

Ben Gurion University of the Negev

Department of Electrical and Computer Engineering

Multimedia Compression SEM 1 – Final Project

Image Compression – From DCT to Wavelets

Tal Shkuri: 316411974

Itamar Guerchon: 209703941

Guy Mayost: 316414697

מוטיבציה

עם התקדמות הטכנולוגיה, איכות המולטימדיה הולכת ומשתפרת - תמונות וסרטוני וידאו מצולמים באיכותיות גבוהות יותר, הרזולוציות מתחדדות וכתוצאה מכך כמות המידע בכל תמונה או סרטון וידאו שכזה גדלה גם כן. הדבר מעלה קושי לאחסן ולשדר את נתוני המולטימדיה ברוחב פס בגודל נתון, ועם כמות אחסון מוגבלת שלא גדלו באופן פרופורציונלית לכמות המידע. לפיכך, נרצה למצוא דרך אפקטיבית יותר לדחוס נתוני מולטימדיה.

רקע תאורטי

כפי שלמדנו בכיתה, טרנספורמציה ה-DCT (Discrete Cosine Transform) שימשה כבסיס לדחיסת JPEG מאז 1992. התמרה זו מאופיינת בקירוב נתונים שכן היא ממירה ממרחב המיקום למרחב התדר באמצעות סדרת קוסינוסים. התמונה מחולקת ל- 8×8 בלוקים עליהם מתבצעת התמרת ה-DCT. מאפיין עיקרי נוסף הוא שרוב האנרגיה של האות מרוכזת בתדרים הנמוכים, מה שמאפשר לדחוס נתונים ע"י התעלמות מהתדרים הגבוהים. את התמרת ה-DCT ניתן ליישם על ידי אלגוריתמים עם סיבוכיות זמן נמוכה (FFT) מה שהופך אותה ליעילה בשימוש בזמן אמת.

Subband Coding היא שיטה לדחיסת אותות שבה האות מחולק לתחומי תדר שונים הנקראים subbands. בתדרים הנמוכים מוכל רב המידע לעומת התדרים הגבוהים שמכילים פרטים עדינים יותר הנקראים details. ה-DCT הנמוכים מוכל רב המידע לעומת התדרים הגבוהים שמכילים פרטים עדינים יותר הנקראים details. ה-DWT (Wavelet Transform) הוא סוג של Subband Coding המבצע את החלוקה לתחומי התדרים השונים. ה-DWT מתאפיין בתכונת הרב רזולוציה – הוא מפרק את האות לרמות תדר שונות ומאפשר שליטה על הרזולוציה. תכונה נוספת של התמרה זו היא הפירוק ההיררכי שהיא מבצעת. ה-DWT מחלק את האות ל-Subbands ברמות שונות כאשר התדרים הנמוכים נקראים approximation והתדרים הגבוהים הם ה-details. כמו כן, ההתמרה מותאמת למערכת הראייה האנושית משום שהאחרונה מסוגלת להבחין בטווח תדרים סופי. היכולת של ה-DWT לפרק ולברור רמות תדרים מסוימות מיעלת את הדחיסה ברמות דחיסה אגרסיביות כאשר תבחר להישאר עם התדרים הרלוונטיים בלבד.

הצגת הבעיה

המאמר מפרט על כך ששיטות מבוססות DCT מוגבלות ביכולתן להתמודד עם מספר בעיות. בעיה ראשונה היא שכאשר דוחסים בקצבי שידור ביטים נמוכים, כלומר כאשר אנו נאלצים לדחוס באופן אגרסיבי יותר - ככל הנראה בשל חוסר זמינות ברוחב הפס, אנו נתקלים בתופעת ה-Block-based artifacts. אלה נגרמים בגלל החלוקה לאזורים קטנים (בלוקים בגודל 8×8 פיקסלים) ושיטת הקוונטיזציה שמבוצעת על כל בלוק בנפרד. הסיבה העיקרית לכך היא ששיטת ה-DCT אוספת מידע על תדרים עבור כל בלוק בנפרד ולא על התמונה כולה, מה שעלול להוביל לאובדן קשרים בין הפיקסלים בקצוות של הבלוקים. בעיה שנייה היא העמידות המוגבלת שיש לאותם מקודדים מבוססי DCT. עמידות מתייחסת ליכולת של מערכת דחיסת תמונות לתפקד באופן תקין ולשמור על איכות תמונה גבוהה גם בתנאים קשים כגון שגיאות בשידור, רעש נתונים או מחיקות חלקיות של מידע. משתמע מכך, שמערכות דחיסה עמידות יודעות להתמודד עם אובדן נתונים ויודעות לשחזר ולקבל תמונה איכותית גם אם חלק מהמקטעים נפגעים. בעיה נוספת היא שלאותם מקודדים מבוססי DCT אין תמיכה ניהולית טובה של דרגות דחיסה שונות וקבלת רזולוציות שונות של התמונה בקלות. במילים אחרות, הם בעלי סקלאביליות מוגבלת.

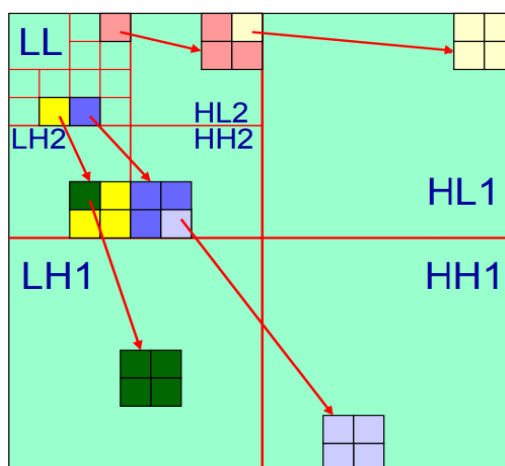
הפתרון

הפתרון שהמאמר מציג הוא שימוש בשיטות דחיסה מבוססות DWT. טרנספורמצית ה-wavelet הבדידה, לעומת ה-DCT, מספקת דחיסה איכותית יותר גם כאשר משדרים בקצבי ביטים נמוכים ונאלצים לדחוס באופן אגרסיבי. הדבר בא לידי ביטוי בכך שאותן שיטות מבוססות DWT נמנעות מתופעת ה-Block-based artifacts ובנוסף תומכות בשידור פרוגרסיבי – המידע הבסיסי נשלח תחילה, ומידע נוסף לשיפור איכות התמונה נשלח בהמשך במידה ואנחנו רואים שהתפנה מקום ברוחב פס השידור. לדוגמה, שידור תמונה ברזולוציה נמוכה שמושלמת עם הזמן עד לכדי רזולוציה גבוהה יותר.

המאמר מציג מספר טכניקות מתקדמות שמנצלות את התכונות של קידוד באמצעות DWT.

Embedded Zerotree Wavelet (EZW)

השיטה ראשונה היא שיטת ה- Embedded Zerotree Wavelet (EZW) שכאמור פותחה על ידי שאפירו בשנת 1993. תחילה, מקדמי ה-DWT מפרוקים באמצעות מבנה פס-אוקטבה (Octave-Band) שמחלק את המקדמים ל-subbands (תתי-פסים) ברמות תדר שונות. פירוק זה מאפשר ללכוד גם פריטים כלליים וגם פריטים עדינים בתמונה.

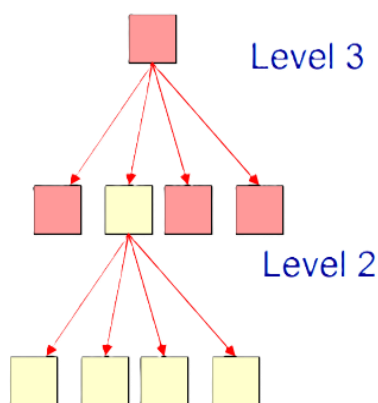


Octave-Band Wavelet Decomposition

בפירוק Octave-Band Wavelet Decomposition כפי שניתן לראות באיור שלעיל, מתקבל שכל level בנוי מה-approximation coefficients היכן שהמידע הכללי, כדוגמת קווי המתאר של עצמים בתמונה שמור. מידע זה שמור בתדרים הנמוכים – LL (Low Low). מסביבם נמצא ה-details בתדרים הגבוהים, היכן ששמורים הפרטים העדינים יותר כמו הקצוות והטקסטורה של התמונה. אלה הם ה- HL, HH, LH.

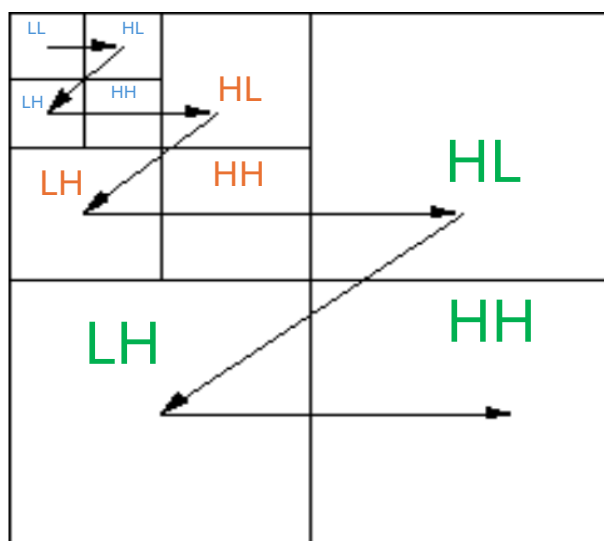
התחכום כאן הוא שתהליך הפירוק עצמו אינו דורש ידע מקדים על סוג התמונה או מאפייניה, אלא הוא מבוסס על התכונות המובנות של ה-DWT שמייצרות הפרדה טבעית בין מבנה כללי של התמונה לבין פרטיה. המקדמים החשובים ביותר ממוקמים בתדרים הנמוכים ונשמרים בעדיפות גבוהה, בעוד שהמקדמים בתדרים הגבוהים מוזנחים אם אינם משמעותיים. ניצול תכונת הרב-רזולוציה של ה-DWT שקיים מלכתחילה, הוא זה שמאפשר לייצג את התמונה ברמות פירוט שונות באופן שכזה.

עם זאת, מבנה הפס-אוקטבה מצריך שיטה חכמה יותר לקידוד מקדמי ה-wavelet על מנת להשיג תוצאות דחיסה טובות יותר. בשנת 1992 לואיס ונואלס הציגו לראשונה מבנה נתונים דמוי-עץ לייצוג המקדמים של פירוק פס-אוקטבה. תכונת ההיררכיה המובנת של ה-DWT מארגנת את המקדמים כך שאלו שנמצאים בתתי הפסים ברזולוציה הנמוכה פועלים כ"הורים" למקדמים שבתתי הפסים ברזולוציה הגבוהה יותר, ובכך מאפשרים ייצוג קומפקטי.



Quad-trees

ברגע שהערך של אחד הצאצאים של תת עץ מסוים הוא מתחת לסף threshold מסוים T שנקבע מראש, סבירות גבוהה שגם הילדים של אותו צאצא יהיו מתחת לאותו T. במצב זה, נוכל להזין ערך End-of-block ולא נצטרך להמשיך לרמות הנמוכות יותר בתת העץ שמתאימות לצאצא ממנו יצאתי. כתוצאה מכך אנו מקבלים חיסכון במידע. המחשה ויזואלית לכך ניתן לראות באיור שמוצג לעיל.



אופן השחזור מתחיל מה- approximation coefficients (כלומר מה-LL) של ה- level הגבוה ביותר, וממשיך לפענח ולהוסיף מידע, את ה- details, על פי הצורך לפירוט נוסף, בסדר הבא: $LL \rightarrow HL \rightarrow LH \rightarrow HH$. במידה וסיים לפענח את כל ה- details של אותו ה- level אז ימשיך ל- LL של ה- level הבא באותו האופן הנ"ל. המחשה לכך ניתן לראות באיור שבסוף העמוד הקודם: השחזור מתחיל מה-LL של level 3 ועובר ל- details שלו. כאשר סיים, הוא ממשיך ל- level 2 בצורה דומה כך שכל level 3 מהווה בעצם את ה-LL של level 2, וכך ממשיך הלאה ל- level 1 עד שמסיים את תהליך השחזור.

היתרון באופי תהליך השחזור הנ"ל הוא שאנו יכולים לעצור אותו בכל עת כרצוננו, על פי זמינות רוחב הפס. כך אנו מאפשרים בעצם את תכונת השידור הפרוגרסיבי שהסברנו קודם לכן.

Set Partitioning in Hierarchical Trees (SPIHT)

השיטה השנייה שפותחה בעקבות ה- EZW היא שיטת ה- Set Partitioning in Hierarchical Trees (SPIHT). שיטת ה- SPIHT לוקחת את עקרונות ההיררכיה והיעילות של ה- EZW ומשפרת אותם באמצעות מיון וחלוקה חכמים יותר של קבוצות המקדמים. החלוקה מתבצעת בשני שלבים: זיהוי מקדמים משמעותיים ולאחר מכן חלוקתם לקבוצות. בשלב הראשון, מקדמים בעלי ערך גדול מסף threshold מסוים מסומנים כמשמעותיים. בשלב השני, הקבוצות ההיררכיות (trees) מחולקות שוב ושוב, כאשר כל שלב מתמקד במקדמים החשובים ביותר.

לשיטת ה- SPIHT קיימים שיפורים ניכרים על פני שיטת ה- EZW. הראשון מבניהם הוא סדר המיון הדינמי. בעוד שה- EZW מסתמך על "עצי אפסים סטטיים" (Zerotrees), ה- SPIHT ממין מקדמים לפי חשיבותם באופן דינמי. כמו כן, ה- SPIHT יעיל יותר בשל חלוקת הקבוצות שנעשית באלגוריתם שזו מצמצמת את הצורך במידע נוסף לקידוד ומפחיתה את כמות הביטים הנדרשת. יתרה מזאת, שיטת ה- SPIHT משפרת את יכולת השידור הפרוגרסיבי בכך שמאפשרת שידור מדורג שבו התמונה משתפרת ככל שמתקבלים יותר נתונים.

Scalable Image Compression with EBCOT

השיטה השלישית שפותחה בעקבות ה- EZW ושיטת ה- SPIHT הוא אלגוריתם ה- EBCOT (Embedded Block Coding with Optimized Truncation).

ה- EBCOT משתמש בטרנספורמציה ה- Wavelet הבדידה לפירוק התמונה, ולאחר מכן מבצע קוונטיזציה וקידוד של מקדמי ה- Sub-bands. האלגוריתם פועל על ידי חלוקת כל Sub-band לבלוקים קטנים (Code-Blocks), כאשר כל בלוק מקודד בנפרד, מה שמאפשר גמישות גבוהה בדחיסה ושיפור הסקלאביליות. נזכיר כי סקלאביליות (מדרגיות) מתייחסת ליכולת של מערכת לתמוך במגוון רזולוציות או קצבי ביטים, כך שהתמונה יכולה להישמר וניתן לשחזר אותה באיכויות שונות ללא צורך בדחיסה חוזרת. המשמעות היא שניתן להפיק מהתמונה גרסאות שונות בהתאם לצורך, למשל לשימושים בעלי מגבלות רוחב פס או דרישות רזולוציה שונות.

אחד היתרונות של ה-EBCOT הוא הסקלאביליות מלאה. כלומר האפשרות לדחוס ולשדר תמונה ברזולוציות שונות וביחסי דחיסה מגוונים ללא צורך בדחיסה מחדש. זה מאפשר שליטה גמישה יותר על איכות ורזולוציה. יתר על כן קיימת הגישה אקראית. כלומר, ניתן לגשת לבלוקים בודדים באופן עצמאי, מה שמקל על עיבוד ותצוגה של חלקי התמונה. שליטה מדויקת יותר בשחזור נתונים מתאפשר בשל החלוקה של כל Sub-band לבלוקים קטנים. ושיפור השידור הפרוגרסיבי. בזכות הגישה האקראית ניתן לשדר רק את הזרמים הרלוונטיים, מה שמוסיף דיוק לבחירת הנתונים ומייעל את קצב השידור בהתאם לדרישות הרזולוציה והאיכות.

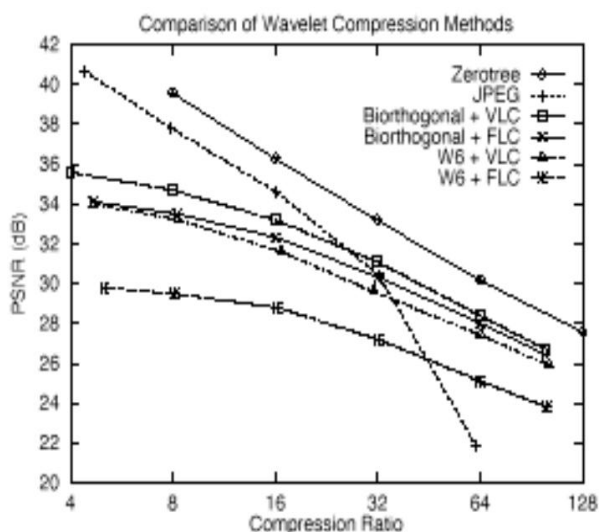
השיפורים העיקריים ביחס לשיטות הקודמות הם צמצום בזבז הביטים ושחזור תמונה איכותי יותר ברמות שונות. צמצום בזבז הביטים מתאפשר בזכות חלוקה חכמה של הזרם הדחוס ויכולת התאמה מדויקת של הדחיסה לכל בלוק. שחזור תמונה איכותי יותר ברמות שונות מתאפשר בזכות החלוקה לבלוקים עצמאיים ושימוש בטרנספורם ה-Wavelet הבדיד. זה מאפשר קוונטיזציה מדויקת ושיפור יחס אות לרעש (SNR).

Lossless Image Compression using Integer to Integer WT

השיטה הרביעית שפותחה היא דחיסה ללא איבוד נתונים (Lossless Compression) חיונית לתמונות רפואיות, נתוני סייסמוגרפיה (מדידות של גלי סיסמיים הנוצרים בפעילות גיאולוגית כמו ברעידות אדמה), תמונות לוויין ותכני וידאו באיכות גבוהה. ה-JPEG מציע קידוד Lossless על בסיס קידוד הפרשי פיקסלים, אך עקב גודל התמונות הגדול, יש צורך בשיטות המאפשרות אחסון ושידור פרוגרסיבי יעיל.

למרות שהשימוש ב-Wavelet מתאים לדחיסה פרוגרסיבית, אחת הבעיות המרכזיות היא שה-Output שעבר סינון מתקבל בערכים עם נקודה צפה (Floating Point). זה מקשה על שחזור Lossless של הנתונים. הפתרון לכך הוא טרנספורם Wavelet הפיך מסוג Integer-to-Integer, שממיר נתונים שלמים לערכים שלמים תוך שמירה על היכולת לפענח את התמונה באופן מדויק לחלוטין.

שיטה זו מתבססת על שלבי הרמה (Lifting Steps) המאפשרים בנייה גמישה של טרנספורמים שלמים לכל תמונה. מספר טרנספורמים כאלה יושמו ונבדקו, והמסקנה היא שהם מאפשרים דחיסת תמונות ללא איבוד נתונים תוך תמיכה בשידור פרוגרסיבי. כלומר, שידור הדרגתי שבו ניתן להעביר גרסה ברזולוציה נמוכה של התמונה תחילה ולאחר מכן לשדר פרטים נוספים לשיפור האיכות.



Comparison of Wavelet Compression

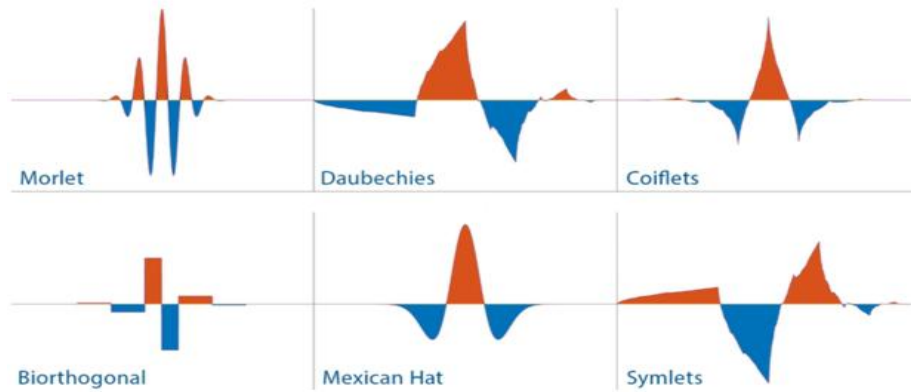
יחסי דחיסה :

ביחס דחיסה נמוך של עד 1:25 ניתן לראות כי דחיסת JPEG שמבוססת על DCT מציגה ביצועים טובים יותר מספרית בהשוואה לשיטות דחיסה מבוססות DWT (פרט ל-Zerotree).
ביחס דחיסה גבוה של מעל 1:30 ניתן לראות כי הביצועים של ה-JPEG מתדרדרים במהירות בשל מגבלות הבלוקים הקבועים של DCT, מה שיוצר את בעיית ה-Block-based artifacts. לעומת זאת, קידודים המבוססים על ה-Wavelet הבודד מדורדרים באופן הדרגתי ומתון יותר גם ביחסי דחיסה של 1:100, תוך שמירה על איכות תמונה טובה יותר. ה-Zerotree ודחיסה מבוססת Wavelet מתאימים יותר לדחיסה גבוהה בזכות עמידותם ותכונת השחזור האיכותית שלהם גם ברמות דחיסה גבוהות מאוד.

תלות במקדמים :

מקדמי ה-DCT קבועים ותדריו מיוצגים במטריצה של 8×8 לעומת ה-DWT שהוא תלוי בסוג המסנן. מכיוון שאין מסנן יחיד שמתאים לכל התמונות באופן אופטימלי, אז קידוד תמונות מבוסס DWT דורש התאמה של מסנן ה-Wavelet בהתאם למאפיינים הסטטיסטיים של התמונה. בנוסף ניסויים הראו שביצועי הדחיסה (Lossy ו-Lossless) תלויים בסוג התמונה. כלומר, מסננים מסוימים עובדים טוב יותר על תמונות ספציפיות, אך אין מסנן אחד שמתפקד בצורה מיטבית על כל סוגי התמונות.
(עופר גם בהרצאה שלנו הוסיף הערת ידע כללי ואמר שבאופן כללי המסנן של Daubechies עובד הכי טוב כיום עבור תמונות רנדומליות אך זה לא כחלק מהחומר של המאמר, סתם הערת ידע כללי).
לכן, בחירה אדפטיבית של מסנן ה-Wavelet חיונית. בנוסף, גם שימוש בקוונטיזציה ומודלים סטטיסטיים מתקדמים לא יפצה על מסנן לא מתאים.

תמונה של דוגמאות של משפחות אם של Wavelet ובניהם משפחת ה-Wavelet של Daubechies עליה עופר נתן את הערת הידע הכללי.



Mother wavelet functions

השורה התחתונה של ההשוואה והמאמר הוא :

DWT אינו מהווה תחליף ל-DCT אלא משמש לשיפורו, במיוחד כאשר יש צורך להפחית ארטיפקטים ולהתמודד עם דחיסות גבוהות יותר.

ביקורת על המאמר

המאמר מספק סקירה מצוינת על המעבר מ-DCT ל-DWT, אבל סובל ממיקוד חד-צדדי ומחוסר ניתוח מעשי מספק. הוא מועיל להבנת השיטות החדשות, אך חסר נתונים אמפיריים מעמיקים, ניתוח חישובי, והשוואה לטכניקות דחיסה אחרות. היה ניתן לשפר אותו למשל עם היו מוספים השוואות מקיפות יותר, ניתוח של חסרונות ה-DWT, ודיון על היישום המעשי של השיטות השונות.

נחלק ליתרונות וחסרונות:

• יתרונות:

- סקירה מקיפה ומעמיקה – יש תיאור מפורט על האבולוציה של שיטות הדחיסה, כולל היתרונות והחסרונות של טכניקות שונות, תוך שילוב רקע מתמטי ואנליזה של הביצועים.
- פרקטיות ויישומים מגוונים – המאמר מדגיש כיצד שיטות כמו SPIHT ו-EBCOT מתאימות ליישומים קריטיים, כולל שידור פרוגרסיבי, דחיסה סקלאבילית, דחיסת תמונות רפואיות ולווייניות, ומסביר כיצד תכונות אלו מספקות גמישות רבה יותר לעומת JPEG המסורתי.

• חסרונות:

- מיעוט דוגמאות מעשיות - חסר נתונים אמפיריים רחבים או הדגמות מעשיות של ביצועי השיטות השונות על תמונות שונות. השוואות ה-PSNR מוצגות באופן תיאורטי בלבד, אך לא מספקות ניתוח מפורט של התוצאות בפועל.
- מורכבות חישובית לא מוסברת - אין התייחסות מספקת לדרישות החישוביות בשיטות כמו SPIHT ו-EBCOT. אין ניתוח לגבי צריכת הזיכרון, זמן הריצה או ההשפעה של פרמטרים שונים על הביצועים, מה שעלול להקשות על הבנת היישום המעשי של השיטות.

- היעדר ביקורת על DWT - המאמר משבח את יתרונות ה-DWT אך אינו מציג את האתגרים הטמונים בו, כמו הצורך בבחירת פילטרים מתאימים והשפעתם על איכות השחזור. כמו כן, לא מוזכרת בעיית המורכבות של התמרת ה-Wavelet הבדיד לעומת DCT, מה שמונע נקודת מבט מאוזנת יותר.
- מיקוד חד-צדדי – המאמר מתמקד כמעט לחלוטין בטכניקות מבוססות Wavelet, ואינו משווה אותן לטכניקות דחיסה מודרניות אחרות, כגון למידת מכונה או אלגוריתמים היברידיים המשלבים DWT ו-DCT. כתוצאה מכך, הקורא אינו מקבל תמונה רחבה של התחום כולו.

המלצות להמשך העבודה

- (1) שיפור טכניקות ה-DWT – ניתן לשפר את בחירת מסנני ה-Wavelet ואת ההתאמה שלהם לסוגי תמונות שונים (לוויין, רפואיות, טבע וכו'), להפחית artifacts ולשפר את ה-PSNR. ניתן לעשות זאת על ידי שימוש בלמידה מכונה לבחירת מסננים דינמיים שמתאימים לכל תמונה באופן אדפטיבי וניתן לבצע אופטימיזציה של המסננים באמצעות מאגרי תמונות גדולים ולבחון השפעת פילטרים שונים על איכות התמונה לאחר דחיסה ושחזור.
- (2) אינטגרציה עם למידת מכונה - ניתן לשפר את יעילות הדחיסה וההתאמה האוטומטית בהתאם למאפייני התמונה. וניתן לשפר את השיטות כמו ה-SPIHT ו-EBCOT באמצעות אלגוריתמים חכמים יותר. ניתן לעשות זאת על ידי אימון רשתות נוירונים לניתוח מאפייני התמונה ולבחור את פרמטרי הדחיסה האופטימליים. בנוסף, לשלב למידת מכונה עם השיטות כמו ה-SPIHT ו-EBCOT כדי לייעל את הדחיסה בזמן אמת ולשפר את השחזור תוך שמירה על איכות גבוהה.
- (3) העברת מידע ועמידות לשגיאות בסביבות רועשות - ניתן לשפר את עמידות הדחיסה לשגיאות בתנאי תקשורת לא יציבים, למשל, העברת תמונות בלויינים, טלפונים סלולריים ורשתות אלחוטיות על ידי שילוב דחיסת Wavelet עם פרוטוקולי תיקון שגיאות כדי לשמור על איכות התמונה גם בתנאים של איבוד נתונים או רעשים חיצוניים. וניתן גם לפתח אלגוריתמים מותאמים לשידור תמונות בסביבות רועשות.
- (4) דחיסה רב-מודאלית - ניתן ליישם טכניקות דחיסה גמישות יותר שיכולות להתמודד עם סוגי מדיה שונים כמו טקסט, תמונות, וידאו ותמונות משולבות וניתן לשפר את האחסון והשידור של מידע מגוון בצורה אופטימלית. ניתן לעשות זאת על ידי פיתח מודלים המשלבים דחיסת Wavelet עם טכניקות כמו קידוד הופמן (Huffman Coding) או דחיסה פרדיקטיבית. בנוסף, ניתן ליישם שיטות דחיסה שמתאימות למידע רב-מודאלי, למשל עיבוד תמונה וטקסט יחד, מה שיכול לשמש ליישומים מתקדמים כמו זיהוי טקסט מתמונות.

*מצורף קוד הממחיש פירוק ל-subbands