής, της VCGS. Ακόμα, για τις υπόλοιπες τιμές του  $T$ , παρατηρήθηκε ότι οι μετρικές ήταν πολύ κοντά στις προβλέψεις της VCGS, υποδεικνύοντας την αξιοπιστία της μεθόδου που χρησιμοποιήθηκε.

#### 4.3.3 Χρήση πληροφορίας λεπτομέρειας με στάθμιση - μηδενισμός πληροφορίας προσέγγισης

Στη συγκεκριμένη ενότητα προτείνεται η εφαρμογή της μεθοδολογίας όπου κάνουμε χρήση πολλών επιπέδων ανάλυσης με την εφαρμογή του μετασχηματισμού κυματιδίων 4 επιπέδων. Επιπλέον κάνουμε έλεγχο χρησιμοποιώντας είτε τη πληροφορία από το χρωματικό μοντέλο CIELAB που σχετίζεται με τη 'ζωντάνια'/Vividness ή το 'βάθος'/Depth καθώς και κάθε κανάλι του συγκεκριμένου χρωματικού μοντέλου ξεχωριστά (LChannel, AChannel, BChannel).

**Πίνακας 4.4** Επίδοση της μεθοδολογίας που χρησιμοποιεί πληροφορία λεπτομερειών μετά την εφαρμογή του μετασχηματισμού κυματιδίων 4 επιπέδων.

<b>VCGS</b>	<b>0.8256</b>	<b>0.6296</b>	<b>0.8277</b>	<b>7.9483</b>	<b>SCID</b>
<b>VCSW -Vab</b>	0.8099	0.6114	0.8135	8.2363	SCID
<b>VCSW -Dab</b>	0.8157	0.6178	0.8191	8.1249	SCID
<b>VCSW - LChannel</b>	0.8150	0.6164	0.8180	8.1457	SCID
<b>VCSW - AChannel</b>	0.7839	0.5851	0.7859	8.7574	SCID
<b>VCSW - BChannel</b>	0.7933	0.5930	0.7943	8.6035	SCID

#### **4.3.4 Χρήση πληροφορίας λεπτομέρειας με στάθμιση - μηδενισμός πληροφορίας προσέγγισης - χρήση σιγμοειδούς συνάρτησης**

Στη συγκεκριμένη ενότητα προτείνεται η εφαρμογή της μεθοδολογίας όπου κάνουμε χρήση στάθμισης και εφαρμογή της σιγμοειδούς συνάρτησης στους συντελεστές του αντίστροφου μετασχηματισμού κυματιδίων.

**Πίνακας 4.5** Επίδοση της μεθοδολογίας που χρησιμοποιεί πληροφορία λεπτομερειών μετά την εφαρμογή του μετασχηματισμού κυματιδίων 4 επιπέδων. Κατά την αντιστροφή στο τελευταίο επίπεδο χρησιμοποιείται η σιγμοειδής συνάρτηση.

<b>VCGS</b>	<b>0.8256</b>	<b>0.6296</b>	<b>0.8277</b>	<b>7.9483</b>	<b>SCID</b>
<b>VCSW_sig-moid -Vab</b>	0.8155	0.6167	0.8188	8.1296	SCID
<b>VCSW_sig-moid -Dab</b>	0.8094	0.6103	0.8124	8.2586	SCID
<b>VCSW_sig-moid -LChannel</b>	0.8127	0.6134	0.8157	8.1921	SCID
<b>VCSW_sig-moid -AChannel</b>	0.7926	0.5932	0.7955	8.5814	SCID
<b>VCSW_sig-moid -BChannel</b>	0.8012	0.6018	0.8033	8.4343	SCID

#### 4.4 Πειράματα για SCI

Με βάση τα προηγούμενα πειράματα συμπεραίνουμε τα εξής:

Ο αλγόριθμος VCSW χωρίς τη χρήση της σιγμοειδούς παρουσιάζει γενικά υψηλές τιμές SROCC και PLCC, αλλά ελαφρώς χαμηλότερες από τον VCGS. Οι διάφορες εκδόσεις του VCSW (Vab, Dab, LChannel, AChannel, BChannel) έχουν παραπλήσιες επιδόσεις, με μικρές διακυμάνσεις μεταξύ τους.

Η προσθήκη της σιγμοειδούς βελτιώνει ελαφρώς τις τιμές PLCC και RMSE σε ορισμένες περιπτώσεις, αλλά δεν φαίνεται να υπάρχει σημαντική διαφορά στις τιμές SROCC και KROCC. Οι εκδόσεις με sigmoid έχουν επίσης παραπλήσιες επιδόσεις μεταξύ τους, με την LChannel να φαίνεται και πάλι να έχει την καλύτερη απόδοση.

## ΚΕΦΑΛΑΙΟ 5

### ΣΥΜΠΕΡΑΣΜΑΤΑ

Στην παρούσα εργασία, βασικό αντικείμενό της ήταν η ανάπτυξη υπολογιστικών μεθοδολογιών στο πρόβλημα της αξιολόγησης της ποιότητας των εικόνων έχοντας πλήρη πρόσβαση στα δεδομένα της αρχικής -μη υποβαθμισμένης- εικόνας. Στο πλαίσιο αυτό προτάθηκε και διερευνήθηκε η χρήση οπτικά εξεχουσών περιοχών (visual salient regions) τόσο κατά τον υπολογισμό των διαφόρων χαρτών ομοιότητας που προέκυπταν όσο και κατά το τελικό στάδιο της συγκέντρωσης (pooling) των τιμών και την εξαγωγή του τελικού σκορ. Να σημειωθεί ότι η ανθρώπινη προσοχή που συνδέεται με την έννοια των οπτικά εξεχουσών περιοχών δεν αποδίδει μόνο το πως λαμβάνει μία οπτική σκηνή το ανθρώπινο μάτι αλλά υποδεικνύει τη σημασία κάθε pixel στο πλαίσιο της πρότασης μίας νέας μετρικής.

Για να επιτευχθεί κάτι τέτοιο διερευνήθηκε η χρήση του μετασχηματισμού wavelet επιλέγοντας το CIELAB χρωματικό μοντέλο γιατί είναι πιο κοντά στην ανθρώπινη διαίσθηση των χρωμάτων. Επιπλέον, χρησιμοποιήθηκαν οι στατιστικές ιδιότητες των συντελεστών ενδιαφέροντος από τον προαναφερόμενο μετασχηματισμό οπότε και εκμεταλλευτήκαμε την προσαρμογή στατιστικών κατανομών όπως η Student's t. Έτσι πραγματοποιήθηκαν πολλά πειράματα διερευνώντας διάφορες παραμέτρους του εν λόγω προβλήματος.

Πρέπει να σημειωθεί ότι εκτός από τις γνωστές βάσεις δεδομένων που σχετίζονται με το υποσύνολο των φυσικών εικόνων, πραγματοποιήθηκαν πειράματα με σύνολα δεδομένων όπου οι εικόνες έχουν περιεχόμενο που σχετίζεται με το τι βλέπουν οι χρήστες στις οθόνες (screen content images).

Η ανάλυση των πειραμάτων κατέδειξε διάφορα συμπεράσματα σχετικά με την επίδραση των διαφορετικών χαρακτηριστικών και μεθοδολογιών στις βάσεις δεδομένων TID2013, KADID-10K και SCID.

Στην πρώτη ομάδα πειραμάτων, διαπιστώθηκε ότι η προσθήκη του χαρακτηριστικού BDIP (Block Difference of Inverse Probabilities) βελτιώνει σημαντικά την πρόβλεψη της ποιότητας της εικόνας. Οι μετρικές απόδοσης, όπως το SROCC, το KROCC και το PLCC, παρουσίασαν βελτίωση και στις τρεις βάσεις δεδομένων, ενώ το RMSE παρέμεινε σε αποδεκτά επίπεδα. Αντίθετα, η προσθήκη του χαρακτηριστικού BVLC (Block Variation of Local Correlation Coefficients) δεν αποδείχθηκε επωφελής, καθώς δεν βελτίωσε τις μετρικές απόδοσης και αύξησε το RMSE, υποδεικνύοντας μεγαλύτερη αβεβαιότητα στις προβλέψεις.

Στη δεύτερη ομάδα πειραμάτων, τα αποτελέσματα έδειξαν ότι οι διάφορες βάσεις δεδομένων επηρεάζονται διαφορετικά από την παρουσία ή την απουσία των συντελεστών προσέγγισης ( $cA$ ). Για παράδειγμα, ο μηδενισμός των  $cA$  στο κανάλι  $b$  έδωσε καλύτερα αποτελέσματα στη βάση δεδομένων TID2013, ενώ η διατήρηση των  $cA$  στο κανάλι  $b$  ήταν πιο αποδοτική στη βάση δεδομένων KADID-10K. Στη βάση δεδομένων SCID, ο μηδενισμός των  $cAs$  βελτίωσε την ακρίβεια των μετρήσεων σε σύγκριση με τη διατήρησή του. Συνολικά, δεν κατέστη δυνατή η εξαγωγή ενός ενιαίου συμπεράσματος για όλες τις περιπτώσεις, υποδεικνύοντας ότι η επίδραση του  $cA$  εξαρτάται από τη συγκεκριμένη βάση δεδομένων και τις παραμορφώσεις της.

Η τρίτη ομάδα πειραμάτων έδειξε ότι η μείωση των διαστάσεων και η αφαίρεση της μέσης τιμής οδήγησαν σε βελτιωμένα αποτελέσματα, παρέχοντας ακριβέστερες μετρήσεις και ταχύτερη εκτέλεση των αλγορίθμων. Η χρήση πρώτων διαφορών σε διάφορες γωνίες (0, 45, 90, 135 μοίρες) βελτίωσε την ακρίβεια των προβλέψεων της ποιότητας εικόνας και στις τρεις βάσεις δεδομένων, με τις βελτιώσεις να είναι πιο εμφανείς στη βάση δεδομένων SCID.

Στην τέταρτη ομάδα πειραμάτων, η εφαρμογή της σιγμοειδούς συνάρτησης για την τροποποίηση των συντελεστών λεπτομέρειας αποδείχθηκε ιδιαίτερα αποδοτική. Τα αποτελέσματα έδειξαν ότι αυτή η μεθοδολογία είχε υψηλή αξιοπιστία και ακρίβεια, παρουσιάζοντας συγκρίσιμες ή καλύτερες επιδόσεις σε σχέση με τις καθιερωμένες μετρικές. Στη βάση δεδομένων TID2013, η μετρική με την εφαρμογή της σιγμοειδούς συνάρτησης είχε SROCC 0.8769 και PLCC 0.8823, ενώ στη βάση δεδομένων KADID-10K είχε SROCC 0.8705 και PLCC 0.8680. Στη βάση δεδομένων

SCID, η χρήση της σιγμοειδούς συνάρτησης βελτίωσε επίσης την ακρίβεια και την αξιοπιστία των μετρήσεων ποιότητας εικόνας.

Συνολικά, τα αποτελέσματα των πειραμάτων δείχνουν ότι η επιλογή και ο συνδυασμός των καναλιών του χρωματικού χώρου CIELAB (L, a, b) και των χαρακτηριστικών βάθους και ζωντάνιας επηρεάζουν διαφορετικά την αξιολόγηση της ποιότητας εικόνας στις βάσεις δεδομένων TID2013, KADID-10K και SCID. Ανάλογα με το κάθε πείραμα, διαφορετικά κανάλια και χαρακτηριστικά αποδεικνύονται πιο αποδοτικά, βελτιώνοντας την ακρίβεια και αξιοπιστία των αλγορίθμων. Η κατάλληλη επιλογή και συνδυασμός των καναλιών και χαρακτηριστικών μπορεί να βελτιώσει σημαντικά την απόδοση των αλγορίθμων αξιολόγησης ποιότητας εικόνας, παρέχοντας πιο ακριβείς και συνεπείς μετρήσεις.

### **5.1 Συμπεράσματα 3<sup>ου</sup> Κεφαλαίου και Μελλοντική εργασία**

Στη συγκεκριμένη εργασία στα πρώτα δύο κεφάλαια διερευνήθηκε η τρέχουσα κατάσταση στον τομέα της αξιολόγησης της ποιότητας των εικόνων, αλλά και τις προκλήσεις που πρέπει να αντιμετωπίσουν οι ερευνητές του τομέα. Έτσι έγινε αναφορά σε μεθοδολογίες, σύνολα δεδομένων και τρόπους προσέγγισης του προβλήματος της αξιολόγησης της ποιότητας της εικόνας. Διερευνήθηκαν διάφορες μετρικές πλήρους αναφοράς (full-reference) για την αντικειμενική αξιολόγηση της αντιληπτικής ποιότητας των εικόνων. Βασικό συστατικό στοιχείο ήταν η διερεύνηση της εφαρμογής των οπτικά εξεχουσών περιοχών (visual salient areas) και η εφαρμογή τους στο πλαίσιο του προς εξέταση προβλήματος της αξιολόγησης της ποιότητας. Σε αυτό το πλαίσιο διερευνήθηκε η εφαρμογή της Student's-t ως στατιστικής κατανομής που μπορεί να περιγράψει τους συντελεστές του μετασχηματισμού κυματιδίων. Η κατάλληλη τροποποίηση των τιμών των διαφόρων συντελεστών του ακολουθώντας κάποιους κανόνες. Οι προτεινόμενες μετρικές είχαν πολύ καλή απόδοση. Με βάση τα πειράματα που διεξήχθησαν μπορούμε να πούμε τα εξής:

(1) Αρχικά σε όλα τα πειράματα χρησιμοποιείται μια απλή με κριτήριο την ερμηνευσιμότητα προσέγγιση στον ορισμό της μετρικής. Εχμεταλλεύεται δομικά

συστατικά των εικόνων μέσω της κλίσης και συστατικά χρωματικότητας από το επιλεγμένο χρωματικό μοντέλο.

(2) Επιλέγεται η διερεύνηση συγκεκριμένων ιδιοτήτων ή καναλιών του χρωματικού μοντέλου CIELAB καθώς και η επίδοση της μετρικής μέσω της χρήσης του μετασχηματισμού wavelet καθώς και της χρήσης της στατιστικής κατανομής Student's -t στις ιδιότητες της «ζωντάνιας» και του «βάθους».

(3) Διερευνάται η χρήση χαρακτηριστικών υφής και συμπεραίνουμε την επίδρασή τους στην επίδοση της προτεινόμενης μετρικής.

(4) Εξετάστηκε η επιλογή διαφόρων χρωματικών καναλιών.

(5) Η δημιουργία χαρτών ομοιότητας με τη χρήση της στατιστικής κατανομής μας έδωσε την δυνατότητα της στάθμισης ανάλογα με τον προτεινόμενη περιοχή εξέχουσας περιοχής.

Από άποψη χρονικής πολυπλοκότητας όλοι οι εξεταζόμενοι αλγόριθμοι είναι αποδοτικοί και συγκρίσιμης ταχύτητας για παράδειγμα με τον αλγόριθμο VCGS ή VSSC. Όταν όμως χρησιμοποιείται η προσαρμογή της Student's t στους wavelet συντελεστές τότε εμφανίζεται μία σχετική καθυστέρηση.

Όπως και να έχει υπάρχουν ορισμένοι περιορισμοί στην προσέγγισή μας και ευκαιρίες βελτίωσης του αλγορίθμου. Αυτοί σχετίζονται με την εφαρμογή διαφορετικών στρατηγικών συγκέντρωσης οι οποίοι μπορεί εν δυνάμει να αυξήσουν την επίδοση. Με άλλα λόγια ο κατάλληλος συνδυασμός των χαρτών που προκύπτουν και η περαιτέρω στάθμιση μπορεί να δώσει καλύτερα αποτελέσματα.

## 5.2 Συμπεράσματα 4<sup>ου</sup> Κεφαλαίου και Μελλοντική εργασία

Οι περισσότερες υπάρχουσες προσεγγίσεις αξιολόγησης επινοήθηκαν για φυσικές εικόνες. Διερευνήθηκε η αποτελεσματικότητά τους στην αξιολόγηση της ποιότητας SCI χωρίς να ληφθεί υπόψη η διαφορά στο περιεχόμενο και τα χαρακτηριστικά της εικόνας μεταξύ των SCI και των φυσικών εικόνων. Αναγνωρίζοντας αυτή τη δυσκολία, χαρακτηριστικά που σχετίζονται με

- τις διαφορετικές αναλογίες περιεχομένου συνεχούς τόνου (δηλ. σχετικά ομαλές ακμές και χρώματα φυσικών σκηνών)

- ασυνεχείς τόνοι περιεχομένου (δηλ. εξαιρετικά αιχμηρές άκρες με περιορισμένες χρωματικές παραλλαγές των κειμένων, σχημάτων και γραφημάτων) μπορεί να δώσουν διαφορετικά αποτελέσματα.

Η συνεισφορά του συγκεκριμένου κεφαλαίου συνίσταται

(1) στη χρήση μιας απλής και ερμηνεύσιμης προσέγγισης για να αξιολογηθεί η ποιότητα της εικόνας περιεχομένου οθόνης,

(2) στη διερεύνηση των καναλιών χρωματισμού του CIELAB χρωματικού μοντέλο

Για το λόγο αυτό χρησιμοποιούνται όλα τα προαναφερόμενα χαρακτηριστικά του προηγούμενου κεφαλαίου με σκοπό να προκύψουν χάρτες ομοιότητας καθώς και η στρατηγική συγκέντρωσης στάθμισης για να συνδυαστούν οι χάρτες που λαμβάνονται και παράγουν την τελική βαθμολογία ποιότητας

Στο μέλλον, εστιάζουμε στην επινόηση μιας πιο αποτελεσματικής μεθόδου βασισμένης στα χαρακτηριστικά χρωματισμού προσπαθώντας να συμπεριλάβουμε παραλλαγές τύπου περιεχομένου, οπότε και να γενικευτούν αποτελεσματικά σε διάφορες μορφές εικόνων.

Εν τέλει και σε συνδυασμό με το προηγούμενο κεφάλαιο, αξίζει να διερευνηθεί γιατί τα παραδοσιακά μοντέλα NIQA δεν μπορούν να εφαρμοστούν αποτελεσματικά στους SCI και πώς να δημιουργηθούν ενοποιημένα μοντέλα IQA για το χειρισμό τόσο των φυσικών εικόνων όσο και των Screen Content εικόνων, τα οποία μπορούν να ονομαστούν μοντέλα IQA τύπου cross-content.



## ΒΙΒΛΙΟΓΡΑΦΙΑ

---

- [1] T. D. Nguyen, Q. Tran Thanh, T. M. Duc, T. Nguyen Quynh and T. M. Hoang, "SVM classifier based face detection system using BDIP and BVLC moments", The 2011 International Conference on Advanced Technologies for Communications (ATC 2011), Da Nang, pp. 264-267, September 2011.
- [2] I. Bakurov , M. Buzzelli , R. Schettini , M. Castelli , and L. Vanneschi, "Full-Reference Image Quality Expression via Genetic Programming", IEEE Transactions on Image Processing, vol. 32, pp. 1458-1473, February 2023.
- [3] K. Gu , Shiqi Wang , Huan Yang , Weisi Lin , Guangtao Zhai , Xiaokang Yang , Wenjun Zhang , "Saliency-Guided Quality Assessment of Screen Content Images," IEEE Transactions on Multimedia, vol. 18, no. 6, pp. 1098-1110, June 2016.
- [4] L. S. Chow, R. Paramesran, "Biomedical Signal Processing and Control", Elsevier, vol 27, Pages 145-154, May 2016.
- [5] Bo Yan , Bahetiyaer Bare , Weimin Tan , "Naturalness-Aware Deep No-Reference Image Quality Assessment", IEEE Transactions on Multimedia, vol. 21, no. 10, pp. 2603-2615, October 2019.
- [6] Z. Wang and A. C. Bovik, "Mean Squared error: Love It or leave It? A new look at signal fidelity measures", IEEE Signal Process. Mag., vol 26, no.1, pp. 98-117, January 2009.
- [7] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600-612, April 2004.
- [8] Z. Wang, E. P. Simoncelli and A. C. Bovik, "Multi-scale structural similarity for image quality assessment", Invited Paper, IEEE Asilomar Conference on Signals, Systems and Computers, vol. 2, pp. 1398-1402, November 2003.

- [9] L. Zhang, L. Zhang, X. Mou and D. Zhang, "FSIM: A Feature Similarity Index for Image Quality Assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378-2386, August 2011.
- [10] F. Gao , Y. Wang , P. Li , M. Tan , J. Yu , Y. Zhu , " DeepSim: Deep similarity for image quality assessment", Elsevier, vol. 257, pp. 104-114, September 2017.
- [11] S. Bosse, D. Maniry, K. -R. Müller, T. Wiegand and W. Samek, "Deep Neural Networks for No-Reference and Full-Reference Image Quality Assessment", *IEEE Transactions on Image Processing*, vol. 27, no. 1, pp. 206-219, January 2018.
- [12] T. Song , L. Li , J. Wu , Y. Yang , Y. Li , Y. Guo, G. Shi , "Knowledge-Guided Blind Image Quality Assessment With Few Training Samples", *IEEE Transactions on Multimedia*, vol. 25, pp. 8145-8156, December 2023.
- [13] S. Athar and Z. Wang, "A Comprehensive Performance Evaluation of Image Quality Assessment Algorithms", *IEEE Access*, vol. 7, pp. 140030-140070, September 2019.
- [14] H. R. Sheikh, Z. Wang, L. Cormack and A. C. Bovik, "LIVE Image Quality Assessment Database Release 2", 2005, Online: <http://live.ece.utexas.edu/research/quality>.
- [15].D. Ghadiyaram and A.C. Bovik, "LIVE In the Wild Image Quality Challenge Database", 2015, Online: <http://live.ece.utexas.edu/research/ChallengeDB/index.html>.
- [16] <http://database.mmsp-kn.de/kadid-10k-database.html>.
- [17] H. Lin, V. Hosu and D. Saupe, "KADID-10k: A Large-scale Artificially Distorted IQA Database," 2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX), Berlin, Germany, pp. 1-3, June 2019.
- [18] N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, C.-C. Jay Kuo, "Image database TID2013: Peculiarities, results and perspectives", Elsevier, vol. 30, pp. 57-77, January 2015.

- [19] H. Z. Nafchi, A. Shahkolaei, R. Hedjam and M. Cheriet, "Mean Deviation Similarity Index: Efficient and Reliable Full-Reference Image Quality Evaluator", *IEEE Access*, vol. 4, pp. 5579-5590, August 2016.
- [20] X. Zhang, X. Feng, W. Wang, W. Xue, "Edge Strength Similarity for Image Quality Assessment, *IEEE Signal Processing Letters*", vol. 20, no. 4, pp. 319–322, April 2013.
- [21] D. Varga, "Saliency-Guided Local Full-Reference Image Quality Assessment", *MDPI*, no. 28 vol. 3, pp. 483-496, July 2022.
- [22] C. Shi, Y. Lin, "Full Reference Image Quality Assessment Based on Visual Saliency With Color Appearance and Gradient Similarity", *IEEE Access*, vol. 8, pp. 97310-97320, May 2020.
- [23] L. Zhang, Y. Shen, H. Li, "VSI: A Visual Saliency-Induced Index for Perceptual Image Quality Assessment", *IEEE Transactions on Image Processing*, vol. 23, no. 10, pp. 4270-4281, August 2014.
- [24] W. Xue, L. Zhang, X. Mou and A. C. Bovik, "Gradient Magnitude Similarity Deviation: A Highly Efficient Perceptual Image Quality Index," *IEEE Transactions on Image Processing*, vol. 23, no. 2, pp. 684-695, February 2014.
- [25] L. Sevgi, "Fourier Transform and Fourier Series," *Electromagnetic Modeling and Simulation*, IEEE, pp.71-94, March 2014.
- [26] T. Guo, T. Zhang, E. Lim, M. López-Benítez, F. Ma and L. Yu, "A Review of Wavelet Analysis and Its Applications: Challenges and Opportunities," *IEEE Access*, vol. 10, pp. 58869-58903, June 2022.
- [27] V. Sygouni, C. D. Tsakiroglou, A. C. Payatakes, "Capillary pressure spectrometry: Toward a new method for the measurement of the fractional wettability of porous media", *Physics of Fluids*, vol 18, May 2006.
- [28] Laurent Itti , "Visual salience", (2007), *Scholarpedia*, 2(9):3327
- [29] <https://medium.com/@reshusingh/2d-dwt-a-brief-intro-89e9ef1698e3>

- [30] S. Gornale, A. Patil, C Veersheety," Fingerprint based Gender Identification using Discrete Wavelet Transform and Gabor Filters", International Journal of Computer Applications, vol. 152, no. 4, October 2016.
- [31] Y. D. Chun, S. Y. Seo, and N. C. Kim," Image Retrieval Using BDIP and BVLC Moments" IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, no. 9, September 2003.
- [32] R. S. Berns, "Extending CIELAB: Vividness,  $V_{ab}^*$ , depth,  $D_{ab}^*$ , and clarity,  $T_{ab}^*$ ", Color Research & Application, vol. 39, no. 4, pp. 322–330, August 2014.
- [33] Y. Fang, J. Yan, J. Liu, S. Wang, Q. Li and Z. Guo, "Objective Quality Assessment of Screen Content Images by Uncertainty Weighting," IEEE Transactions on Image Processing, vol. 26, no. 4, pp. 2016-2027, April 2017.
- [34] S. Wang, K. Gu, X. Zhang, W. Lin, S. Ma and W. Gao, "Reduced-Reference Quality Assessment of Screen Content Images," IEEE Transactions on Circuits and Systems for Video Technology, vol. 28, no. 1, pp. 1-14, January 2018.

## ΣΥΝΤΟΜΟ ΒΙΟΓΡΑΦΙΚΟ

---

Ονομάζομαι Ελένη Παρασκευά και είμαι μεταπτυχιακή φοιτήτρια του Τμήματος Μηχανικών Ηλεκτρονικών Υπολογιστών και Πληροφορικής, στην Πολυτεχνική Σχολή του Πανεπιστημίου Ιωαννίνων. Αποφοίτησα από το Τμήμα Μαθηματικών του Πανεπιστημίου Ιωαννίνων με ειδίκευση στις Πιθανότητες, τη Στατιστική και την Επιχειρησιακή Έρευνα. Τα ενδιαφέροντά μου αφορούν κυρίως την ανάλυση δεδομένων.