

עיבוד תמונות בעזרת מחשב (236327)

הרצאה מס' 1

מבוא כללי

כללי

התחום "עיבוד אותות" החל להתפתח לפני יותר ממאה שנים כשהרעיון היה למדוד תופעות ולבצע עיבוד אנלוגי על מדידות אלה (לניקוי רעש, לניתוח, ועוד). במהלך ה-50 שנה האחרונות חל מהפך בדרך בה אנו מעבדים נתונים בשל התפתחות המחשב, ועם מהפך זה גם התחום "עיבוד אותות" השתנה לבלי הכר. די מהר הבינו כי אם האותות יידגמו ויומרו לקובצי מחשב (סדרות של מספרים), ניתן יהיה לעשות מגוון הרבה יותר עשיר של עיבודים.

אחד האותות בהם יש עניין רב הוא תמונות, כיוון שראיה הינה החוש החזק ביותר שלנו וזה המעסיק את מרבית חלקיו של המוח. "עיבוד תמונה" עוסק בטיפול ממוחשב בתמונות. שאלות כגון כיצד תמונה הופכת להיות קובץ מחשב, מה ניתן לעשות על קובץ זה, וכיצד מפיקים מהקובץ תמונה חזרה, הם הנושאים בהם נעסוק. זהו תחום שמשלב תיאוריה עמוקה, אלגוריתמים, ויישומים מעניינים. בקורס זה נציג את המבוא ל-"עיבוד תמונה" ונתחיל קודם בבניית השפה לתחום ע"י היכרות בסיסית עם "עיבוד אותות".

במסגרת פרק זה נערוך היכרות מעט יותר קרובה עם התחום הקרוי "עיבוד תמונה" וסביבתו, על מנת לתת רקע כללי לקורס, ומוטיבציה להמשך. המטרה היא בעיקר לתת את תמונת התחום ממעוף הציפור ללא כניסה לפרטים, אותם נמלא היטב במהלך הקורס. בשל כך, לפרק זה גישה "שיווקית" מעט. כך למשל, בניגוד לפרקים הבאים, לא נראה כאן משוואות ומסגרת מתמטית, אלא נתמקד במלל, תיאורים והדגמות.

מהי תמונה במחשב?

נניח כי אנו עומדים בחדר ובו חלון ודרכו נסתכל אל העולם. מכל נקודה הנראית מהחלון ייפלטו פוטונים (אור) וכמותם תעיד על בהירות אותה נקודה. נניח כי אנו יכולים לספור כמות פוטונים זו בכל נקודה וליחס לה מספר בין 0 ל-

1 כאשר 0 מתייחס לשחור מוחלט (אין פוטונים) ו-1 ללבן מוחלט – כמות פוטונים רבה שנותנת תחושה של בהירות מוחלטת.

בדרך זו הראינו כי תמונה היא פונקציה דו-ממדית $f(x,y)$ המוגדרת לכל נקודה (x,y) בחלון, וערכיה הם מספרים ממשיים בתחום $[0,1]$. תמונה כזו לא תוכל להיות מיוצגת כקובץ מחשב! יש אינסוף נקודות בעלי ערכים, וכל ערך הוא מספר ממשי. על מנת להמיר תמונה כזו לקובץ נזדקק לשתי פעולות בסיסיות:

- דגימה – במקום לומר את ערך הפונקציה לכל (x,y) , נתייחס רק לנקודות על השריג. בצורה זו קיבלנו מספר סופי של נקודות. ברור כי למרווח הדגימה של השריג השפעה מכרעת על איכות התמונה שתתקבל. מקובל להתייחס למושג ה-"רזולוציה" כמידת הדיוק של פרטים קטנים בתמונה, וברור כי הרזולוציה מושפעת מרמת צפיפות השריג.
- קוונטיזציה (כימות) – כל ערך הקרוי פיקסל (Pixel=Picture Element) ייוצג ע"י כמות סופית של סיביות. לדוגמה, אם ניתן לכל פיקסל 8 ביטים, יהיו 256 ערכי "רמות אפור" המתייחסים לגוונים שחור ועד לבן. מסתבר כי לעין אנושית די במספר הזה על מנת לקבל תחושה של אי-אובדן.



פעולת הדגימה בתמונה – חישוב ממוצע על-פני שטחי הגלאים

בכפוף לשתי פעולות אלו יתקבל הקובץ במחשב המתאר את התמונה. לכן, מבחינת המחשב תמונה אינה אלא אוסף מספרים המסודרים כמערך דו-מימדי של שורות ועמודות (מטריצה), כשכל מספר הוא מספר המיוצג בכמות סיביות קבועה וסופית.

כאשר נכנס צבע לתמונה, יש מגוון דרכים לייצגו. גישה מקובלת אחת היא להחזיק מטריצה כמקודם בה לכל פיקסל יש רמת בהירות (שחור-לבן), ערך שני לכל פיקסל שיגיד מה הגוון (על ציר הגוונים - אדום, צהוב, ירוק, כחול, סגול), ומספר שלישי בכל פיקסל שיקבע את עומק הצבע. אלטרנטיבה מקובלת היא ייצוג הפיקסל בשלושה צבעים שמתארים את כמות האדום הירוק והכחול שיש לערבב על מנת לקבל את צבעו ובהירותו של הפיקסל. נושא הצבע מורכב ונדון במסגרת מתקדמת יותר, ואנו נמעיט בדיון בו.

תמונות קיימות במגוון גדלים וסוגים. ניתן מספר דוגמאות על מנת להמחיש גדלים אופייניים של תמונות, כשנניח כי הייצוג כולל צבע ונעשה שימוש ב- 8 סיביות לכל ערך:

- אייקון במחשב – מקובל להחזיק Icons בגודל קטן של עזרות בודדות על עשרות בודדות של פיקסלים. לדוגמה, עבור 50 על 50 פיקסלים יידרש קובץ של 7500Bytes.
 - תמונת פספורט – תמונת פספורט איכותית תדרוש כ- 200 על 200 פיקסלים, המובילים לגודל קובץ של 120Kbytes.
 - תמונת טלוויזיה – בשידורי לוויין דיגיטאליים כיום גודל התמונה המוצגת הוא כ- 640 על 480 פיקסלים. לכן כל תמונה הינה בגודל של 900KB. אם נניח כי משודרות כ- 30 תמונות בשנייה לשם קבלת תחושת רצף בזמן, אז כל שנייה דורשת כ- 30Mbytes, וסרט באורך שעה כולל נפח של 108GBytes. בשנים הקרובות יעבור שוק שידורי הטלוויזיה מהפיכה ל-HDTV בה כמות הפיקסלים גדולה פי 4, ועימה הגדלים הנ"ל הופכים להיות בלתי נתפסים.
 - תמונה ממצלמה דיגיטאלית באיכות טובה – כיום ניתן לרכוש במחיר סביר מצלמה דיגיטאלית בה תמונה כוללת 2000 על 2000 פיקסלים – 12Mbytes לכל תמונה.
 - תמונת הדמאת לוויין SPOT – הלוויין הצרפתי SPOT מצלם את כדור הארץ מגובה של 800 ק"מ כשהוא מייצר תמונות בגודל של 6000 על 6000 פיקסלים. נפח תמונה אחת כזו הוא 100MB. ישנם לווייני צילום אחרים בהם התמונות אף גדולות יותר.
- לפני שנודעזע עמוקות מהגדלים הללו, חשוב להבהיר כי אמנם תמונות הינן בגודל המתואר, ודבר זה מעיד על הכובד שבעיבוד מידע מסוג זה. אך נפחים

אלו רלוונטיים רק במהלך רכישתן ועיבודן של התמונות. כאשר הן נאגרות בזיכרון או משודרות מפעילים תהליכי דחיסה שאותם נזכיר בהמשך.

כיצד נוצרת תמונה?

מצלמה דיגיטאלית מייצגת דרך מודרנית ויחסית ישירה לרכישת תמונה והעברתה למחשב. בליבה של מצלמה כזו קיים מערך גלאים דו-מימדי גדול שסופר בכל גלאי (כל גלאי יהפוך אחר-כך לפיקסל) פוטונים וממיר אותם לזרם אלקטרוניים – זרם חזק פירושו הרבה פוטונים שפירושו גוון בהיר, וחוסר בפוטונים נותן את הגוונים הכהים. על מנת לקבל תמונה באיכות טובה נדרשת אופטיקה לשם מיקוד התמונה על-פני מישור הגלאים. כמו כן נדרשת סידרה של פעולות עיבוד תמונה על מוצא הגלאים לשם ניקוי המערך מבעיות שונות של אי-תאימות הגלאים, גלאים חסרים, הכנסת צבע, תיקון לפגמים הנובעים מתאורה לקויה ועוד.

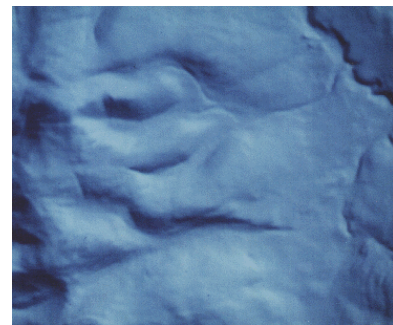
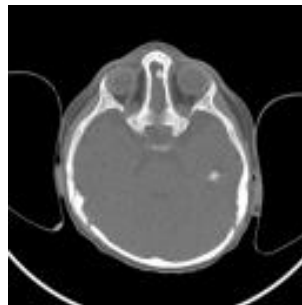
כאשר בידנו תמונה בעותק קשה (הדפסה), הדרך להבאתה למחשב היא דרך סורק (Scanner). סורק הינו בעצם מצלמה דיגיטאלית אך כזו הפועלת בתנאי תאורה מבוקרים, וללא תנועת האובייקט המצולם. לכן ניתן ליישם הצילום ע"י מערך חד-מימדי זול יותר שנע לאורכה של התמונה וקורא ערכים, שורה אחר שורה.

תמונה יכולה להיות מלאכותית, יציר מחשב הנבנה במגוון דרכים אפשריות:

- ע"י פעולת rendering בגרפיקה, המקבלת מודל תלת-מימדי של אובייקטים ומציגה כיצד הגוף יראה מזווית מסוימת עם תאורה מסוימת.
- ציורים וסימנים, כגון אייקונים במחשב או כל יצירה של גרפיקאי.
- ע"י מניפולציות על נתונים לשם הצגתם, כגון תיאור מפת גבהים של אזור גיאוגרפי מסוים כתמונה בה בהיר זה גבוה ושחור נמוך.
- ע"י עיבוד נתונים בעלי אופי לא תמונתי והמרתם לתמונה כמו בצילום CT, אולטרה-סאונד, MRI, מידע סיסמי, ועוד.

בבואנו להפוך תמונה לקובץ מחשב נצטרך להתלבט בסוגיות כגון מרווח הדגימה וקביעת הרזולוציה של התמונה, הדרך הנכונה לבצע קוונטיזציה, התייחסות לתנועה של האובייקט המצולם, תכנון הגלאים והאופטיקה לשם

מניעת תופעות קיפול, תיקון גוונים מקומי, ועוד. כל אלו מהווים פרק אחד רחב בעיבוד תמונה.



תמונות מסוגים שונים – תמונת אדם, גרפיקה, תמונת נוף, תמונת CT, ויזואליזציה של מידע עומק

מה נעשה עם תמונה במחשב?

אחד הנושאים החמים ביותר (ועם הצלחה מסחרית מוכחת!) הוא דחיסת תמונות. כבר ראינו כי הנפחים בתמונות מגיעים למימדים בלתי נתפסים. תמונת צבע בגודל 1000 על 1000 עם 8 סיביות לפיקסל דורשת 3MB. האם זה מחויב המציאות? התשובה למזלנו היא חד-משמעית לא! ניתן לדחוס תמונות כאלה פי 10 ועדיין לא לראות כמעט כל פגם. הרעיון הוא לייצר אלגוריתם דחיסה המקבל קובץ תמונה כפי שתיארנו ומפיק קובץ קטן יותר משמעותית המייצג את התמונה הדחוסה. כדי לראות את התמונה נחוץ אלגוריתם פריסה המשחזר את התמונה ממיעוט הביטים שנשמרו.

הרעיון הבסיסי בדחיסה הוא מיצוי יתירות הקיימת בין הפיקסלים השונים שנוטים להיות דומים זה לזה אם הם שכנים. מסתבר כי קשרים אלו משמעים כי המידע האמיתי הקיים בתמונה קטן לאין-ערוך.



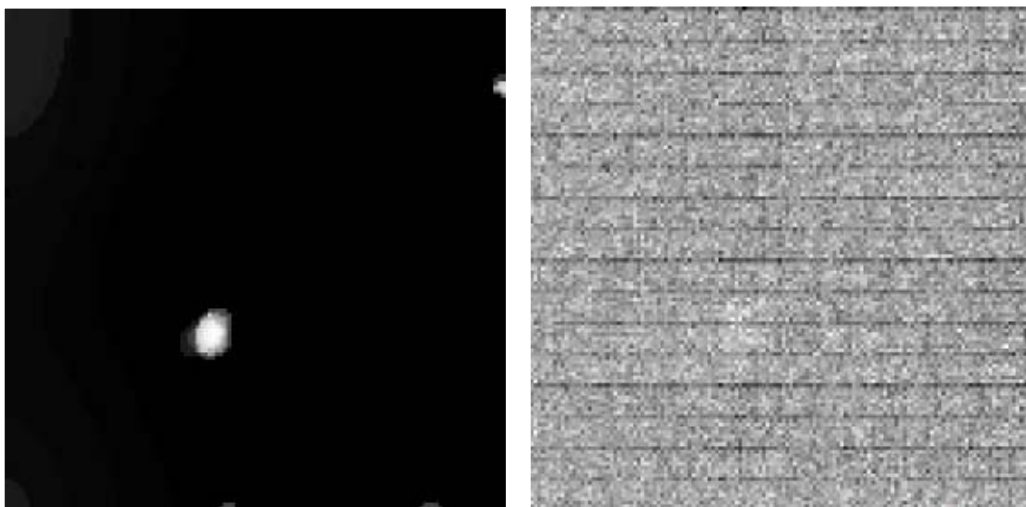
דחיסת תמונות – שימוש באלגוריתם JPEG בדחיסה ביחס 35:1

ישנם מצבים בהם דחיסה מאוד רצויה אך ללא התרת שגיאות. כך למשל הוא הדבר בתמונות רפואיות בהם אסור לגרוע מהאיכות מחשש לאבחנה לא נכונה. במקרים כאלו משתמשים באלגוריתמי דחיסה ללא אובדן המבטיחים כי התמונה לאחר דחיסה ופריסה זהה בדיוק למקור. ברור כי בכפוף לכך, יחסי הדחיסה האפשריים קטנים יותר ותלויים באופן חזק בתוכן התמונה.

דחיסה נחוצה שבעתיים בסרטי וידאו. בסרט וידאו יש יתירות בין פיקסלים גם בזמן וגם במקום ולכן ניתן להשיג דחיסה ביחס 100 ויותר. ואמנם, נושא דחיסת הווידיאו מזמן עבר מדיון אקדמי עקרוני למגוון מוצרים ולשוק המגלגל מיליארדים. שידורי טלוויזיה דיגיטאליים, מערכות DVD, צילום במצלמות, שירות וידאו-פון, תוכנות לתקשורת וידאו דרך האינטרנט, ועוד – כל אלה חייבים להשתמש בדחיסת וידאו, ואמנם עושים זאת. בתחום זה לא פלא כי ריבוי המוצרים הביא לתקנים וסטנדרטים, על מנת לאפשר למערכות שונות לדבר זו עם זו ברמת הקבצים הדחוסים.

נושא פופולארי שני בטיפול בתמונות במחשב הוא נושא שיפור איכות. במקרים רבים נתונה תמונה כשהיא פגומה, אם בשל תנאי הצילום (אובך, ערפל, גשם, תאורה לקויה, צילום דרך מים, ועוד), או בשל שימוש במערכת צילום לקויה. דוגמה קלאסית למצב כזה התרחשה לפני כ-10 שנים עם שיגור הטלסקופ הבל (Hubbel). בשל פגם בהצבת המראות המרכזות התקבלו תמונות מטושטשות ונחוץ היה לתקן. עד לשיגור אסטרונוטים לתיקון הטלסקופ עשו שימוש באלגוריתמי שיפור תמונות לשם הוצאת המרב מהמידע הפגום ששודר לארץ.

גלאים זולים המוצבים במצלמות זולות (webcam) יוצרים מגוון רחב של תופעות מפריעות המחייבות תיקון. גלאים במערך דו-מימדי אינם בהכרח אחידים בעקומת ההיגב שלהם לאור. פירוש הדבר שאור אחד על הגלאים צפוי לתת תמונה עם פיקסלים בערכי רמות אפור משתנים. תיקון תופעה זו, ותיקון תופעה קיצונית יותר של גלאי מקולקל הינם הכרח בטרם הצגת התמונה למשתמש. יצרני מטריצות הגלאים יכולים בצורה זו למכור (בזול יותר) גם מטריצות גלאים עם מספר גלאים פגומים בידיעה שערכי אלו יושלמו בתהליך מתקן באינטרפולציה.



תמונה מקולקלת מ-Gemini ותיקונה

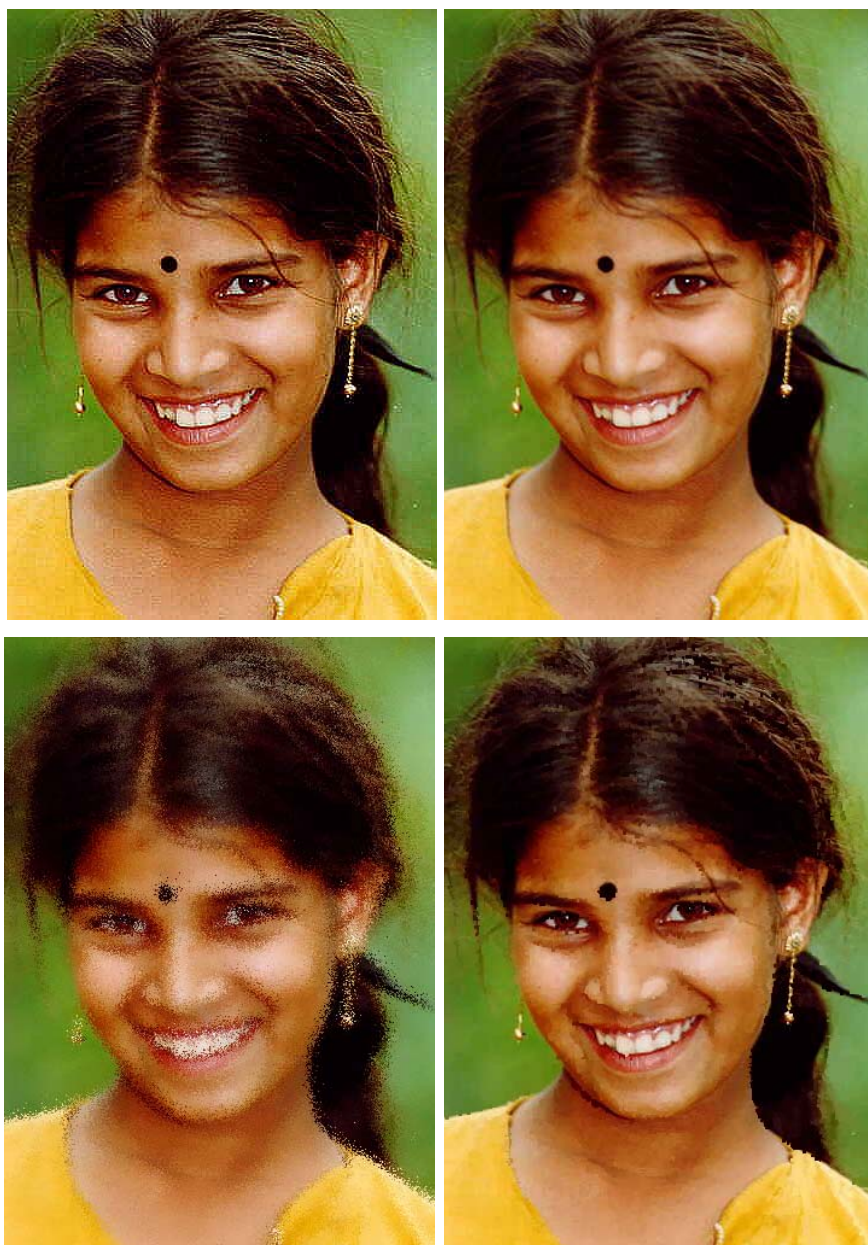
בהינתן תמונה ישנה ושרוטה, או בהינתן תמונה שעליה קושקש, נרצה לעיתים להחזיר התמונה למצב מקורי. תהליכים כאלו מקובלים עד לאחרונה להיפטר באמצעים ידניים ע"י גרפיקאים. במוזיאונים שחזורי תמונות מפגעי הזמן נעשים ע"י אומנים. כיום ישנם אלגוריתמים היודעים לתקן תמונות דיגיטליות עם פגמים כאלו עם הצלחה מפתיעה של מילוי חלקים חסרים ומחיקת שריטות. בתוכנות רבות בהן עושים שימוש לצפייה בתמונות (Photoshop, Irfanview, ואין-ספור אחרות) ישנן ספריות שלמות עם מגוון עיבודים לשם חידוד תמונה, וביצוע מגוון רחב של אפקטים. לדוגמה, נניח כי בידינו מגוון עשיר של תמונות של ואן-גוך, ויש לנו תמונה שאותה נרצה לעבד כך שתיראה כציור של ואן-גוך. אלגוריתמים לגימיקים מעין זה קיימים והצלחתם מפתיעה למדי.



מילוי חורים ותיקון פגמים אחרים בתמונה

שיפור תמונה נדרש לא פעם בשל ניצול לקוי של התחום הדינאמי. נניח כי מתוך חדר צילמנו כלפי חוץ ביום שמש באופן שגם חלקים מתוך החדר וגם הנוף בחוץ מופיעים. אם הפלאש הופעל פנים החדר יראה טוב והנוף בחלון יראה לבן בוהק ורווי ללא יכולת לראות פרטים. אם הפלאש לא יופעל הנוף יראה טוב אך פנים החדר יראה חשוך לחלוטין קיימים אלגוריתמים שמקבלים תמונה פגומה כזו ומתקנים אותה, יוצרים את האפקט שלכל פיקסל יש פלאש אישי שלו שמתאים לצרכים.

באופן דומה, כשנצלם סצנה בה יש אובייקטים קרובים ורחוקים, מנגנון המיקוד יבחר מרחק אחד במיקוד מיטבי כשכל מה שקרוב יותר או רחוק יותר מופיע כמטושטש. תהליכי תיקון יכולים להביא לתמונה חדה בכל המרחקים.



עיבודים שונים על תמונה ב-IrfanView – מקור (למעלה מימין), חידוד תמונה (למעלה משמאל), אפקט ציור שמן (למטה מימין), ופיזור (למטה משמאל).



תיקון אפקט תאורה לקויה

תחום עשיר בתיאוריה ותוצאות מעשיות הוא נושא שחזור תמונה. נניח כי נסענו לחו"ל, צילמנו תמונות רבות, ועם הגיענו חזרה גילינו כי מנגנון המיקוד של המצלמה היה פגום וכל התמונות מטושטשות. כיוון שתהליך הקלקול ידוע, ניתן לבצע שחזור לתמונות בטרם הקלקול, ובמקרים רבים התוצאות משביעות רצון. בעיה זו הינה קצהו של קרחון של בעיות הקרויות בעיות היפוך ואלו רבות ופופולאריות בעיבוד תמונה. בעיה אחרת ממשפחה זו היא הבעיה הבאה: נתונה תמונה באיכות טובה אך קטנה. רצוננו להגדילה לשם הדפסה. כל מי שהתנסה בהגדלת תמונות יודע כמה מתסכל לראות את התוצאה שתהיה בדרך-כלל מטושטשת. בשנים האחרונות מוצעות גישות שונות לייצר הגדלות תוך שמירת איכות, ויצירת האשליה שהתוצאה היא תמונה טבעית.

וריאציה אחרת של בעיית היפוך היא בעיית הסופר-רזולוציה. נניח כי יש לנו 100 תמונות באיכות ירודה של אובייקט מצולם. האם ניתן לאחד את המידע בהן לשם יצירת תמונה אחת ברזולוציה משופרת? מסתבר כי התשובה היא חיובית והתוצאות מדהימות! מערכות בטחון מצלמות לא פעם פושע בפעולה. עם הקפאת התמונה והדפסתה האיכות לא מאפשרת זיהוי. טכניקות סופר-רזולוציה תאפשרנה לא פעם בניית תמונה ממנה הזיהוי אפשרי.



תוצאת שחזור תמונה – מקור (למעלה משמאל), גרסה פגומה (למעלה מימין)
ושחזור בשתי גישות

מעבר לדחיסה ושיפור תמונות ישנם אלגוריתמים רבים שתכליתם ניתוח תוכן התמונה או סרט. בסרט וידאו למשל ניתן לשערך ולנתח תנועה בין תמונות. על בסיס זה למשל, ניתן לגלות חדירות לאזור מוגן. ניתוח תנועה כזה נמצא גם בבסיסה של מערכת ניווט מטוס המצלם את הקרקע (כפי שטילי טומהוק עושים). מתנועה ואופייה נוכל להקיש על המבנה התלת-מימדי של האובייקטים המצולמים. לשחזור העומק מקום חשוב בעיבוד תמונה, וגישות שונות לבניית העומק מתמונות קיימות (מסטריון, מתנועה, מהצללה, מפוקוס, מתאורה משתנית, ועוד).



תוצאת סופר-רזולוציה

משימה חשובה בתמונות היא פירוק תמונה סמנטי (בחלוקה לאובייקטים טבעיים) וסגמנטציה. הרעיון הוא לחלק את התמונה לאזורים המתייחסים לאובייקטים שונים. נושא זה מורכב כיוון שאם למשל צולם אדם וחולצתו נושאת מירקם, יתכן כי החולצה תחולק לנתחים, סביר להניח כי לא תצורף לראש והידיים כי אלה בגוונים שונים ועוד. משימה דומה היא משימת וקטוריזציה של תמונות המתארות שרטוטים טכניים או מפות, שם המטרה היא לגלות אובייקטים בסיסיים כמו קו, חץ, עיגול, ועוד, ולהמיר התיאור הרסטרי (Raster) בתיאור של אובייקטים עניינים ללא אובדן מידע.



תוצאת עקיבה אחר אדם הולך בסרט וידאו (תוצאות מעבודת הדוקטורט של רומן גולדנברג מהפקולטה למדעי המחשב בטכניון)

ניתן לחפש אובייקטים בתמונה כגון פרצופים, וניתן אף להציע זיהוי. כך למשל פועלת כל תוכנה המלווה סורק בביצוע OCR (Optical Character Recognition), בה דף עם טקסט מצולם כתמונה ולאחר מכן מומר ל-ASCII ע"י זיהוי כל אות בתמונה. אם מדובר בסרט וידאו, לא די בגילוי האובייקט אלא נחוצה גם עקיבה אחריו ואיתורו בכל תמונה תוך בחינה שהתנועה הנצפית ריאלית.



גילוי פרצופים בתמונות

כיצד נפיק תמונה מקובץ?

דיברנו עד כה על הכנסת התמונה למחשב ועל עיבודה בו. לעיתים די בכך, אם בכוונתנו לסיים העיבוד בלקיחת מסקנות מהתמונה באופן אוטומטי, כגון גילוי אוטומטי של גידולים סרטניים בתמונות, או זיהוי האדם הנכנס לחדר מאובטח ע"י תווי פניו ופתיחת הדלת בהתאם. לעומת זאת, במקרים רבים הצרכן במערכת הוא אדם ואז יש להציג התמונה החוצה. אפשרות אחת העומדת בפנינו היא שימוש במוניטור להצגת התמונה. מוניטור אופייני מבצע את הפעולה ההפוכה לפעולת הגלאי במצלמה דיגיטאלית – בהינתן ערכו של פיקסל מוזרם זרם חשמלי פרופורציוני לו וזה מומר לפוטונים הנפלטים מתא קטן – זה יהיה הפיקסל שאותו בפועל נראה. אם המסך כולל המון נקודות כאלו (תתקרבו למסך הטלוויזיה ותראו!), לצופה האנושי תיווצר התחושה של תמונה רציפה ונעימה. קצב הרענון של המסך, כמות התאים הפלואורוצנטיים בו וצפיפותם, ומרחב הצבעים אותם הם יכולים לייצר – כל אלה מכתיבים את איכות התמונה המופקת.

לא פעם נדרשת פעולת התאמה של התמונה בטרם הצגתה למסך על מנת למונע תופעות קיפול. לדוגמה, קו אלכסוני חד בתמונה עלול להופיע במסך כשהוא מקוטע ומשונן. כשהתמונה מוקטנת על מנת להיכנס למסך, מרקמים עדינים יכולים לקבל עיוותים איומים. פעולות תיקון מתאימות נעשות בדרך-כלל על ידי הכרטיס הגרפי הדוחף את המסך.

הפקת תמונה מקובץ יכולה להיעשות גם ע"י הדפסה. עלינו לצבוע על-גבי נייר בצבעים שונים בהתאם לערכי הפיקסל. מדפסות Continuous Tone כאלה קיימות אך מאוד יקרות. האלטרנטיבה היא לקבוע מספר מחסניות של צבעים בסיסיים במדפסת ולכפות על תהליך הדפסת התמונה להשתמש רק בצבעי הקיימים תוך ניסיון להתקרב ככל האפשר לתמונה המקורית. דוגמה קיצונית למצב זה היא הדפסת תמונות בעיתון שחור-לבן, שם מוחלף ערך הפיקסל בכתם שחור בגודל משתנה – גדול בשביל גוון כהה וקטן עבור ערך בהיר. תהליך זה הקרוי Dithering או Error-Diffusion עומד מאחורי הצלחתן הכבירה של מדפסות ה-ink-jet ויכולתן להדפיס תמונות באימות כה רבה.



תהליך Dithering בהדפסת תמונה

תחומים סמוכים

כפי שכבר הסברנו קודם, "עיבוד תמונות" הוא תחום אשר צמח מתוך "עיבוד אותות" בו דנים ברמה הבסיסית באפיונם של אותות, ואפיון מערכות המטפלות בהם. מטבע הדברים, כשהדיון ניתן ברמה כללית כזו, הדגש הוא על דיון מתמטי עקרוני ופחות באלגוריתמים התפורים לאפליקציות מיוחדות. מתוך

עיבוד אותות נוצרו תתי-תחומים כמו הטיפול באותות דיבור, אותות מוזיקליים, אותות מכ"מיים, ואחרים. בתתי-תחומים אלה ישנה התמקצעות בתכונות הייחודיות לאותות הנדונים. לדוגמה, ניתוח זיהוי של אות-דיבור מורכב למדי ולמרביתו אין כל רלוונטיות כשעוסקים בתמונות.

עד כה עשינו שימוש במושג "עיבוד תמונה" כמושג המקיף את כל הפעילות הדנה בטיפול ממוחשב בתמונות, ואמנם כך הייתה הגישה בתחילת הדרך. עם התרחבות הידע בטיפול בתמונות החלו להיווצר תתי-התמחויות ותחומים חדשים. הראשונה והגדולה שבהם היא "ראיה ממוחשבת". ההבחנה בין "עיבוד תמונה" ל-"ראיה ממוחשבת" אינה חדה וברורה לחלוטין (ועל כך יעידו, למשל, החפיפות בנושאים הנדונים בכנסים בתחומים השונים). עם זאת ישנה הסכמה על החלוקה העקרונית הבאה – אם מדובר באלגוריתם המקבל תמונה ומחזיר תמונה (כמו שיפור או דחיסה) זה "עיבוד תמונה", בעוד שאם המוצא אינו תמונתי (גילוי מטרות, שערור תנועה, חלוקה לאזורים, שערור עומק) זוהי "ראיה ממוחשבת". ברוח זו גם נאמר כי "גרפיקה חישובית" דנה בתחום בו מפותחים אלגוריתמים המקבלים מידע שאינו תמונתי ומפיקים תמונה במוצאם.

גם "ראיה ממוחשבת" צמחה עם השנים והתפצלה לתתי-פעילויות. נתח מחקרי עצמאי ורחב שנפרד מתחום זה הוא "זיהוי תבניות" שכל תכליתו גילוי זיהוי עצמים בתמונה. פעילות זו משיקה וקרובה לתחום אחר מרתק הקרוי "למידה חישובית" בו עוסקים סטטיסטיקאים, מתמטיקאים, ואנשי מחשבים.

מכיוון שעיבוד תמונה נתונה כרוך לא פעם בהבנת דרך קבלתה, ישנו קשר הדוק בין "עיבוד תמונות" ו-"אופטיקה"/"אלקטרו-אופטיקה". תהליך בניית תמונה מתחיל באופטיקה נאותה, עבור דרך גלאים אלקטרו-אופטיים, ורק אז נתון לנו כמטריצת פיקסלים.

הפעילות בעיבוד תמונה ענפה, ולרוב חוקרים ומהנדסים העוסקים בתחום מתמחים בחלקה אחת, ויודעים מעט על יתר הפעילויות. ברמת היכרות בסיסית עם התחום נדמה כי די באינטואיציה טובה כדי לפתור בעיות בטיפול בתמונות. עם השנים ניכר כי גישה זו יכולה להוביל לכברת דרך קצרה וחסומה.

ידע רב בתחום "עיבוד תמונה" נשען על שימוש בכלים מתמטיים מתקדמים. ואמנם, בשנים האחרונות הפכה "עיבוד תמונה" לכר פעילויות של מתמטיקאים המחפשים דרכים לרתום את הידע שלהם. פרקים כגון אלגברה ליניארית, חשבון דיפרנציאלי, גיאומטריה, גיאומטריה דיפרנציאלית, משוואות דיפרנציאליות, תורת הקבוצות, מורפולוגיה, תורת האינפורמציה, סטטיסטיקה,

תורת ההסתברות, תורת האותות והמערכות, כאוס ופרקטלים, תורת הקירובים, ועוד – כולם משמשים כיום בעיבוד תמונה. כך יוצא כי סיבה מרכזית לקושי ללמוד "עיבוד תמונה" לעומק כיום היא הצורך להכיר מגוון גדול מאוד של כלים מתמטיים מתקדמים.

יישומים ומערכות עם עיבוד תמונה

יש מגוון רחב של מוצרים כיום בשוק בהם יש שימוש בטכנולוגיית עיבוד תמונה. נזכיר כאן רשימה חלקית:

- תמונות ווידאו באינטרנט – כל מידע מסוג תמונתי המופיע באינטרנט מופיע כשהוא דחוס. מעבר לכך, יש לא אחת צורך בשמירת זכויות יוצרים, ורצון של יוצר התמונה/סרט להטביע בו "חתימה אישית" שתעיד על מקוריות ותמנע הכחשה של מעתיקים. גם הטבעת חתימה כזו נעשית בטכניקות עיבוד תמונה.
- וידאו-פון ומערכות Video-Conferencing – בדומה לתקשורת אינטרנט, דחיסה חיונית כדי לאפשר תקשורת יעילה בין הצדדים. מעבר לכך, נחוצות מצלמות ודרישה להתמודד עם תנאי תאורה לקויים, מעבר אוטומטי מדובר לדובר, ועוד.
- טלוויזיה בכבלים – גם כאן דחיסה חיונית. מעבר לכך בטלוויזיות חדישות ישנו מגוון רחב של עיבודים כדי להקפיא תמונה באיכות טובה, להמיר מבנה מ-interlace ל-progressive, הגדלת תמונה, גילוי פרסומות והורדתן ב-TIVO, ועוד.
- מצלמה דיגיטלית – כבר הזכרנו כי נדרשת שורה ארוכה של עיבודים כדי לקחת את המידע מהגלאים ולהופכו לתמונה נעימה לעין.
- מדפסת – ראינו כי נחוצים אלגוריתמים שידעו להפיק את המרב מסט צבעים נתון. בנוסף מקובל לתת כלי עיבוד בתוכנה למשתמש לשם שיפור התמונה בטרם הפקתה.
- גילוי פגמים בייצור מעגלים מודפסים – טכנולוגיה שלמה הוקמה סביב יכולות עיבוד תמונה לשם גילוי אוטומטי של פגמים במעגלים מודפסים, ובתעשיית המיקרו-אלקטרוניקה. הרעיון הוא לספק למערכת דוגמה גנרית של המראה הנכון ולבקש מהמערכת לזהות חריגות ולאמוד אוטומטית את משקלן.

- צילום רפואי וטומוגרפיה – כל תחום הצילום הרפואי נשען חזק על עיבוד תמונות. הדבר נכון במיוחד בתחום שזכה להיות תחום עצמאי בעיבוד תמונה, הוא הטומוגרפיה החישובית. בתחום זה נמדדים היטלים על האובייקט המצולם, ומאלה נבנית תמונה.

עיבוד תמונות בטכניון בישראל

הטכניון מהווה מרכז ידע ברמה עולמית בתחום עיבוד תמונה. חברי הסגל העוסקים בתחום זה באופן פעיל הם:

- מדעי המחשב - פרדי ברוקשטיין, אודי ריבלין, מיכה לינדנבאום, רון קימל, עירד יבנה, מיכאל אלעד.
- חשמל - דוד מלאך, יהושע זאבי, אריה פויאר, אלן טננבאום, משה פורת, יואב שכנר, מיכאל צ'יבולבסקי.
- תעשייה וניהול - אילן שמשוני, דוב דורי.

וישנם גם אחרים העוסקים בתחום זה לפרקים ולא באופן מרכזי.

גם באוניברסיטאות אחרות בישראל רבה הפעילות בתחום זה. באוניברסיטה העברית יושבת קבוצה גדולה המתמחה בעיבוד תמונות וראיה ממוחשבת, וכך גם הדבר במכון ויצמן, באוניברסיטת ת"א, באוניברסיטת חיפה, ובבאר-שבע.

בתעשייה הישראלית יש מקום מכובד לפעילויות בעיבוד תמונה. בכל החברות הטכנולוגיות הגדולות (אלביט, אלאופ, התעשייה האווירית, אלתא, אינטל, KLA, אפלייד-מטריאלס, HP, צורן, נס, ורבות אחרות) יש פעילות עשירה בתחום זה. ישנן גם חברות סטרט-אפ קטנות רבות המריצות רעיונות בתחום זה ליישומי תקשורת, ציום ביתי, צילום רפואי, ועוד.

עיבוד תמונות בעזרת מחשב (236327)

הרצאה מס' 2

אותות ומערכות 1 – אותות בקואורדינטות רצף

מאפייני אותות

החיים מזמנים לנו שפע רחב של אותות מסוגים שונים. אות הינו פונקציה המתארת גודל (או גדלים) מסוימים כפונקציה של הזמן, המקום, או תחום הגדרה אחר. לדוגמה, גובה של אדם מיום לידתו ועד מותו היא אות המתאר אורך כפונקציה של הזמן. דוגמה אחרת לאות היא הלחץ הברומטרי במדינה מסוימת בכל מקום ובכל נקודת זמן. כפי שכבר הבהרנו בפרק קודם, תמונה מתארת גוון או בהירות כפונקציה של המקום.

כזכור, כשעוסקים בפונקציה כלשהי, משתמשים בסימון $f(x): \Omega \rightarrow \Theta$ האומר כי התחום ממנו באים משתני הפונקציה הוא Ω והטווח של ערכי המוצא של הפונקציה הם מהשדה Θ . שדות אלו (Ω ו- Θ) יכולים להיות שדה הממשיים, השלמים, שדות סופיים ועוד.

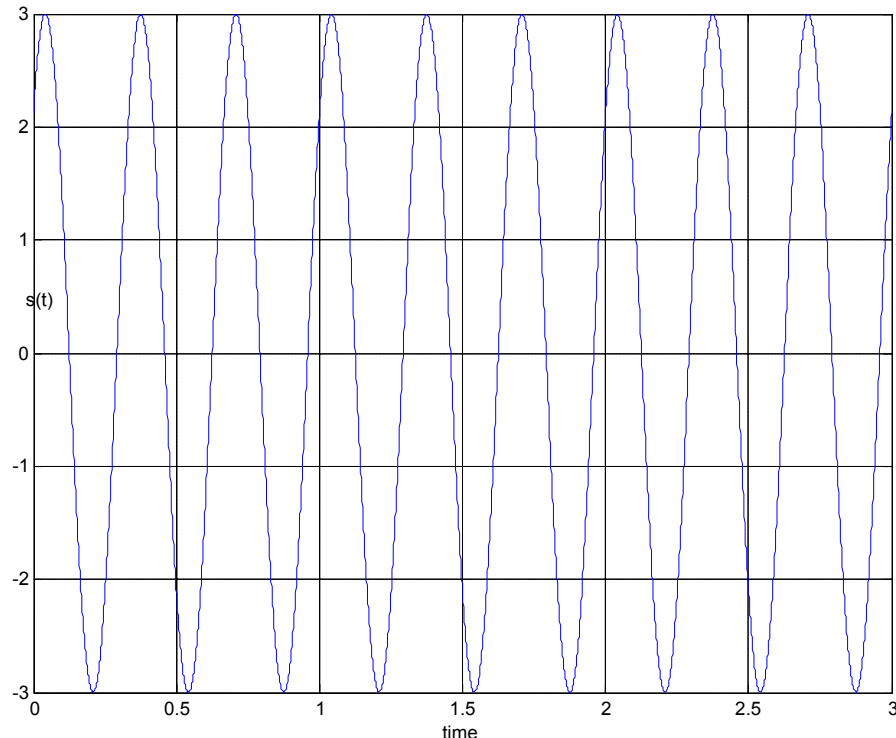
ברוח זו, אם נסמן את האות כפונקציה $S: F^D \rightarrow G^L$ פירוש הדבר שפונקציה זו מקבלת D פרמטרים (מקום, זמן, או כל גורם התייחסות אחר) מהשדה F בתחום, ומייצרת L ערכי מוצא בשדה G בטווח. לדוגמה, אם זוהי פונקציה של הזמן, שדה התחום F יהיה המספרים הממשיים. אם מדובר במקום, נוכל להגדיר את שדה התחום כסקלר מרוכב או כשני מספרים ממשיים. יתכן גם ששדה התחום יהיה בדיד, כמו התייחסות לחץ הדם של חולה כפונקציה של הזמן אך רק בשעה עגולה ביום, היכולה לקבל 24 ערכים אפשריים.

באופן דומה, הטווח של הפונקציה יכול להיות דיסקרטי – לדוגמה בינארי. לדוגמה, אות המתאר את מצבו של סכר כפתוח או סגור כפונקציה של היום בשנה ייתן כי שדה הטווח G הוא בינארי. המוצא של הפונקציה – ערכי

האות - יכול להיות ממשי, מדומה, ועוד. לכן, אות יאופיין ע"י הבחירה F - שדה ההגדרה, הבחירה G - שדה הטווח, D - מימד התחום, ו- L - מימד הטווח.

ניתן להבחין בין אותות גם ע"י התייחסות לאופיו של מקור האות - מקור טבעי או מקור מלאכותי, מקור אקראי או מקור דטרמיניסטי, ועוד. לאבחנות אלו חשיבות רבה בבואנו לבחור את כלי העיבוד לאות, אך אנו לא ניכנס לנושאים אלו במסגרת זו.

לעת עתה ועד הודעה אחרת, אנו נניח כי האות שלנו הוא אות פשוט וחד מימדי בתחום ובטווח, הניתן להתייחסות כפונקציה $s(t)$ של הזמן המייצרת ערכי מוצא ממשיים וסקלריים. לדוגמה, האות $s(t) = 3\sin(6\pi t + 0.25\pi)$ הוא אות סינוסואידלי המאופיין ע"י אמפליטודה בגובה 3, תדר השתנות הנותן ביחידת זמן אחת 3 מחזורים שלמים, ופאזה של רבע מחזור. הציור הבא ממחיש איך נראה אות זה. זהו אות מחזורי - עוד תכונה שיכולה לאפיין אותות. אות זה יכול להיות למשל מקור מתח AC המסופק למערכת חשמלית.



תיאור של אות סינוסואידלי

גרף זה התקבל ע"י הפקודות הבאות ב-Matlab:

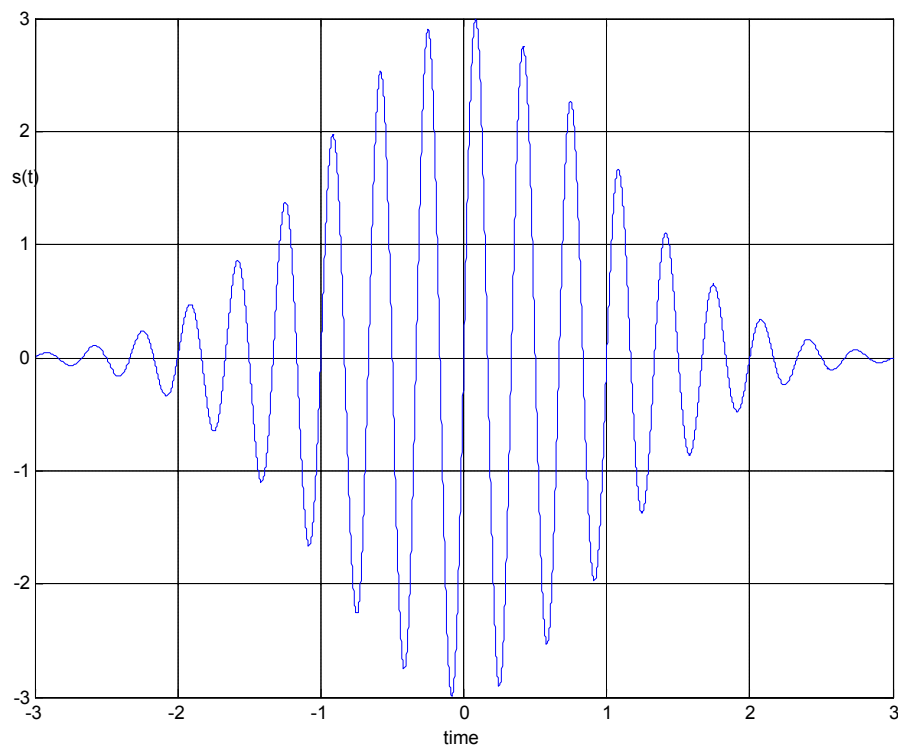
```
plot(0:0.001:3,3*sin(6*pi*(0:0.001:3)+pi/4));  
grid on; xlabel('time'); ylabel('s(t)');
```

הציור הבא מראה אות מורכב יותר שבו נתקלים לעתים קרובות בעיבוד אותות – אות סינוסואידלי מחזורי המדוכה ע"י מישקול גאوسي (אות זה קרוי פונקציית גאבור ומרבים להשתמש בה מסיבות שלא יפורטו כאן):

$$s(t) = 3 \exp\left\{-\frac{t^2}{2}\right\} \sin(6\pi t)$$

הפעם נעשה שימוש בפקודות הבאות ב-Matlab לייצור הגרף:

```
t=-3:0.001:3; s=3*exp(-t.^2/2).*sin(6*pi*t);  
plot(t,s); grid on; xlabel('time'); ylabel('s(t)');
```



תיאור של אות סינוסואידלי מוכפל בגאוסיאן החוסם את טווח קיומו בזמן

בהינתן אות נרצה לדעת את תכונותיו. מעבר למאפיינים אותם הזכרנו קודם, נוכל להתעניין בערכו הממוצע המנורמל למשל. גודל זה יהיה

$$.M = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T s(t) dt$$

כך למשל, ברור כי ממוצעם המנורמל של שני האותות שתוארו בגרפים קודם הוא אפס, אם כי אין זה טריוויאלי להשתמש בהגדרה ישירות כדי לראות זאת. במקרה של האות הסינוסואידלי, כאשר T גדל הממוצע אינו מתקרב באופן עקבי לערך האפס אלא מתנדנד תוך כדי התקרבות לאפס, כי

$$\begin{aligned} M &= \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T 3 \sin(6\pi t + \pi/4) dt = \\ &= \lim_{T \rightarrow \infty} \frac{3}{2T} \int_{-T}^T [\sin(6\pi t) \cos(\pi/4) + \cos(6\pi t) \sin(\pi/4)] dt = \\ &= \lim_{T \rightarrow \infty} \frac{[-\cos(6\pi t) \cos(\pi/4) + \sin(6\pi t) \sin(\pi/4)]_{-T}^T}{4\pi T} = \\ &= -\lim_{T \rightarrow \infty} 3 \sin(\pi/4) \frac{\sin(6\pi T)}{6\pi T} = 0 \end{aligned}$$

הפיתוח עבור האות השני יוצא יותר קל בשל היותו אנטי-סימטרי, $s(t) = -s(-t)$. בשל כך יוצא כי האינטגרל על תחום סימטרי סביב האפס מתאפס זהותית לכל T , כי

$$\begin{aligned} M &= \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T s(t) dt = \lim_{T \rightarrow \infty} \frac{1}{2T} \left[\int_{-T}^0 s(t) dt + \int_0^T s(t) dt \right] = \\ &\stackrel{\substack{\uparrow \\ a=-t}}{=} \lim_{T \rightarrow \infty} \frac{1}{2T} \left[-\int_T^0 s(-a) da + \int_0^T s(t) dt \right] = \\ &= \lim_{T \rightarrow \infty} \frac{1}{2T} \left[\int_0^T s(-t) dt + \int_0^T s(t) dt \right] = 0. \end{aligned}$$

ובאופן דומה נוכל להתעניין בערכי הממוצע הכללי (שאינו מנורמל), בהנחה שהאינטגרל מתכנס:

$$M = \int_{-\infty}^{\infty} s(t) dt$$

ועבוד האות הסינוסואידלי ממוצע זה אינו מוגדר (הוא משתנה כפונקציה של דרך הגדרת גבולות האינטגרל בדרכם לאינסוף).

ממוצע האות השני הוא אפס בשל האסימטריה שלו. נשים לב לנקודה עדינה כאן - גם אות סינוסואידלי רגיל הוא א-סימטרי ולמרות זאת לא ניתן להשתמש בטיעון זה כדי לומר שממוצעו אפס. הסיבה היא שהאינטגרל מאפס לאינסוף צריך להתכנס כדי לקזז את האינטגרל על חצי הציר השני. דבר זה קורה באות השני שלנו אך לא באות סינוסואידלי רגיל שאינו דועך.

כלי חשוב שמרבים להשתמש בו בניתוח אותות הוא מדד מרחק בין אותות שונים. מקובל להשתמש במדד שגיאה ריבועית ממוצעת המוגדר לפי

$$MSE = \int_{-\infty}^{\infty} |s(t) - g(t)|^2 dt$$

כמקרה פרטי של מדד זה, נגדיר את האנרגיה המצויה באות כ:

$$E = \int_{-\infty}^{\infty} |s(t)|^2 dt$$

בהגדרה זו אנו בעצם מחשבים את המרחק בין האות $s(t)$ ובין אות האפס (שערכו אפס בכל נקודת זמן). בהפעלת ערך מוחלט אנו מתגוננים כנגד האפשרות שהאות בעל ערכים מרוכבים. לבחירה להעלות את האות בריבוע בשתי ההגדרות הללו קשר לעובדה שאנו בוחרים (שרירותית!) להגדיר מרחק בין אותות ואנרגית אות בעזרת נורמת L_2 של הפונקציה. נשים לב כי האנרגיה של האות הסינוסואידלי היא אינסופית בשל היותו קיים בכל נקודת זמן. לאותות כאלה מקובל להגדיר כתחליף את המושג הספק שהינו אנרגיה ליחידת זמן, המוגדרת לפי:

$$P = \frac{1}{2T} \int_{-T}^T |s(t)|^2 dt$$

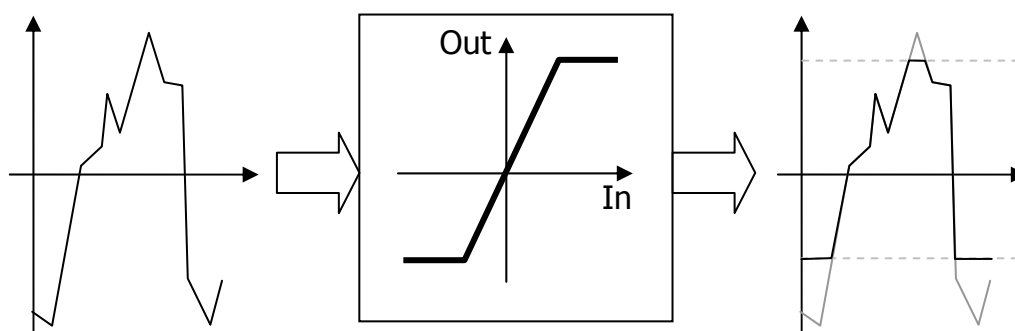
אורך הקטע הוא $2T$, ובאות מחזורי מעין זה טבעי לקבוע אותו להיות מחזור אחד, דהיינו $T = 1/6$ בדוגמה זו. זה ייתן

$$\begin{aligned} P &= \frac{6}{2} \int_{-1/6}^{1/6} |3 \sin(6nt + \pi/4)|^2 dt = 27 \cdot \int_{-1/6}^{1/6} \sin^2(6nt + \pi/4) dt = \\ &= 27 \cdot \int_{-1/6}^{1/6} \frac{1 - \sin(12nt + \pi/2)}{2} dt = 27 \cdot \int_{-1/6}^{1/6} \frac{1}{2} dt = \frac{27}{2} \cdot \frac{1}{3} = 4.5. \end{aligned}$$

פעולות על אותות ומערכות

מה נוכל לעשות על אותות? מגוון רחב של פעולות! בהינתן האות $s(t)$ נוכל להכפילו בקבוע a ולקבל אות חדש $a \cdot s(t)$. נוכל להעלות האות בריבוע, כלומר ליצור אות חדש $g(t) = s^2(t)$, ובעצם נוכל להציע כל פונקציה מוכרת לנו $\sin, \cos, \log, \exp$ (ועוד) אשר תפעל על האות ותיתן $g(t) = f\{s(t)\}$. כל אפשרויות אלו קרויות מערכות חסרות זיכרון כיוון שלשם קביעת g בזמן t כל

שנדרש הוא לדעת את s באותו זמן ואת עקומת ההתאמה של הערכים f , אותה ניתן לתאר כגרף שמראה כיצד מומר כל ערך כניסה לערך יציאה.



מערכת חסרת זיכרון המיועדת לקטימת אות הכניסה

נוכל גם להציע מערכות עם זיכרון. לדוגמה, מערכת מהצורה $g(t)=s(t)+s(t-1)$ היא מערכת המפעילה זיכרון. כך גם המערכת

$$g(t) = \int_{-1}^1 s(t-x)dx$$

אשר מחשבת את השטח מתחת לאות באינטרוול $[-1, +1]$ סביב נקודת הזמן x .

מערכות עם זיכרון נשענות על מרכיב בעל חשיבות מרכזית – הזזת זמן. המערכת הבסיסית $g(t)=s(t-d)$ משהה את אות הכניסה בשיעור d ואז מספקת אותו במוצאה. באופן דומה, המערכת $g(t)=s(t+d)$ מקדימה את אות הכניסה בשיעור דומה.

נסתכל על שתי מערכות T_1 ו- T_2 המשתמשות בזיכרון ונבחין בהבדל מהותי ביניהן:

$$g_1(t) = T_1\{s(t)\} = \int_0^1 s(t-x)dx$$

$$g_2(t) = T_2\{s(t)\} = \int_{-1}^1 s(t+x)dx$$

המערכת T_1 משתמשת בערכי אות הכניסה $s(t)$ בזמנים $[t-1, t]$ על מנת לקבוע את מוצא המערכת בזמן t . על מערכת זו נאמר כי היא סיבתית. המערכת השנייה אינה סיבתית כיוון שלשם חישוב מוצאה בזמן t היא צריכה לדעת את ערכי הכניסה בעבר ובעתיד, בתחום הזמנים $[t-1, t+1]$. סיבתיות הינה תכונה

חשובה למערכות הפועלות ב-"זמן אמת" (כלומר, בכל רגע נתון מוציאות מוצא המתייחס לכל המידע שהצטבר עד לרגע זה).

מערכת יכולה לכלול מרכיבי משוב בתוכה. לדוגמה, המערכת $g(t) = g(t-1)s(t)$ משתמשת בערכי העבר כדי לקבוע את מוצאה בזמן t . מערכות כאלה, הנקראות מערכות רקורסיביות, מאופיינות בכך שהזיכרון האפקטיבי שלהן אינסופי (הן כוללות השפעת הכניסה לאורכי זמן ארוכים מאוד).

תכונה חשובה נוספת היא יציבות המערכת. היינו רוצים לדעת שאם מערכת ניזונה מאות חסום (כזה שערכיו מצויים בתוך תחום ערכים סופי), גם מוצאה כזה. תכונה זו קרויה יציבות BIBO (Bounded Input – Bounded Output). לדוגמה, המערכת $g(t) = s^2(t)$ יציבה, בעוד שהמערכת הרקורסיבית $g(t) = s(t) + g(t-1)$ אינה יציבה. אם אות הכניסה קבוע ושווה 1, מוצא המערכת יגדל ויגדל באופן ליניארי במדרגות עד לאינסוף. גם מערכת פשוטה יותר כמו $g(t) = \log[s(t)]$ תאבד יציבות, כי עבור אות כניסה בתחום $(0,1]$ ערכי המוצא יגיעו ל $-\infty$.

נאמר על מערכת כי היא קבועה בזמן אם הפעלתה על אות המוזן בזמן מניבה את אותו המוצא כמו הפעלתה על האות המקורי והזת מוצאה בזמן. כלומר, עבור המערכת T , אם לאות שרירותי $s(t)$ המוצא הוא $g(t) = T\{s(t)\}$, אזי אם מתקיים כי $\forall d, g(t+d) = T\{s(t+d)\}$, הרי שהמערכת קבועה בזמן. אינטואיטיבית, פירוש הדבר שהמערכת פועלת בצורה זהה בזמנים שונים. לדוגמה, המערכת $g(t) = t \cdot s(t)$ אינה קבועה בזמן כי הזת מוצאה נותנת

$$g(t+d) = (t+d) \cdot s(t+d)$$

בעוד שהזת הכניסה והזנתה תיתן

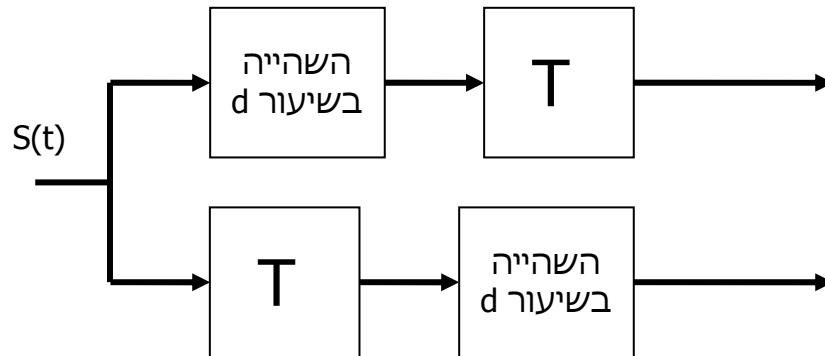
$$t \cdot s(t+d)$$

לעומת זאת, המערכת $g(t) = s(t-1) \cdot s(t)$ קבועה בזמן כיוון שהזת המוצא נותנת

$$g(t+d) = s(t-1+d) \cdot s(t+d)$$

והזת הכניסה מביאה לאותו ביטוי.

בעצם, בהגדרת התכונה "קביעות בזמן" אנו שואלים אם החלפת הסדר בין שתי מערכות – האחת השהייה והאחרת המערכת שבמבחן – תיתן התנהגות זהה.



שחלוף השהייה והמערכת כמבחן להיות המערכת קבועה בזמן

נושא זה של שרשור מערכות וחיבורם בדרכים שונות חשוב. אנו נאמר על שתי מערכות T_1 ו- T_2 כי הן משתחלפות עם מתקיים כי

$$T_1\{T_2\{s(t)\}\} = T_2\{T_1\{s(t)\}\}$$

לדוגמה, עבור המערכות שהגדרנו קודם

$$g_1(t) = T_1\{s(t)\} = \int_0^1 s(t-x)dx$$

$$g_2(t) = T_2\{s(t)\} = \int_{-1}^1 s(t+x)dx$$

יתקבל כי

$$T_1\{T_2\{s(t)\}\} = \int_0^1 \left(\int_{-1}^1 s(t-u+x)dx \right) du$$

$$T_2\{T_1\{s(t)\}\} = \int_{-1}^1 \left(\int_0^1 s(t+u-x)dx \right) du = \int_0^1 \int_{-1}^1 s(t-u+x)dxdu$$

ושתי מערכות אלו זהות. באופן דומה, שתי המערכות חסרות הזיכרון

$$g_1(t) = s^2(t) \text{ ו- } g_2(t) = \sqrt{|s(t)|} \text{ גם כן משתחלפות כיוון ש-}$$

$$T_1\{T_2\{s(t)\}\} = \left(\sqrt{|s(t)|} \right)^2 = |s(t)|$$

$$T_2\{T_1\{s(t)\}\} = \sqrt{|s^2(t)|} = |s(t)|$$

לעומת זאת, שתי המערכות $g_1(t) = s^2(t)$ ו- $g_2(t) = s(t) + s(t-1)$ אינן משתחלפות כי

$$\begin{aligned} T_1\{T_2\{s(t)\}\} &= (s(t) + s(t-1))^2 \\ T_2\{T_1\{s(t)\}\} &= s^2(t) + s^2(t-1) \end{aligned}$$

כשם שניתן לחבר מערכות בטור בזו אחר זו, ניתן גם לחברן במקביל ולסכם את מוצאיהן. בצורה זו, אם נתונות שתי מערכות T_1 ו- T_2 , חיבורן ייתן את המערכת השקולה

$$.g(t) = T_1\{s(t)\} + T_2\{s(t)\}$$

מערכות ליניאריות וקבועות בזמן

מתוך מגוון המערכות האפשריות, ישנה תת-משפחה רחבה בה נרבה לעסוק, והיא המשפחה הליניארית. מערכת מוגדרת כליניארית אם היא מקיימת את שתי התכונות הבאות:

- תכונת ההומוגניות: עבור כל אות $s(t)$ וסקלר a מתקיים כי

$$.T\{a \cdot s(t)\} = a \cdot T\{s(t)\}$$

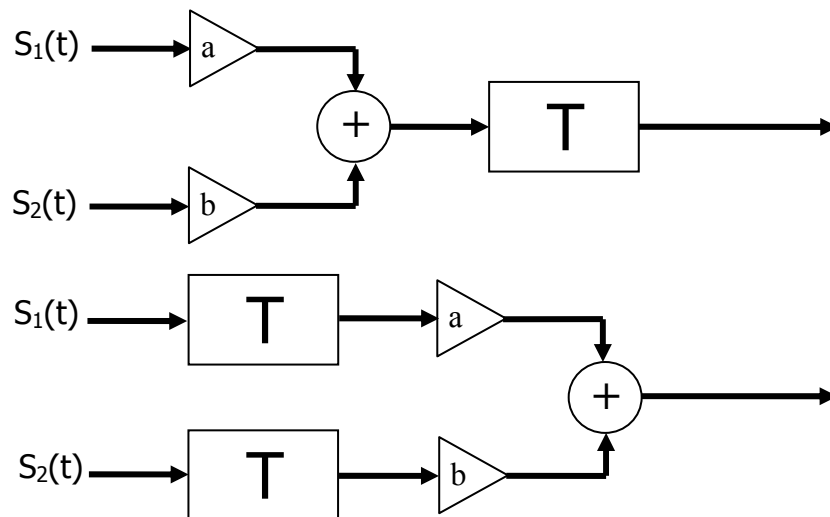
- תכונת האדיטיביות: עבור כל שני אותות $s_1(t)$ ו- $s_2(t)$ מתקיים כי

$$.T\{s_1(t) + s_2(t)\} = T\{s_1(t)\} + T\{s_2(t)\}$$

ניתן להחליף שתי תכונות אלו בתכונה אחת שקולה לחלוטין האומרת שלכל שני אותות $s_1(t)$ ו- $s_2(t)$ ולכל שני סקלרים a ו- b , יתקיים כי

$$.T\{a \cdot s_1(t) + b \cdot s_2(t)\} = a \cdot T\{s_1(t)\} + b \cdot T\{s_2(t)\}$$

ברמה אינטואיטיבית נאמר כי מערכת היא ליניארית אם מוצאה הוא צירוף ליניארי של ערכי הכניסה. כשאנו אומרים צירוף ליניארי, מותר שיהיו משקלים שונים בצירוף זה, ואף משקלים תלויי זמן.



מערכת T ליניארית אם שתי המערכות הנ"ל שקולות

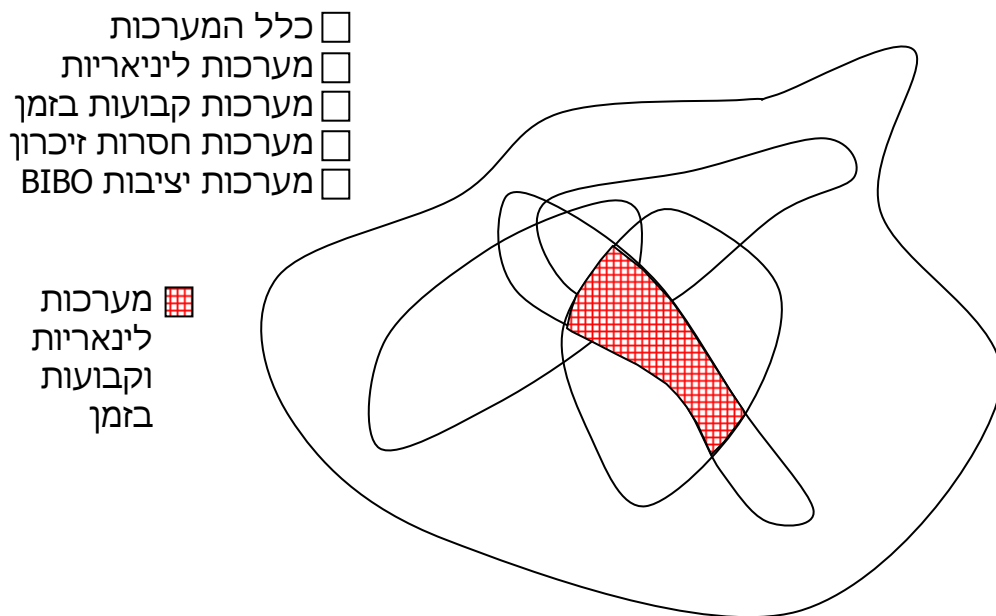
ניתן מספר דוגמאות להמחיש זאת:

- המערכת $g(t) = s(t) + 3s(t-1)$ היא מערכת ליניארית. למעשה זו מערכת סיבתית, עם זיכרון, והיא גם קבועה בזמן.
- המערכת $g(t) = s(t) + \sqrt{t} \cdot s(t-1)$ תלויה בזמן, אך היא עדיין ליניארית, כי היא בונה את מוצאה כצירוף ליניארי של אות הכניסה.
- המערכת $g(t) = 2s(t) + 3$ אינה ליניארית. קל לראות זאת לפי ההגדרה, כיוון שתכונת ההומוגניות $T\{a \cdot s(t)\} = 2a \cdot s(t) + 3 \neq a \cdot T\{s(t)\} = 2a \cdot s(t) + 3a$ לא מתקיימת.
- המערכת $g(t) = \int_{-\infty}^t s(x)dx$ היא מערכת ליניארית. זוהי מערכת לא יציבה BIBO – אות קבוע יביא את מוצאה לאינסוף.
- באופן דומה, המערכת $g(t) = \int_{-\infty}^t f(x)s(x)dx$ עבור פונקציה כלשהי ידועה מראש f תהיה ליניארית.
- המערכת $g(t) = s(t)s(t-1)$ אינה ליניארית כי היא מצרפת את ערכי אות הכניסה באופן שאינו צירוף ליניארי.
- המערכת $g(t) = s'(t) = ds(t)/dt$ (גזירה) היא מערכת ליניארית! נראה זאת לפי ההגדרה:

$$\begin{aligned} T\{a \cdot s_1(t) + b \cdot s_2(t)\} &= \frac{d}{dt}[a \cdot s_1(t) + b \cdot s_2(t)] = \\ &= a \frac{ds_1(t)}{dt} + b \frac{ds_2(t)}{dt} = \\ &= a \cdot T\{s_1(t)\} + b \cdot T\{s_2(t)\} \end{aligned}$$

למעשה, מערכת זו גם קבועה הזמן כיוון שאם $g(t) = ds(t)/dt$, אזי הזזת הכניסה תביא להזזה דומה במוצא, דהיינו $g(t-d) = ds(t-d)/dt$.

כפי שכבר אמרנו, אנו נתעניין במסגרת קורס זה במערכות ליניאריות וקבועות בזמן. אנו נראה כי קיימים כלים מתמטיים נוחים לטיפול במערכות אלו.



"עולם" המערכות והיחסים ביניהן

תכונה מעניינת של מערכת ליניארית היא היכולת לשחלפה עם פעולת סיכום. זוהי הבחנה טריוויאלית הנובעת מההגדרה

$$T\{a \cdot s_1(t) + b \cdot s_2(t)\} = a \cdot T\{s_1(t)\} + b \cdot T\{s_2(t)\}$$

גם אינטגרל היא פעולת סיכום, וככזו, גם היא יכולה להשתחלף עם המערכת הליניארית. לכן, עבור אות $s_2(t)$ שהתקבל ממוצאה של המערכת (זוהי מערכת ליניארית, קבועה בזמן)

$$s_2(t) = \int_{-1}^1 s_1(t-x) dx$$

אם בידינו מערכת ליניארית שנייה T אשר תפעל על $s_2(t)$, נקבל כי

$$T\{s_2(t)\} = T\left\{\int_{-1}^1 s_1(t-x)dx\right\} = \int_{-1}^1 T\{s_1(t-x)\}dx$$

ופירוש התוצאה הוא שניתן להפעיל את T קודם ואחר-כך את האינטגרל. אנו נפגוש תכונה זו שוב בצורתה הרחבה יותר בהמשך.

פונקצית ההלם

לפני שנתחיל להרחיב על מערכות ליניאריות וקבועות בזמן, נדבר על אות מיוחד שיעניין אותנו לצורך ניתוח מערכות אלו – פונקצית ההלם. פונקציה זו מוזרה מעט וחולנית מבחינה מתמטית. לכן דרך הגדרתה עקיפה, ואנו נראה שתי דרכים לאפיונה. נגדיר את האות המלבני הבא

$$s_T(t) = \begin{cases} 1/2T & -T < t < T \\ 0 & \text{אחרת} \end{cases}$$

נסתכל על אות זה לערכי T משתנים ההולכים וקטנים ושואפים לאפס. נקבל כי המלבן גדל לגובה אך מתכווץ ברוחבו, ובכל מקרה, לכל בחירה של T שטחו יהיה 1. כלומר,

$$\forall T, \int_{-\infty}^{\infty} s_T(t)dt = 1$$

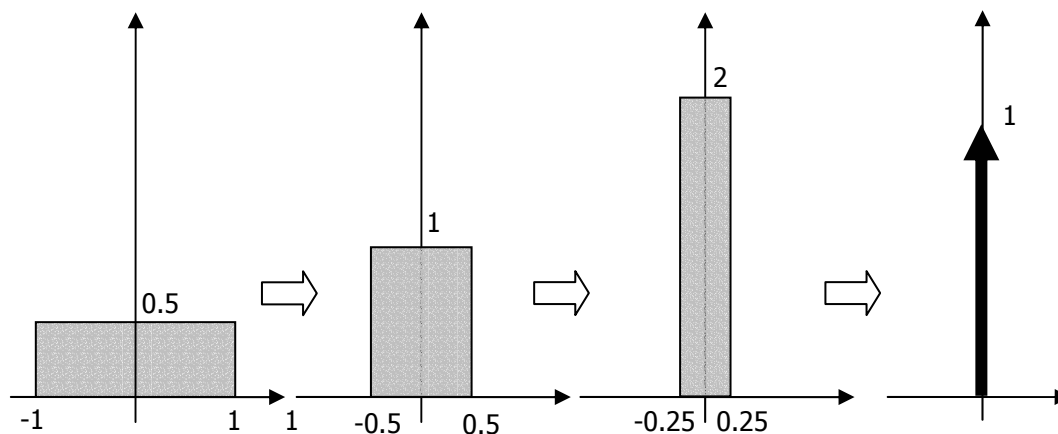
בגבול נקבל פונקציה שרוחבה אפס וגובהה אינסוף וזוהי פונקצית ההלם, המסומנת ב- $\delta(t)$:

$$\delta(t) = \lim_{T \rightarrow 0} s_T(t) = \lim_{T \rightarrow 0} \begin{cases} 1/2T & -T < t < T \\ 0 & \text{אחרת} \end{cases}$$

נוכל לפיכך לומר כי השטח מתחת לפונקצית ההלם גם הוא 1, כלומר,

$$\int_{-\infty}^{\infty} \delta(t)dt = 1$$

בציור נתאר את אות ההלם כחץ אנכי ונציין את שטחו במקום גובהו, כמו בציור הבא:



בניית פונקצית ההלם ע"י סדרת פונקציות מלבניות

מהדיעה ש- $\int_{-\infty}^{\infty} \delta(t) dt$ ברור כי

$$a \cdot \int_{-\infty}^{\infty} \delta(t) dt = \int_{-\infty}^{\infty} [a \cdot \delta(t)] dt = a$$

כלומר, ניתן לכפול פונקציה זו בקבוע a ולקבל פונקציה חדשה בעלת ערך אינסוף בראשית, אך השטח מתחתיה גדול פי a .

נניח כי בידינו שני אותות, $s(t)$ כלשהו ופונקצית ההלם, ורצוננו לחשב את השטח מתחת מכפלתם. שטח זה יהיה

$$\int_{-\infty}^{\infty} \delta(t) s(t) dt = s(0)$$

כיוון שערכי האות בכל מקום מלבד בראשית אינם רלוונטיים כי הם מוכפלים באפס. במקום האפס $s(0)$ מכפיל את שיא ההלם, וכך משפיע את השטח הכולל, ממש כמו שהקבוע a עשה זאת קודם. באופן דומה, התוצאה הבאה

$$\int_{-\infty}^{\infty} \delta(t) s(t+a) dt = \int_{-\infty}^{\infty} \delta(z-a) s(z) dz = s(a)$$

קובעת כי ערך האות $s(t)$ רלוונטי רק במקום בו שיא ההלם פוגש את ערך הפונקציה, ובמקרה זה ב- a . האגף המרכזי במשוואה מתקבל משינוי משתנה האינטגרציה. אם נכתוב את המשוואה שקיבלנו בצורה שונה מעט, נאמר

$$\int_{-\infty}^{\infty} \delta(x) s(t+x) dx = \int_{-\infty}^{\infty} \delta(t-x) s(x) dx = s(t)$$

משמעות משוואה זו שהאות $s(t)$ יכול להיות מתואר כצירוף ליניארי של אותות הלם מוזות $\delta(t-x)$ כשהן משוקללות ע"י ערכי האות עצמו $s(x)$. לייצוג זה תהיה חשיבות רבה בבואנו לנתח מערכות הפועלות על אותות.

למרות שפונקצית ההלם חולה, הפונקציה הקדומה לה (האינטגרל שלה) מוגדרת היטב והינה פונקצית מדרגה פשוטה

$$\int_{-\infty}^t \delta(x) dx = \mu(t) = \begin{cases} 1 & t \geq 0 \\ 0 & \text{אחרת} \end{cases}$$

לא נוכל ברוח דומה לומר מה הנגזרת של פונקצית ההלם, אך נוכל להראות את הקשר הבא

$$\int_{-\infty}^{\infty} \frac{d\delta(x)}{dx} s(x) dx = [\delta(x)s(x)]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \frac{ds(x)}{dx} \delta(x) dx = -s'(0)$$

כשהשתמשנו בתכונת האינטגרציה בחלקים.

התגובה להלם

נחזור כעת למערכות ונניח כי בידינו מערכת ליניארית וקבועה בזמן. אם מזין למערכת זו הלם $\delta(t)$, נקבל מוצא אשר ייקרא התגובה להלם ויסומן $h(t)$, כלומר $T\{\delta(t)\} = h(t)$.

לדוגמה, עבור המערכת $g(t) = s(t) + s(t-1)$ תגובתה להלם פשוטה למדי ומתקבלת פשוט ע"י הצבה $s(t) = \delta(t)$ הנותנת כי $h(t) = \delta(t) + \delta(t-1)$. כדוגמה שנייה, מתקבל כי עבור המערכת

$$g(t) = \int_{-1}^2 s(t-x) dx$$

התגובה להלם היא

$$h(t) = \int_{-1}^2 \delta(t-x) dx \stackrel{\substack{\uparrow \\ t-x=z}}{=} \int_{t+1}^{t-2} \delta(z) dz = \int_{t-2}^{t+1} \delta(z) dz = \begin{cases} 0 & t > 2 \\ 0 & t < -1 \\ 1 & -1 \leq t \leq 2 \end{cases}$$

המקרה הראשון מתייחס לגבול תחתון חיובי ולכן אנרגיית ההלם לא מוכלת בתחום האינטגרציה. באופן דומה, המקרה השני מתייחס לגבול עליון שלילי.

נניח כעת כי הזנו אות כלשהו אחר $s(t)$ למערכת T . מהו הקשר בין מוצא המערכת כעת והתגובה להלם $h(t)$? לצורך בניית התשובה נשתמש בתכונה שהראינו קודם לכן על היכולת לפרק אות לצירוף ליניארי של פונקציות הלם

$$\int_{-\infty}^{\infty} \delta(t-x)s(x)dx = s(t)$$

לכן,

$$\begin{aligned} g(t) = T\{s(t)\} &= T\left\{\int_{-\infty}^{\infty} s(x)\delta(t-x)dx\right\} = \\ &= \int_{-\infty}^{\infty} T\{\delta(t-x)\}s(x)dx = \int_{-\infty}^{\infty} h(t-x)s(x)dx \end{aligned}$$

כשהכנסנו את פעולת המערכת לתוך האינטגרל עשינו שימוש בתכונת הליניאריות של המערכת שמשמעה כי מערכת ליניארית על סכום של איברים שקולה להפעלת המערכת על כל אחד מאיברים אלה וסיכום המוצאים. ההחלפה $T\{\delta(t-x)\} = h(t-x)$ מותרת ונכונה בשל הנחתנו כי המערכת T קבועה בזמן. המסקנה היא שלמערכות ליניאריות וקבועות בזמן, ידיעת התגובה להלם פירושה ידיעת המוצא של המערכת לכל כניסה אפשרית לפי הקשר

$$g(t) = \int_{-\infty}^{\infty} h(t-x)s(x)dx = \int_{-\infty}^{\infty} h(x)s(t-x)dx$$

נאמר יותר מזאת – כל מערכת ליניארית ניתנת לאפיון ע"י תגובתה להלם. מסתבר כי תיתכנה שתי מערכות ליניאריות שונות אשר להן אותה תגובה להלם, אך אנו נתעלם מבעיה זו ונניח התאמה חד-חד-ערכית בין שני המושגים.

קונבולוציה

המשוואה שנכתבה כקשר בין תגובת ההלם לאות הכניסה ידועה בשם "קונבולוציה". אנו מסמנים אותה ע"י הפעולה $\otimes$,

$$g(t) = \int_{-\infty}^{\infty} h(t-x)s(x)dx = \int_{-\infty}^{\infty} h(x)s(t-x)dx = h(t) \otimes s(t)$$

נמחיש כיצד מחושבת הקונבולוציה בפועל באופן גרפי. אם נתייחס לביטוי

$$g(t) = \int_{-\infty}^{\infty} h(t-x)s(x)dx = h(t) \otimes s(t)$$

הרי שעלינו לבצע השלבים הבאים:

- אנו נתייחס לאינטגרל ע"י ציור האותות מעל הציר x .
- נצייר את $s(x)$ על מערכת צירים זו.
- יש לקחת את התגובה להלם $h(x)$ ולהפוך אותה בתמונת מראה לקבלת $h(-x)$. לאחר מכן יש להזיזה בשיעור t קדימה לקבלת $h(t-x)$.
- יש לכפול את $h(t-x)$ באות $s(x)$.
- יש לחשב האינטגרל מתחת שטח המכפלה.
- ע"י הסעת h מעל s (לסריקת ערכי הזמן t) נקבל את המוצא בכל נקודת זמן.

נראה דוגמה לשם המחשת התהליך ונבנה אותה עבור המערכת שראינו כבר:

$$h(t) = \int_{-1}^2 \delta(t-x) dx = \begin{cases} 1 & -1 < t < 2 \\ 0 & \text{אחרת} \end{cases}$$

נניח כי אות הכניסה שלנו הוא

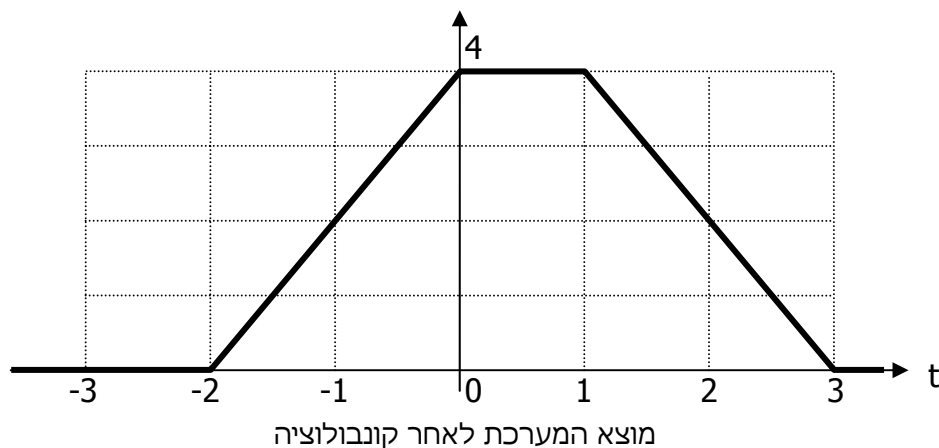
$$s(t) = \begin{cases} 2 & -1 < t < 1 \\ 0 & \text{אחרת} \end{cases}$$

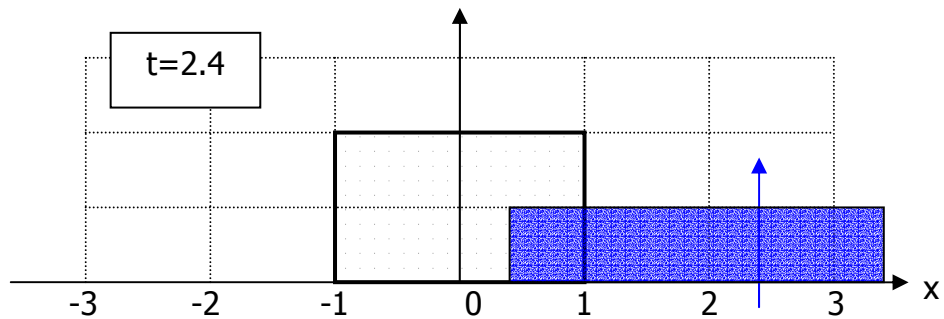
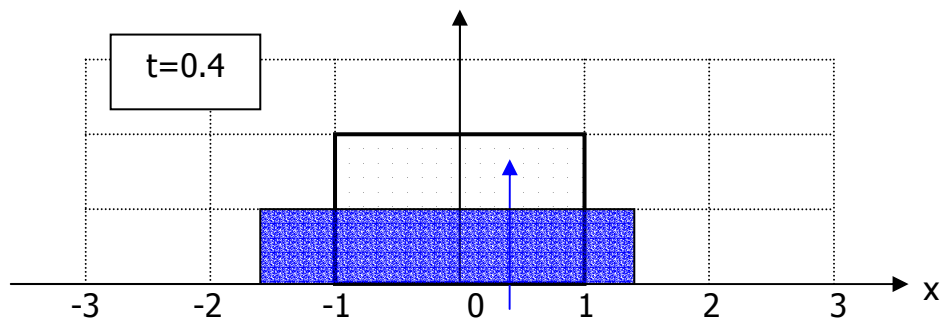
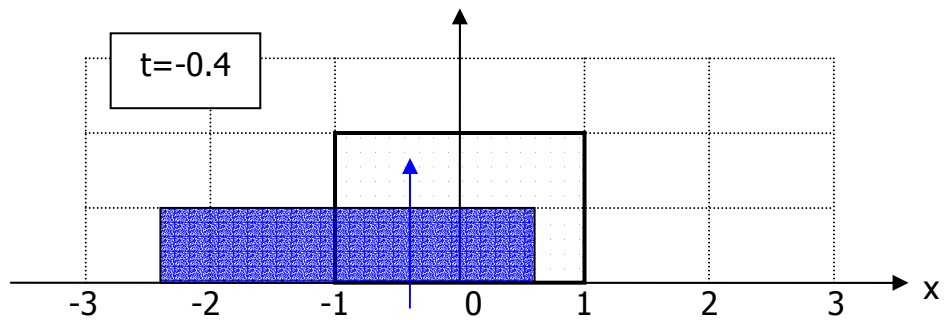
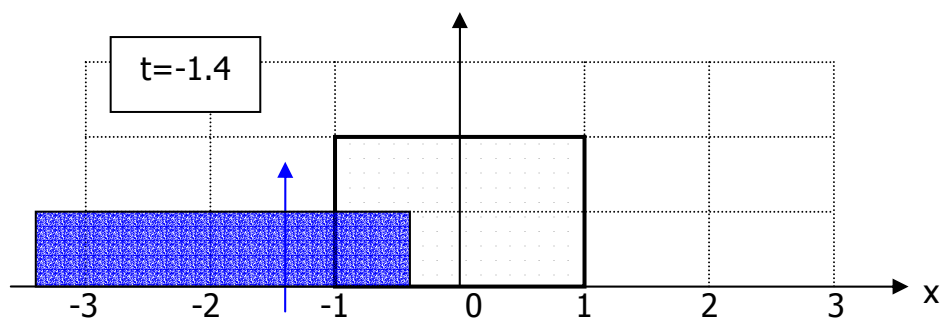
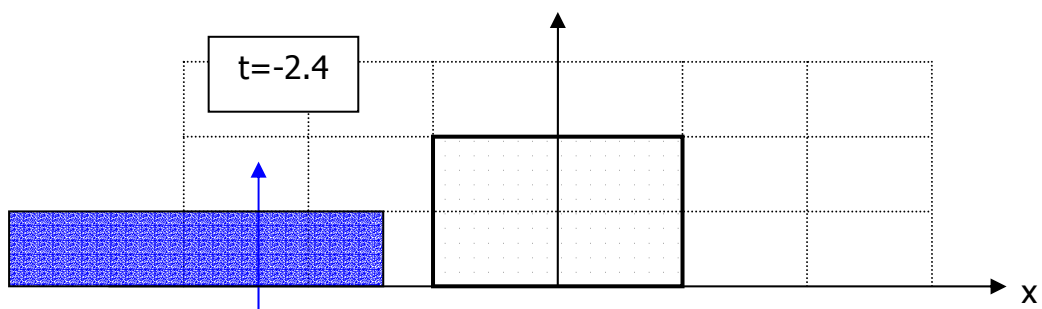
נציע פתרון אנליטי ופתרון גרפי של הבעיה. נתחיל ברישום אנליטי של

הקונבולוציה:

$$g(t) = \int_{-\infty}^{\infty} h(t-x)s(x)dx \underset{\substack{\uparrow \\ \text{limits} \\ \text{of } s(x)}}{=} 2 \int_{-1}^1 h(t-x)dx \underset{\substack{\uparrow \\ t-x=z}}{=} 2 \int_{t-1}^{t+1} h(z)dz$$

מביטוי זה פשוט יחסית לחשב את $g(t)$ – זהו השטח מתחת תגובת ההלם באינטרוול סימטרי סביב t ברוחב 2, ולכן המוצא ייראה כך:





קונבולוציה ע"י הסעת פונקציה אחת על-גבי האחרת (משבצת=1)

בהתייחס לציור בעמוד הקודם, ברור כי עבור $t < -2$ המוצא הוא אפס. כך גם הדבר עבור $t > 3$. ערכי השיא במוצא מתקבלים עבור מצב של חפיפה מלאה בין שתי הפונקציות וזה קורה עבור $0 < t < 1$, ואז הערך הוא 4. בתחום $-2 < t < 0$ מטפס ערך הקונבולוציה ליניארית עם הגדלה של החפיפה, מאפס עד 4. באופן דומה, בתחום $1 < t < 3$ דועך המוצא מ-4 לאפס.

נתאר בקצרה מספר תכונות כלליות של קונבולוציה ללא הוכחתן:

- קומוטטיביות - $s(t) \otimes g(t) = g(t) \otimes s(t)$.
- אסוציאטיביות - $s(t) \otimes (g(t) \otimes f(t)) = (s(t) \otimes g(t)) \otimes f(t)$.
- דיסטריביוטיות - $s(t) \otimes [g(t) + f(t)] = s(t) \otimes g(t) + s(t) \otimes f(t)$.
- מכפלה בסקלר - $a \cdot [s(t) \otimes g(t)] = [as(t)] \otimes g(t) = [ag(t)] \otimes s(t)$.
- הזזה - $s(t) \otimes g(t - a) = s(t - a) \otimes g(t)$.
- גזירה - $[s(t) \otimes g(t)]' = [s'(t)] \otimes g(t) = [g'(t)] \otimes s(t)$.

תכונות מערכות ליניאריות וקבועות בזמן

תכונה בסיסית של מערכות ליניאריות קבועות בזמן היא היותן משתחלפות. אנו נפגוש תכונה זו במספר צורות, ולמעשה, ענייננו במערכות מסוג זה נובע בין היתר מתכונה חשובה זו. למעשה, מתכונות הקונבולוציה קל לראות כי התכונה הנ"ל מתקיימת כיוון ש:

$$h_2(t) \otimes (h_1(t) \otimes s(t)) = (h_1(t) \otimes h_2(t)) \otimes s(t) = (h_2(t) \otimes h_1(t)) \otimes s(t)$$

כשעשינו שימוש בתכונות האסוציאטיביות והקומוטטיביות.

תכונה שנייה בעלת חשיבות היא היכולת להבחין אם מערכת סיבתית עפ"י תגובתה להלם. אם התגובה להלם היא אפס בזמנים שליליים, בהכרח מתקיים כי המערכת סיבתית. זאת כיוון שעבור תגובה להלם $h(t)$ המתאפסת לזמנים שליליים מתקבל כי

$$g(t) = h(t) \otimes s(t) = \int_{-\infty}^{\infty} h(x)s(t-x)dx \quad \underset{\substack{\uparrow \\ t < 0: h(x)=0}}{=} \int_0^{\infty} h(x)s(t-x)dx$$

כלומר גבולות האינטגרציה משתנים. לכן מוצא המערכת בזמן t תלוי אך ורק בערכי הכניסה בזמנים $t-x$ עבור x חיובי, דהיינו – סיבתיות.

תכונה שלישית נוגעת ביציבות המערכת במובן BIBO. אם נזין למערכת אות חסום $\forall t, |s(t)| \leq M$, יתקבל כי מערכת ליניארית וקבועה בזמן תהיה יציבה BIBO אם מתקיים כי

$$\int_{-\infty}^{\infty} |h(x)| dx < \infty$$

קל לראות זאת,

$$g(t) = h(t) \otimes s(t) = \int_{-\infty}^{\infty} h(x)s(t-x)dx$$

$$\begin{aligned} \Rightarrow |g(t)| &= \left| \int_{-\infty}^{\infty} h(x)s(t-x)dx \right| \leq \\ &\leq \int_{-\infty}^{\infty} |h(x)| \cdot |s(t-x)| dx \leq M \int_{-\infty}^{\infty} |h(x)| dx < \infty \end{aligned}$$

עיבוד תמונות בעזרת מחשב (236327)

הרצאה מס' 3

אותות ומערכות 2 – התמרת הפורייה האינטגרלית

אנו עוסקים כעת רק במערכות ליניאריות וקבועות בזמן. קיימת משפחה של אותות $s(t)$ מעניינים אשר מצליחים לעבור מערכות מסוג זה "ללא פגע", כלומר מוצא המערכת נראה כמו $s(t)$ למעט שינוי בגודל

$$g(t) = T\{s(t)\} = a \cdot s(t)$$

או נראה כי אותות אלו מאפשרים ניתוח מאוד יעיל של מערכות ליניאריות וקבועות בזמן. לצורך דיון באות זה נחוצה לנו תזכורת בנושא מספרים מרוכבים.

מספרים מרוכבים

מהו השורש של -1 ? במסגרת המספרים הממשיים אין מענה לשאלה זו. הרעיון של מספרים מרוכבים בא להרחיב את שדה המספרים על מנת לאפשר שורש גם למספר שלילי. נסמן $j = \sqrt{-1}$, ונגדיר שדה מספרים חדש דו-מימדי (בניגוד לחד מימד במספרים הממשיים) שבו כל מספר יהיה מהמבנה $c = x + jy$ כאשר x, y מספרים ממשיים. למעשה, כל מספר מרוכב (קומפלקסי) c ניתן לתיאור כנקודה במישור, בניגוד למספרים ממשיים הניתנים לתיאור כנקודה על ישר אינסופי.

בהינתן מספר מרוכב, $c = x + jy$, נאמר כי החלק הממשי בו הוא x ונסמן $\text{Re}\{c\} = x$, ובאופן דומה, החלק המדומה הוא y ונסמן $\text{Im}\{c\} = y$. בהינתן מספר מרוכב, יש להגדיר עליו את פעולות היסוד – חיבור, חיסור, כפל וחילוק. אנו נניח כי x, y, z, w הם מספרים ממשיים, ופעולות אלו מוגדרות ע"י:

- חיבור – $(x + jy) + (z + jw) = (x + z) + j(y + w)$
- חיסור – $(x + jy) - (z + jw) = (x - z) + j(y - w)$
- כפל – $(x + jy) \cdot (z + jw) = (xz - yw) + j(yz + xw)$

• חילוק -
$$(x + jy)/(z + jw) = \frac{(x + jy)(z - jw)}{(z + jw)(z - jw)} = \left(\frac{xz + yw}{z^2 + w^2} \right) + j \left(\frac{yz - wx}{z^2 + w^2} \right)$$

לצורך הגדרת פעולת החילוק כפלנו מונה ומכנה במספר המרוכב הצמוד ל- $(z + jw)$ – המספר הצמוד (conjugate) מוגדר כ- $(z - jw)$, ומסומן כ- $\overline{z + jw}$.

עבור מספרים מרוכבים ניתן לשנות את ייצוגם ולהשתמש בעובדה שמספר מרוכב $(x + jy)$ אינו אלא נקודה במישור. ככזו, נקודה זו מרוחקת מהראשית בשיעור של $|c| = \sqrt{x^2 + y^2}$, זוויתה מהציר האופקי היא $\theta = \text{atan}(y/x)$ (יש לשים לב לכך שפעולת atan כזו תחזיר ערך באינטרוול $(-0.5\pi, 0.5\pi]$, ובכך יש אובדן מידע – יש להתייחס לזווית בתחום $(0, 2\pi]$. ב-Matlab אבחנה זו קשורה לשוני בין הפקודות atan ו- atan2). שני מספרים אלו ימוזגו למספר מרוכב ע"י הרישום

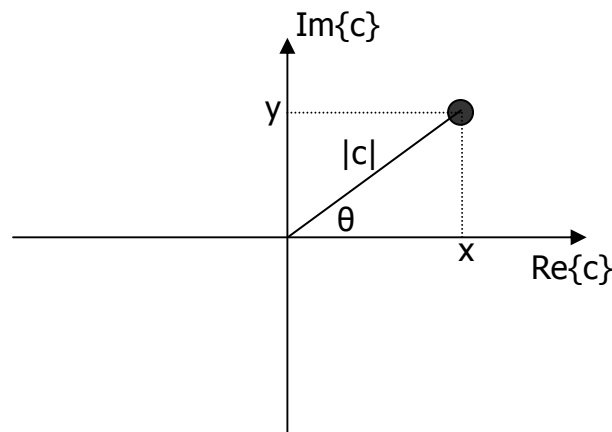
$$c = x + jy = |c| \cdot e^{j\theta} = |c| \cdot (\cos \theta + j \sin \theta)$$

נשים לב כי ברישום הנ"ל עשינו שימוש בעובדה שבהינתן הזווית והמרחק מהראשית, ניתן לחלץ את המרכיב הממשי והמדומה של המספר המרוכב. לכן הגדרנו

$$e^{j\theta} = \cos \theta + j \sin \theta$$

בעזרת הגדרה תחליפית זו של מספר מרוכב, כפל וחילוק הופכים פשוטים יותר כיוון ש:

$$c_1 = |c_1| \cdot e^{j\theta_1}, \quad c_2 = |c_2| \cdot e^{j\theta_2} \Rightarrow \begin{aligned} c_1 \cdot c_2 &= |c_1| |c_2| \cdot e^{j(\theta_1 + \theta_2)} \\ c_1 / c_2 &= |c_1| / |c_2| \cdot e^{j(\theta_1 - \theta_2)} \end{aligned}$$



ייצוג של מספר מרוכב

למעשה, גם פעולות מורכבות יותר כמו $\exp$ ו- $\log$ ניתנות להגדרה ומימוש:

$$c = x + jy = |c| \cdot e^{j\theta} \Rightarrow \begin{aligned} \log(c) &= \log(|c|) + j\theta \\ \exp(c) &= \exp(x + jy) = \exp(x)(\cos y + j \sin y) \end{aligned}$$

הגדרת התמרת הפורייה

נחזור להתחלה – אמרנו כי קיימת משפחה של אותות שעוברת דרך מערכות ליניאריות "ללא פגע". משפחה זו היא משפחת האותות

$$s(t) = \exp\{j2\pi f t\} = \cos(2\pi f t) + j \sin(2\pi f t)$$

זהו אות מרוכב הרמוני (מחזורי) בעל תדר f – כלומר ביחידת זמן אחת הוא משלים f מחזורים שלמים. נזין אות זה למערכת ליניארית וקבועה בזמן המאופיינת ע"י תגובתה להלם $h(t)$ ונקבל

$$\begin{aligned} g(t) &= h(t) \otimes s(t) = \int_{-\infty}^{\infty} h(x)s(t-x)dx = \\ &= \int_{-\infty}^{\infty} h(x)\exp\{j2\pi f(t-x)\}dx = \\ &= \exp\{j2\pi f t\} \int_{-\infty}^{\infty} h(x)\exp\{-j2\pi f x\}dx = H(f) \cdot \exp\{j2\pi f t\} \end{aligned}$$

קיבלנו כי מוצא המערכת גם הוא אות מרוכב בתדר f וכל שהשתנה הוא שהוא כעת מוכפל במספר קבוע התלוי בשני גורמים - תגובת ההלם והתדר f . התקבל ביטוי מעניין המכפיל את אות הכניסה – ביטוי אינטגרלי זה לוקח את התגובה להלם שהינה פונקציה של הזמן, וממיר אותה לפונקציה חדשה שהיא פונקציה של התדר.

ביטוי זה מוכר בשם "התמרת פורייה", והדרך בה הצגנו אותה כאן הייתה גורמת לפורייה להתהפך בקברו! יוסף פורייה הוא מתמטיקאי צרפתי בן המאה השמונה-עשרה אשר פיתח התמרה זו לצד תוצאות מתמטיות רבות אחרות. להתמרה זו חשיבות רבה בעיבוד אותות, בשל היותה מקילה לניתוח ואפיון של מערכות.

נסכם אם כך: בהינתן מערכת ליניארית וקבועה בזמן בעלת התגובה להלם $h(t)$, התמרת הפורייה של התגובה להלם המסומנת $H(f)$ היא פונקציה של התדר המוגדרת ע"י

$$H(f) = F\{h(t)\} = \int_{-\infty}^{\infty} h(t) \exp\{-j2\pi ft\} dt$$

לפעמים בספרות הגדרה זו ניתנת עם מקדם אך אנו נתעלם מכך כאן.



יוסף פורייה

שאלה ראשונה שיש לשאול היא – האם הפונקציה $H(f)$ מהווה ייצוג מלא של המערכת, ממש כשם ש- $h(t)$ הינה? במילים אחרות, האם התמרה זו מאבדת מידע? התשובה היא שאין כל אבדן מידע והדרך להראות זאת היא ע"י תהליך מתמטי מוגדר היטב של התמרה הפוכה אשר יחזיר מ- $H(f)$ ל- $h(t)$.

$$\begin{aligned} F^{-1}\{H(f)\} &= \int_{-\infty}^{\infty} H(f) \exp\{j2\pi ft\} df = \\ &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} h(x) \exp\{-j2\pi fx\} dx \right) \exp\{j2\pi ft\} df = \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x) \exp\{j2\pi f(t-x)\} dx df = \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(t-x) \exp\{j2\pi fx\} dx df = \\ &= \int_{-\infty}^{\infty} h(t-x) \left(\int_{-\infty}^{\infty} \exp\{j2\pi fx\} df \right) dx = \\ &= \int_{-\infty}^{\infty} h(t-x) \delta(x) dx = h(t). \end{aligned}$$

התהליך הנ"ל למעשה הראה את התמרת הפורייה ההפוכה, המעבירה אותנו מ-
 $H(f)$ ל- $h(t)$. השתמשנו בתכונה

$$\int_{-\infty}^{\infty} \exp\{j2\pi f x\} df = \delta(x)$$

כלומר, עבור כל x שאינו אפס יתקבל כי אנו סוכמים מעל אות מחזורי ולכן האינטגרל מתאפס. כאשר $x=0$ אנו סוכמים מעל אות קבוע ואז מקבלים אינסוף.

קיבלנו אם כך דרך אלטרנטיבית לתיאור מערכת ליניארית וקבועה בזמן, ע"י ההתייחסות ל- $H(f)$. מה משמעותה? על מנת להבהיר זאת, נתחיל במשמעות התמרת הפורייה על אות כלשהו $s(t)$:

$$S(f) = F\{s(t)\} = \int_{-\infty}^{\infty} s(t) \exp\{-j2\pi f t\} dt$$

מהי המשמעות של $S(f)$? למעשה אנו מתארים את אות הכניסה בדרך שונה אך ממצה. בעוד $s(t)$ אומר לנו כיצד בנוי האות ע"י תיאור ערכיו בכל נקודת זמן t , הפונקציה $S(f)$ אומרת לנו כיצד בנוי האות כאוסף של אותות הרמוניים $\exp\{j2\pi f t\}$. נראה זאת דרך דוגמא פשוטה – נניח עי האות שלנו כולל הרמוניה אחת בתדר f_0 והוא

$$s(t) = \exp\{j2\pi f_0 t\}$$

אזי התמרת הפורייה שלו היא

$$S(f) = \int_{-\infty}^{\infty} s(t) \exp\{-j2\pi f t\} dt = \int_{-\infty}^{\infty} \exp\{-j2\pi(f - f_0)t\} dt = \delta(f - f_0)$$

כלומר, מהתמרת הפורייה אנו מזהים כי מדובר באות בעל הרמוניה יחידה, ואשר תדרה הוא f_0 . באופן דומה, אם האות שלנו הוא אות סינוסואידלי רציף:

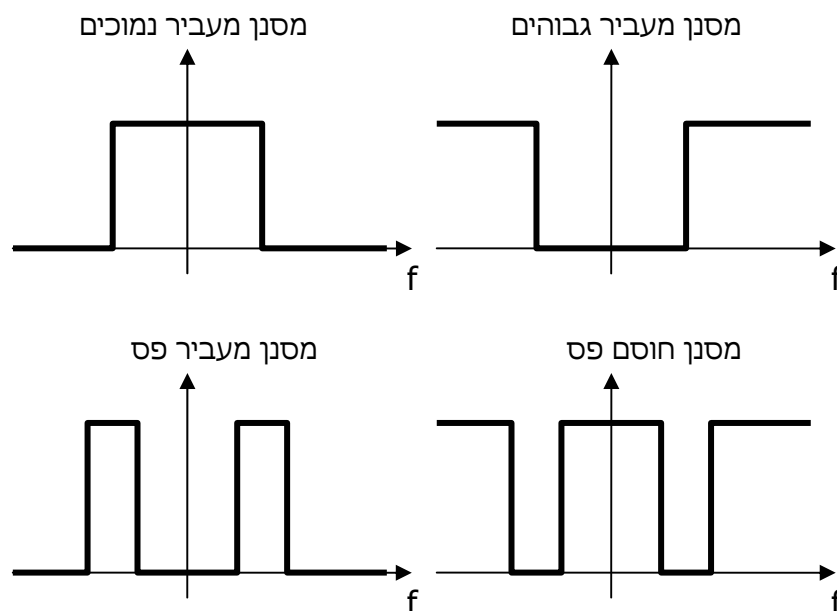
$$s(t) = \sin(2\pi f_0 t) = \frac{\exp\{j2\pi f_0 t\} - \exp\{-j2\pi f_0 t\}}{2j}$$

אזי התמרת הפורייה שלו היא

$$\begin{aligned} S(f) &= \int_{-\infty}^{\infty} s(t) \exp\{-j2\pi f t\} dt = \\ &= \int_{-\infty}^{\infty} \frac{\exp\{j2\pi(f_0 - f)t\} - \exp\{-j2\pi(f + f_0)t\}}{2j} dt = \frac{1}{2j} [\delta(f - f_0) - \delta(f + f_0)] \end{aligned}$$

כלומר, לאות זה יש שתי הרמוניות בתדרים f_0 ו- $-f_0$.

נחזור לשאלה המקורית: בהינתן מערכת עם $H(f)$ כהתמרת הפורייה של תגובתה להלם, מה משמעות של פונקציה זו? אם נזין למערכת זו אות הרמוני יחיד בתדר f_0 ראינו כבר כי מוצאה יהיה אותו אות (באותו תדר) כשהוא מוכפל ב- $H(f)$, כלומר, $H(f)$ היא תגובת התדר של המערכת הנתונה. לדוגמה, אם $H(f)$ מתאפסת לטווח מסוים של תדרים, פירוש הדבר הוא שהמערכת הנתונה לנו חוסמת תדרים אלו. מקובל לצייר את הערך המוחלט של פונקציית תגובת התדר כדי לתאר את פעולת המערכת - ראה את הציור הבא לתיאור מסננים קלאסיים.



מסננים קלאסיים ותיאורם בתדר. לדוגמה - מעביר נמוכים ידכה תדרים גבוהים.

גם לפאזה (הרי $H(f)$ מרוכב!) יש תפקיד חשוב, והיא קובעת כיצד מוזן האות במוצא. לדוגמה, אם $H(f_0) = 0.01 \cdot \exp\{jn/4\}$, פירוש הדבר שמוצא המערכת בתדר זה יונחת פי 100, ויוסט בכדי רבע מחזור קדימה, כיוון ש:

$$\begin{aligned} H(f_0) \exp\{j2\pi f_0 t\} &= 0.01 \cdot \exp\{jn/4\} \exp\{j2\pi f_0 t\} = \\ &= 0.01 [\cos(2\pi f_0 t + n/4) + j \sin(2\pi f_0 t + n/4)] \end{aligned}$$

כאשר נזין למערכת אות כלשהו $s(t)$, הרי שראינו כי ניתן לפרקו לצירוף ליניארי של אותות הרמוניים. בשל ליניאריות המערכת, כל אחד מאלה יוכפל בתגובת התדר בתדר המתאים לו, וכך ייקבע המוצא. אנו נראה מיד דרך פשוטה

לניסוח תכונה זו באופן מתמטי נקי, והיא אחת הסיבות לפופולאריות של התמרת הפורייה.

תכונות בסיסיות של התמרת הפורייה

כעת נציג עם הוכחות בסיסיות סידרה של תכונות של התמרת הפורייה. מסתבר כי בעזרת תכונות אלה ניתן לחשב התמרה של אות בעל הרכב מורכב, ע"י פירוקו.

- הגדרה - עבור האות $g(t)$ התמרת הפורייה נתונה ע"י

$$G(f) = F\{g(t)\} = \int_{-\infty}^{\infty} g(t) \exp\{-j2\pi ft\} dt$$

$$\text{וההתמרה חזרה ע"י } g(t) = F^{-1}\{G(f)\} = \int_{-\infty}^{\infty} G(f) \exp\{j2\pi ft\} df$$

- ליניאריות – התמרת הפורייה ליניארית, כלומר, עבור כל שני אותות $s(t)$ ו- $g(t)$, ושני סקלרים a ו- b , מתקיים כי

$$\begin{aligned} F\{a \cdot s(t) + b \cdot g(t)\} &= \int_{-\infty}^{\infty} (a \cdot s(t) + b \cdot g(t)) \exp\{-j2\pi ft\} dt = \\ &= a \cdot \int_{-\infty}^{\infty} s(t) \exp\{-j2\pi ft\} dt + b \cdot \int_{-\infty}^{\infty} g(t) \exp\{-j2\pi ft\} dt = \\ &= a \cdot F\{s(t)\} + b \cdot F\{g(t)\} \end{aligned}$$

תכונה זו נובעת מהיות ההתמרה מבוססת על סכימה אינטגרלית, בה ניתן להוציא מקדמים החוצה ולפרק הסכימה לכל נתח בנפרד.

- ממוצע – ממוצע האות $g(t)$ אינו אלא $G(0)$:

$$\int_{-\infty}^{\infty} g(t) dt = \int_{-\infty}^{\infty} g(t) \exp\{-j2\pi t \cdot 0\} dt = F\{g(t)\} \Big|_{f=0} = G(0)$$

- ממשיות וסימטריות – בהינתן לנו אות $g(t)$ ממשי טהור (כלומר ללא מרכיב מדומה), להתמרת הפורייה שלו מבנה סימטרי צמוד,

$$\begin{aligned} G(f) &= \int_{-\infty}^{\infty} g(t) \exp\{-j2\pi ft\} dt \\ \overline{G(f)} &= \int_{-\infty}^{\infty} g(t) \exp\{-j2\pi ft\} dt = \int_{-\infty}^{\infty} g(t) \exp\{j2\pi ft\} dt = G(-f) \end{aligned}$$

השתמשנו בתכונה של פעולת הצמוד שהיא מתחלפת עם סכימה. משמעות תוצאה זו היא שעבור אותות ממשיים לא צריך לדעת את ערכי $G(f)$ לערכי f שליליים כיוון שאלו משתמעים מערכי G בתדרים החיוביים. מסיבה זו, אגב, המסננים הקנוניים שהוצגו קודם סימטריים סביב האפס בערכם המוחלט – אנו מניחים כי תגובתם להלם ממשית.

לתוצאה זו גם אפקט הפוך – אם $g(t)$ הוא אות סימטרי צמוד,

כלומר $\overline{g(t)} = g(-t)$, יתקבל כי התמרת הפורייה ממשית, כיוון ש:

$$\begin{aligned} G(f) &= \int_{-\infty}^{\infty} g(t) \exp\{-j2\pi ft\} dt = \int_0^{\infty} g(t) \exp\{-j2\pi ft\} dt + \int_{-\infty}^0 g(t) \exp\{-j2\pi ft\} dt = \\ &= \int_0^{\infty} g(t) \exp\{-j2\pi ft\} dt + \int_0^{\infty} g(-u) \exp\{j2\pi fu\} du \end{aligned}$$

$\uparrow$
 $u=-t$

ע"י שינוי משתנה בחלק הסכימה השני נקבל

$$\begin{aligned} G(f) &= \int_0^{\infty} g(t) \exp\{-j2\pi ft\} dt + \int_0^{\infty} \overline{g(t)} \exp\{j2\pi ft\} du = \\ &= \int_0^{\infty} g(t) \exp\{-j2\pi ft\} dt + \int_0^{\infty} g(t) \exp\{-j2\pi ft\} du \end{aligned}$$

$\uparrow$
 $u=-t$

בתוצאה זו השתמשנו בעובדה שסכום של שני מספרים צמודים זה לזה הוא מספר ממשי טהור.

כמקרה אחרון מעניין, אם האות הנתון לנו $g(t)$ גם ממשי וגם סימטרי יתקבל כי $G(f)$ ממשי וסימטרי אף הוא. הוכחה לכך לא נחוצה כיוון שהיא בנויה על שתי הטענות הקודמות.

- הזזה ואפנון – אם ידוע לנו כי התמרתו של $g(t)$ היא $G(f)$, ההתמרה של אות מוזז תתקבל כשהיא מאופננת:

$$\begin{aligned} \int_{-\infty}^{\infty} g(t-d) \exp\{-j2\pi ft\} dt &= \int_{t-d=u}^{\infty} g(u) \exp\{-j2\pi f(u+d)\} du = \\ &= \exp\{-j2\pi fd\} \int_{-\infty}^{\infty} g(t) \exp\{-j2\pi ft\} dt = \exp\{-j2\pi fd\} G(f) \end{aligned}$$

משמעות האפנון הפעלת התמרת הפורייה בפונקציה הרמונית בתדר עם זמן קבוע ושווה למידת ההשהיה d .

באופן דומה, אם לקחנו את האות $g(t)$ ואפננו אותו ע"י הכפלתו ב-

$\exp\{-j2\pi f_0 t\}$, ההתמרה תהיה מוזזת -

$$\begin{aligned}\int_{-\infty}^{\infty} g(t) \exp\{-j2\pi f_0 t\} \exp\{-j2\pi f t\} dt &= \\ &= \int_{-\infty}^{\infty} g(t) \exp\{-j2\pi(f + f_0)t\} dt = G(f + f_0)\end{aligned}$$

- הגדלה – אם אות הכניסה שלנו עובר שינויי סקלת ציר זמן ההתמרה תשנה סקלה בצורה דומה והפוכה לה. בהנחה ששינוי הסקלה a הוא חיובי נקבל

$$\begin{aligned}F\{g(at)\} &= \int_{-\infty}^{\infty} g(at) \exp\{-j2\pi f t\} dt \\ &= \frac{1}{a} \int_{-\infty}^{\infty} g(u) \exp\{-j2\pi f u / a\} du = \frac{1}{a} G(f/a)\end{aligned}$$

$\uparrow$
 $at=u$

- גזירה – אם ידוע כי $g(t)$ ו- $G(f)$ הם צמד פוריירי ונניח כי ידוע כי $g(t)$ מתאפס לקראת האינסוף משני הכיוונים. אזי התמרת הפורייה של נגזרתו של $g(t)$ תהיה

$$\begin{aligned}F\left\{\frac{d}{dt} g(t)\right\} &= \int_{-\infty}^{\infty} \frac{d}{dt} g(t) \exp\{-j2\pi f t\} dt = \\ &= [g(t) \exp\{-j2\pi f t\}]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \frac{d}{dt} \exp\{-j2\pi f t\} g(t) dt = \\ &= j2\pi f \cdot \int_{-\infty}^{\infty} \exp\{-j2\pi f t\} g(t) dt = j2\pi f \cdot G(f)\end{aligned}$$

נשים לב לכך שהנחנו כי הגורם הראשון באינטגרציה בחלקים מתאפס, והנחה זו נובעת מכך שהפונקציה $g(t)$ דועכת לקראת האינסוף.

ישנן עוד תכונות חשובות ואותן נעלה בהמשך. חשוב להבין כי לא פעם בהינתן פונקציה, חישוב התמרתה לא צריך להיעשות ע"י ההגדרה ישירות. כתחליף, יש להשתמש בתכונות שתוארו קודם, ובצמדים הבסיסיים המוכרים.

נעצור לרגע דיון זה ונציג מסר דוגמאות לצמדים אופייניים לאות $g(t)$ והתמרתו $G(f)$.

ההתמרה $G(f)$	האות $g(t)$
1	$\delta(t)$
$\exp\{-j2\pi f t_0\}$ [תכונת ההזזה]	$\delta(t - t_0)$
$\delta(f)$	1
$\delta(f - f_0)$ [תכונת האפנון]	$\exp\{-j2\pi f_0 t\}$
$[\delta(f - f_0) + \delta(f + f_0)]/2$	$\cos(2\pi f_0 t)$
$[\delta(f - f_0) - \delta(f + f_0)]/2j$	$\sin(2\pi f_0 t)$
$a/(a^2 + f^2)/\pi$	$\exp(-2\pi a t)$
$\sqrt{\pi/a} \cdot \exp(-\pi^2 f^2/a)$	$\exp(-at^2)$ [גאוסיאן]
$2d \cdot \text{sinc}\{2\pi f d\}$	$\mu(t+d) - \mu(t-d)$ [פונקצית מלבן]
$(2d \cdot \text{sinc}\{2\pi f d\})^2$	$\text{tri}(t, 2d)$ [פונקצית משולש]

הצמד הקשור לפונקצית המלבן מתקבל באופן הבא:

$$\begin{aligned}
 \int_{-\infty}^{\infty} (\mu(t+d) - \mu(t-d)) \exp\{-j2\pi f t\} dt &= \int_{-d}^d \exp\{-j2\pi f t\} dt = \left[\frac{\exp\{-j2\pi f t\}}{-j2\pi f} \right]_{-d}^d = \\
 &= \frac{\exp\{-j2\pi f d\}}{-j2\pi f} - \frac{\exp\{j2\pi f d\}}{-j2\pi f} = \\
 &= \frac{\exp\{j2\pi f d\} - \exp\{-j2\pi f d\}}{j2\pi f} = \\
 &= \frac{\sin\{2\pi f d\}}{\pi f} = 2d \cdot \text{sinc}\{2\pi f d\}
 \end{aligned}$$

כשאנו מגדירים $\text{sinc}(x) = \sin(x)/x$, פונקצית המשולש מוגדרת ע"י

$$s(t) = \begin{cases} 0 & |t| > 2d \\ 2d - t & 0 \leq t \leq 2d \\ t + 2d & -2d \leq t \leq 0 \end{cases}$$

ואת התמרתה נצדיק בהמשך, כשנדון בתכונת הקונבולוציה.

תכונת פרסבל

תכונה מאוד חשובה של התמרת הפורייה היא שימור אנרגיה. תכונה זו, המוכרת בשם "משפט פרסבל", טוענת כי

$$\int_{-\infty}^{\infty} |g(t)|^2 dt = \int_{-\infty}^{\infty} |G(f)|^2 df$$

תכונה זו היא למעשה מקרה פרטי של תכונה אחרת חשובה – שימור מכפלה פנימית, אשר אותה נוכיח. נניח כי בידינו שני אותות $g(t)$ ו- $s(t)$, והתמרתם $G(f)$ ו- $S(f)$. מכפלה פנימית בין שתי פונקציות בזמן מוגדרת ע"י

$$\langle g(t), s(t) \rangle_t = \int_{-\infty}^{\infty} g(t) \overline{s(t)} dt$$

נראה כי מכפלה פנימית בתדר של פונקציות ההתמרה תוביל לאותה תוצאה:

$$\begin{aligned} \langle G(f), S(f) \rangle_f &= \int_{-\infty}^{\infty} G(f) \overline{S(f)} df = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} g(t) \exp\{-j2\pi f t\} dt \right) \left(\int_{-\infty}^{\infty} \overline{s(u)} \exp\{j2\pi f u\} du \right) df \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(t) \overline{s(u)} \exp\{j2\pi f u\} \exp\{-j2\pi f t\} du dt df \end{aligned}$$

שינוי סדר הסכימה יביא ל:

$$\begin{aligned} \langle G(f), S(f) \rangle_f &= \int_{-\infty}^{\infty} G(f) \overline{S(f)} df = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(t) \overline{s(u)} \left(\int_{-\infty}^{\infty} \exp\{j2\pi f(u-t)\} df \right) du dt = \\ &= \int_{-\infty}^{\infty} g(t) \left(\int_{-\infty}^{\infty} \overline{s(u)} \delta(u-t) du \right) dt = \int_{-\infty}^{\infty} g(t) \overline{s(t)} dt = \langle g(t), s(t) \rangle_t \end{aligned}$$

לכן, כמקרה פרטי, הצבת $s(t)=g(t)$ תיתן את תוצאת פרסבל:

$$\int_{-\infty}^{\infty} |g(t)|^2 dt = \int_{-\infty}^{\infty} |G(f)|^2 df$$

תכונות הקורלציה והקונבולוציה

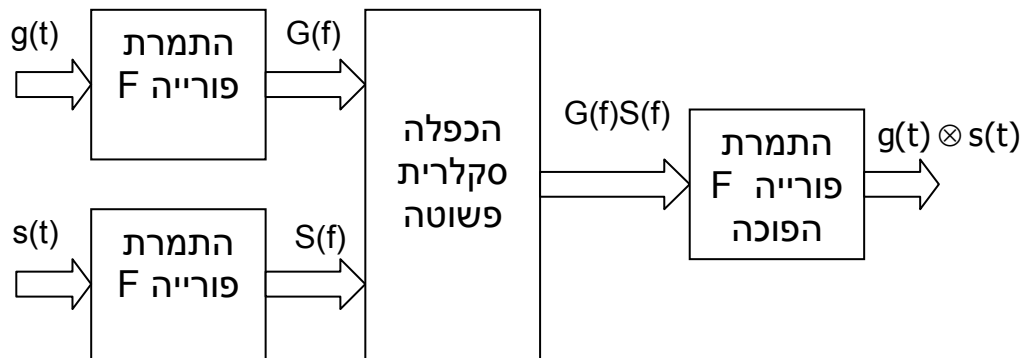
תכונה חשובה אחרת היא תכונת הקונבולוציה, וגירסה דומה לה הקרויה תכונת הקורלציה, אשר דומה לתכונת המכפלה הפנימית שהוצגה קודם. נניח כי בידינו שני אותות $g(t)$ ו- $s(t)$, והתמרתם $G(f)$ ו- $S(f)$. הגדרנו את הקונבולוציה בין שתי הפונקציות בזמן ע"י

$$g(t) \otimes s(t) = \int_{-\infty}^{\infty} g(u) s(t-u) du$$

נראה כי בתדר חישוב הקונבולוציה הופך למכפלה רגילה:

$$\begin{aligned} F\{g(t) \otimes s(t)\} &= F\left\{\int_{-\infty}^{\infty} g(u)s(t-u)du\right\} = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} g(u)s(t-u)du\right) \cdot \exp\{-j2\pi ft\}dt = \\ &= \int_{-\infty}^{\infty} g(u) \left(\int_{-\infty}^{\infty} s(t-u) \cdot \exp\{-j2\pi ft\}dt\right)du = \\ &\stackrel{\substack{\uparrow \\ t-u=z}}{=} \int_{-\infty}^{\infty} g(u) \left(\int_{-\infty}^{\infty} s(z) \cdot \exp\{-j2\pi f(z+u)\}dz\right)du = \\ &= \left(\int_{-\infty}^{\infty} g(u) \exp\{-j2\pi fu\}du\right) \left(\int_{-\infty}^{\infty} s(z) \cdot \exp\{-j2\pi fz\}dz\right) = G(f)S(f) \end{aligned}$$

משמעות תוצאה זו היא שבמקום לחשב קונבולוציה (עניין מורכב בד"כ), נוכל לעבור להתמרות הפורייה של שני האותות המעורבים, להכפילם זה בזה סקלרית – עניין קל – ואז לבצע התמרה הפוכה חזרה לזמן.



ביצוע קונבולוציה ע"י מעבר לתדר

דוגמה: עבור אות כניסה מלבני $s(t) = \mu(t+d) - \mu(t-d)$ המוזן למערכת בעת תגובה להלם $h(t) = \mu(t+d) - \mu(t-d)$, מוצא המערכת יהיה

$$F^{-1}\{F\{\mu(t+d) - \mu(t-d)\} \cdot F\{\mu(t+d) - \mu(t-d)\}\} = F^{-1}\left\{\left(\frac{\sin\{2\pi fd\}}{\pi f}\right)^2\right\}$$

אם נבצע את הקונבולוציה בצורה ישירה נקבל את פונקצית המשולש, כיוון ש:

$$\begin{aligned}
& \int_{-\infty}^{\infty} (\mu(u+d) - \mu(u-d))(\mu(t-u+d) - \mu(t-u-d))du = \\
& = \int_{-d}^d (\mu(t-u+d) - \mu(t-u-d))du = \int_{-d}^d \mu(t-u+d)du - \int_{-d}^d \mu(t-u-d)du = \\
& \stackrel{\substack{\uparrow \\ z=t-u+d}}{=} \int_t^{t+2d} \mu(z)dz \stackrel{\substack{\uparrow \\ z=t-u-d}}{-} \int_{t-2d}^t \mu(z)dz
\end{aligned}$$

נשתמש בתכונה שאינטגרל על-פני פונקצית מדרגה מקיים

$$\int_a^b \mu(z)dz = \begin{cases} b-a & b > a > 0 \\ b & b > 0 > a \\ 0 & 0 > b > a \end{cases}$$

ונקבל את פונקצית המשולש, כי

$$\begin{aligned}
& \int_t^{t+2d} \mu(z)dz \stackrel{\substack{\uparrow \\ z=t-u-d}}{-} \int_{t-2d}^t \mu(z)dz = \\
& = \begin{cases} 2d & t > 0 \\ t+2d & 0 \geq t > -2d \\ 0 & -2d \geq t \end{cases} - \begin{cases} 2d & t > 2d \\ t & 2d \geq t > 0 \\ 0 & 0 \geq t \end{cases} = \\
& = \begin{cases} 2d-2d=0 & t > 2d \\ 2d-t & 2d \geq t > 0 \\ t+2d & 0 \geq t > -2d \\ 0 & -2d > t \end{cases}
\end{aligned}$$

תכונה דומה היא תכונת הקורלציה. פעולת הקורלציה בין אותות מוגדרת

ע"י האינטגרל:

$$\text{Corr}\{g(t), s(t)\} = \int_{-\infty}^{\infty} g(u) \overline{s(u-t)} du$$

חישוב קורלציה חשוב בעיבוד אותות ותמונות – הקורלציה מראה כמה דומים שני אותות. התמרת תדר מובילה ל:

$$\begin{aligned}
F\{\text{Corr}\{g(t), s(t)\}\} &= F\left\{\int_{-\infty}^{\infty} g(u) \overline{s(u-t)} du\right\} = \\
&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} g(u) \overline{s(u-t)} du\right) \cdot \exp\{-j2\pi ft\} dt = \\
&= \int_{-\infty}^{\infty} g(u) \left(\int_{-\infty}^{\infty} \overline{s(u-t)} \cdot \exp\{-j2\pi ft\} dt\right) du = \\
&= \int_{-\infty}^{\infty} g(u) \exp\{-j2\pi fu\} \left(\int_{-\infty}^{\infty} \overline{s(u-t)} \cdot \exp\{-j2\pi f(t-u)\} dt\right) du = \\
&= \left(\int_{-\infty}^{\infty} g(u) \exp\{-j2\pi fu\} du\right) \left(\int_{-\infty}^{\infty} \overline{s(t)} \cdot \exp\{j2\pi ft\} dt\right) = \\
&= \left(\int_{-\infty}^{\infty} g(u) \exp\{-j2\pi fu\} du\right) \overline{\left(\int_{-\infty}^{\infty} s(t) \cdot \exp\{-j2\pi ft\} dt\right)} = G(f) \overline{S(f)}
\end{aligned}$$

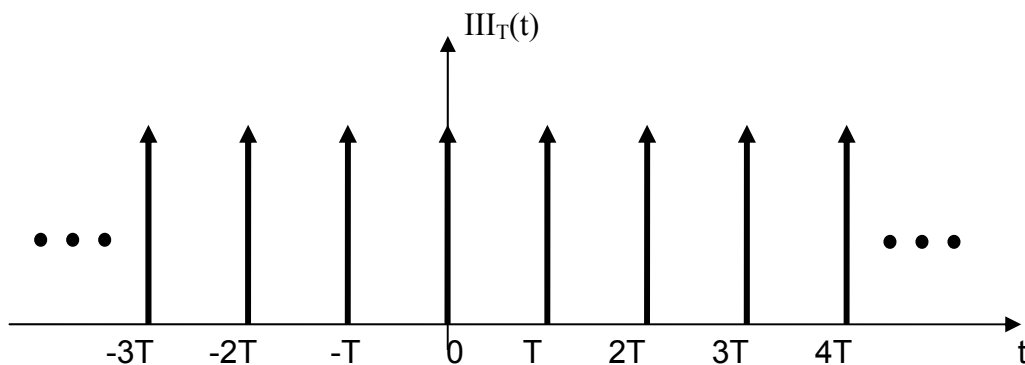
המסקנה היא שגם קורלציה יכולה להיעשות בתדר במכפלה פשוטה.

אותות מחזוריים ופונקצית שה (Shah)

פונקציה מעניינת נוספת שתעניין אותנו היא "רכבת הלמים", או כפי שמתמטיקאים קוראים לה, פונקצית שה (Shah). פונקציה זו נתונה כטור

$$III_T(t) = \sum_{k=-\infty}^{\infty} \delta(t - kT)$$

והיא נראית במתואר בציור הבא:



פונקצית שה (Shah) – רכבת הלמים. השם "Shah" בא מהדמיון של הסימון לאות "ש" בעברית, וגרסאות דומות לה בקירילית ובשפת חרטומים מצריים.

התמרת הפורייה של פונקציה זו הינה גם היא רכבת הלמים. שימוש בהגדרה נותן

$$\begin{aligned} F\{III_T(t)\} &= F\left\{\sum_{k=-\infty}^{\infty} \delta(t - kT)\right\} = \int_{-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \delta(t - kT) \exp\{-j2\pi f t\} dt = \\ &= \sum_{k=-\infty}^{\infty} \left(\int_{-\infty}^{\infty} \delta(t - kT) \exp\{-j2\pi f t\} dt \right) = \\ &= \sum_{k=-\infty}^{\infty} \exp\{-j2\pi f kT\} \end{aligned}$$

עבור תדרים המהווים מכפלה שלמה של $1/T$ סכימה זו תצטבר לאינסוף כיוון שכל אחד מהאקספוננטים יהיה 1, בעוד שלכל בחירת תדר אחרת סכימה זו תתאפס. לכן,

$$F\{III_T(t)\} = F\left\{\sum_{k=-\infty}^{\infty} \delta(t - kT)\right\} = \sum_{k=-\infty}^{\infty} \exp\{-j2\pi f kT\} = \frac{1}{T} \sum_{n=-\infty}^{\infty} \delta\left(f - \frac{n}{T}\right)$$

עניין המקדם $1/T$ נובע מתכונת שימור האנרגיה של פרסבל. בשני המקרים יש סכימה על אינסוף הלמים, אך בתדר יש הלם כל $1/T$ יחידות תדר, בעוד שבזמן הם מופיעים כל T . לכן יש פי T^2 יותר הלמים בתדר. עם הכנסת המקדם הנ"ל סך האנרגיה בשני האותות משתווה (יש לזכור כי בפרסבל אנו מעלים את האות בריבוע, ולכן נחוץ שורש היחס הנ"ל).

תוצאה זו מעניינת מכמה סיבות – קודם כל, מעניין לדעת מיהן הפונקציות שהתמרת הפורייה לא משנה את צורתן. ראינו קודם כי גאוסיאן הוא כזה, ועתה מצאנו פונקציה אחרת בעלת אותה תכונה. מעבר לכך שנה תכונה חשובה יותר לתוצאה הנ"ל, והיא מובילה למושג טורי פורייה שאותם נציג בקצרה לשם השכלה כללית.

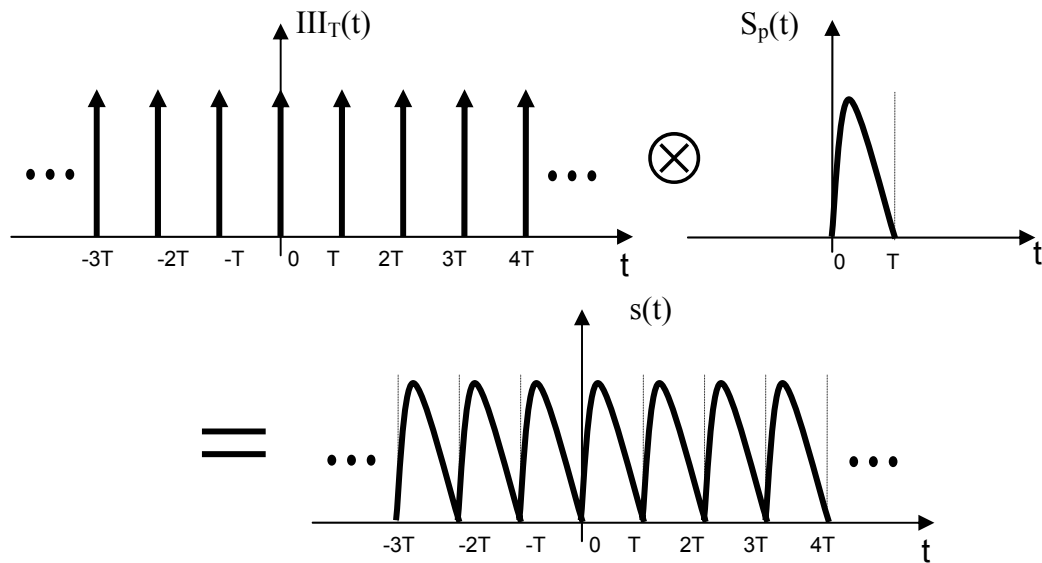
התמרת פורייה לאותות מחזוריים – טורי פורייה

כל אות מחזורי $s(t)$ במרווח T ניתן לתיאור כקונבולוציה של האות במחזור אחד בקטע $[0, T]$ עם הפונקציה $III_T(t)$. נסמן את המחזור היחיד באינטרוול $[0, T]$ ע"י $s_p(t)$ (בשביל period – מחזור יחיד). אזי אנו בעצם קובעים כי

$$s(t) = s_p(t) \otimes III_T(t)$$

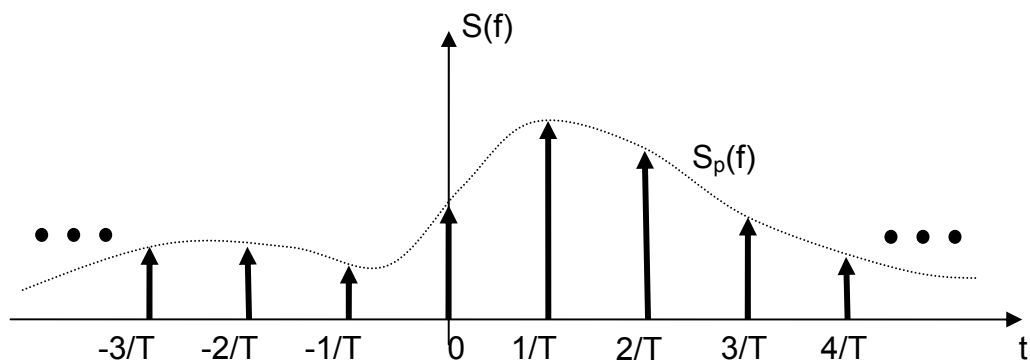
תכונה זו חשובה כיוון שכעת נוכל לקבוע את התמרת הפורייה של אות מחזורי כזה. התמרה זו נתונה בשל תכונת הקונבולוציה כמכפלה בין רכבת הלמים בתדר במרווחים $1/T$ והתמרת הפורייה של מחזור יחיד. לכן, התמרה זו תהיה

$$\begin{aligned} F\{s(t)\} &= F\{s_p(t) \otimes III_T(t)\} = F\{s_p(t)\} \cdot F\{III_T(t)\} = \\ &= S_p(f) \cdot \frac{1}{T} \sum_{n=-\infty}^{\infty} \delta\left(f - \frac{n}{T}\right) = \frac{1}{T} \sum_{n=-\infty}^{\infty} S_p\left(\frac{n}{T}\right) \cdot \delta\left(f - \frac{n}{T}\right) \end{aligned}$$



אות מחזורי ניתן לתיאור כקונבולוציה בין פונקצית שה (Shah) ומחזור יחיד.

קיבלנו כי התמרת הפורייה של אות מחזורי אינה קיימת בכל תדר, אלא רק בנקודות תדר שהינן מכפלות שלמות של $1/T$. כלומר, אות מחזורי בנוי רק מהרמוניות הקשורות באורך המחזור.



התמרת פורייה של פונקציה מחזורית – קו המתאר מתאר את התמרת הפורייה של מחזור יחיד, ודגימת ההלמים באה בשל המחזוריות.

אם נבצע התמרת פורייה הפוכה על התוצאה נקבל כמובן את אות המקור, אך נוכל גם להציגו אחרת – וזוהי הנקודה הראשית בסיפורנו:

$$\begin{aligned} s(t) = F^{-1} \left\{ \frac{1}{T} \sum_{n=-\infty}^{\infty} S_p \left(\frac{n}{T} \right) \cdot \delta \left(f - \frac{n}{T} \right) \right\} &= \frac{1}{T} \int_{-\infty}^{\infty} \sum_{n=-\infty}^{\infty} S_p \left(\frac{n}{T} \right) \cdot \delta \left(f - \frac{n}{T} \right) \exp \{ j 2 \pi f t \} df = \\ &= \frac{1}{T} \sum_{n=-\infty}^{\infty} S_p \left(\frac{n}{T} \right) \cdot \int_{-\infty}^{\infty} \delta \left(f - \frac{n}{T} \right) \exp \{ j 2 \pi f t \} df = \\ &= \frac{1}{T} \sum_{n=-\infty}^{\infty} S_p \left(\frac{n}{T} \right) \cdot \exp \left\{ \frac{j 2 \pi n t}{T} \right\} = \sum_{n=-\infty}^{\infty} c_n \cdot \exp \left\{ \frac{j 2 \pi n t}{T} \right\} \end{aligned}$$

התקבל כי אות המקור ניתן לתיאור כטור של פונקציות הרמוניות מחזוריות בתדרים n/T לכל n שלם, עם מקדמים הניתנים לחישוב ישיר ממחזור יחיד של הפונקציה, לפי

$$c_n = \frac{1}{T} S_p \left(\frac{n}{T} \right) = \frac{1}{T} \int_{-\infty}^{\infty} s_0(t) \exp \{ -j 2 \pi f t \} dt \Big|_{f=n/T} = \frac{1}{T} \int_0^T s(t) \exp \left\{ -\frac{j 2 \pi n t}{T} \right\} dt$$

לסיכום, בהינתן אות מחזורי בעל מחזור T , ניתן להציגו אלטרנטיבית כסכום אינסופי של פונקציות הרמוניות בתדרים n/T (לכל n שלם), ועם מקדמים הניתנים לחישוב מהתנהגות מחזור יחיד.

לדוגמה, עבור האות הפשוט $s(t) = \sin(2 \pi t / T)$, אנו יודעים לפרקו לשתי הרמוניות – מעניין לראות אם הגישה הרחבה יותר כאן תצליח להגיע לאותו פירוק. מקדמי טור הפורייה יהיו

$$\begin{aligned} c_n &= \frac{1}{T} S_p \left(\frac{n}{T} \right) = \frac{1}{T} \int_0^T s(t) \exp \left\{ -\frac{j 2 \pi n t}{T} \right\} dt = \\ &= \frac{1}{T} \int_0^T \sin \left(\frac{2 \pi t}{T} \right) \exp \left\{ -\frac{j 2 \pi n t}{T} \right\} dt = \\ &= \frac{1}{2jT} \left[\int_0^T \exp \left\{ \frac{j 2 \pi t}{T} \right\} \exp \left\{ -\frac{j 2 \pi n t}{T} \right\} dt - \int_0^T \exp \left\{ -\frac{j 2 \pi t}{T} \right\} \exp \left\{ -\frac{j 2 \pi n t}{T} \right\} dt \right] = \\ &= \frac{1}{2jT} \left[\int_0^T \exp \left\{ -\frac{j 2 \pi (n-1)t}{T} \right\} dt - \int_0^T \exp \left\{ -\frac{j 2 \pi (n+1)t}{T} \right\} dt \right] \end{aligned}$$

ברור שעבור כל n מלבד $(+1)$ ו- (-1) ביטוי זה מתאפס זהותית, כיוון שהוא מבצע סכימה על אות מחזורי במחזור שלם בו הממוצע האפס. עבור $n=1$ יתקבל

$$c_1 = \frac{1}{2Tj} \left[\int_0^T dt - \int_0^T \exp\left\{-\frac{j4nt}{T}\right\} dt \right] = \frac{1}{2j}$$

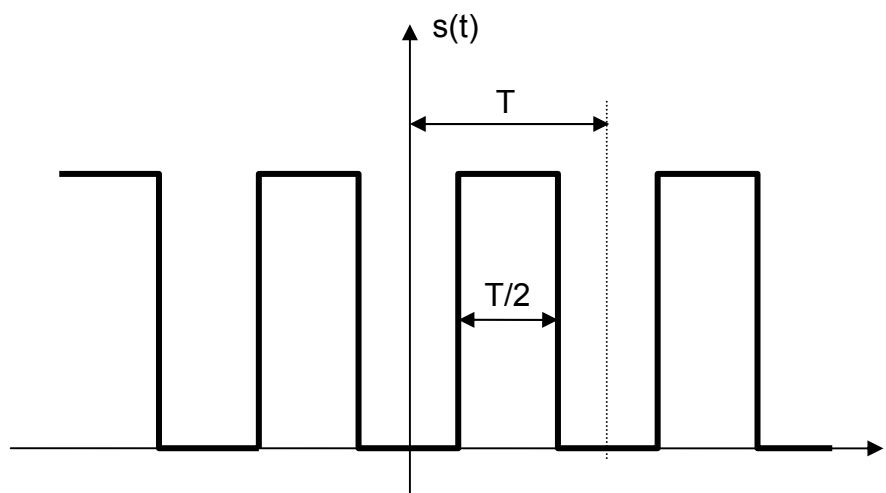
באופן דומה, עבור $n=-1$ יתקבל

$$c_{-1} = \frac{1}{2Tj} \left[\int_0^T \exp\left\{\frac{j4nt}{T}\right\} dt - \int_0^T dt \right] = \frac{-1}{2j}$$

ולכן טור הפורייה של האות $s(t) = \sin(2\pi t / T)$ הוא

$$s(t) = \sum_{n=-\infty}^{\infty} c_n \cdot \exp\left\{\frac{j2\pi n t}{T}\right\} = \frac{1}{2j} \left(\exp\left\{\frac{j2\pi t}{T}\right\} + \exp\left\{-\frac{j2\pi t}{T}\right\} \right)$$

ואמנם זהו הפירוק שלו צפינו. כדוגמה שנייה, נתייחס לאות ריבועי מחזורי המתואר בצירוף הבא.



אות ריבועי מחזורי

ניתן יהיה להציג אות זה ע"י הטור

$$s(t) = \sum_{n=-\infty}^{\infty} c_n \cdot \exp\left\{\frac{j2\pi n t}{T}\right\}$$

כמקדמי טור הפורייה יהיו

$$\begin{aligned}
c_n &= \frac{1}{T} S_p\left(\frac{n}{T}\right) = \frac{1}{T} \int_0^T s(t) \exp\left\{-\frac{j2\pi nt}{T}\right\} dt = \\
&= \frac{1}{T} \int_{T/4}^{3T/4} \exp\left\{\frac{j2\pi nt}{T}\right\} dt = \left[\frac{1}{j2\pi n} \exp\left\{\frac{j2\pi nt}{T}\right\} \right]_{T/4}^{3T/4} = \\
&= \frac{1}{j2\pi n} \left(\exp\left\{\frac{j3\pi n}{2}\right\} - \exp\left\{\frac{j\pi n}{2}\right\} \right) = \\
&= \frac{\exp\{j\pi n\}}{j2\pi n} \left(\exp\left\{\frac{j\pi n}{2}\right\} - \exp\left\{-\frac{j\pi n}{2}\right\} \right) = 2 \exp\{j\pi n\} \operatorname{sinc}\left\{\frac{\pi n}{2}\right\}
\end{aligned}$$

כלומר, האות הריבועי ניתן גם לתיאור כ:

$$s(t) = 2 \sum_{n=-\infty}^{\infty} \exp\{j\pi n\} \operatorname{sinc}\left\{\frac{\pi n}{2}\right\} \cdot \exp\left\{\frac{j2\pi nt}{T}\right\}$$

משמעות הדבר היא שבאות ריבועי יש תדרים גבוהים עד לאינסוף, אך משקלם בסכימה הכוללת דועכת כמו Sinc (כלומר כמו $1/n$ – נשים לב כי ניתן לפשט הלאה את מקדם טור הפוריה כיוון של-sin מוזנת זווית שתמיד תהיה מכפלה שלמה של 90°). בעניין הפאזה בביטוי שהתקבל, ביטוי זה הוא

$$\exp\{j\pi n\} = \cos(\pi n) + j \sin(\pi n) = \cos(\pi n) = (-1)^n$$

מעניין לציין כי מקדמי טור הפורייה סימטריים, $c_n = c_{-n}$, ולכן הגורם המדומה מתבטל ונקבל

$$s(t) = 2 \sum_{n=-\infty}^{\infty} \exp\{j\pi n\} \operatorname{sinc}\left\{\frac{\pi n}{2}\right\} \cdot \exp\left\{\frac{j2\pi nt}{T}\right\} = 2 \sum_{n=-\infty}^{\infty} \exp\{j\pi n\} \operatorname{sinc}\left\{\frac{\pi n}{2}\right\} \cdot \cos\left\{\frac{2\pi nt}{T}\right\}$$

למה חשוב לנו כל כך עניין הטיפול באותות מחזוריים? האם הם באמת

כה נפוצים? כלל וכלל לא! למרות זאת הטיפול שהוצג חשוב משתי סיבות:

1. במקרים רבים נרצה לטפל באות בעל תמך סופי (כלומר כזה שקיים רק במקטע זמן סופי). במקרים אלה מקובל להניח כי אות זה "הומשך מחזורית" כדי להתקיים בכל נקודת זמן.

2. ישנה דואליות מעניינת בין ציר הזמן וציר התדר, ואנו ראינו מספר תופעות

שנובעות מדואליות זו. נשים לב שההגדרה להתמרה קדימה ואחורה מאוד דומות, אם ההתמרה של פונקצית הלים היא קבוע, אז התמרה של קבוע היא הלים, ועוד. כמו שהתמרה של אות מחזורי היא דגומה, אנו נראה כי עבור אות דגום ההתמרה מחזורית, ולכך חשיבות עצומה בדיון באותות ע"י מחשב.

עיבוד תמונות בעזרת מחשב (236327)

הרצאה מס' 4

אותות ומערכות 3 – דגימה

משפט הדגימה ומשמעותו

נתון לנו אות $s(t)$, ורצוננו לדגום אותו במרווחים T שווים על מנת להמירו לאות דיסקרטי היכול להישמר בזיכרון מחשב (ברור כי יש לדאוג שהתמך – תחום הקיום – של האות יהיה סופי וכל ערך דגימה בו ייוצג בכמות סיביות סופית על מנת שנקבל באמת קובץ בגודל סופי של סיביות). פעולת הדגימה שנציע היא

$$s[n] = s(t) \Big|_{t=nT} = s(nT)$$

השאלה המרכזית בה נעסוק במסגרת זו היא אילו תנאים יש להציב על המרווח T כך שניתן יהיה לשחזר את $s(t)$ מתוך דגימותיו $s[n]$ ללא אובדן? לצורך מענה לשאלה זו, נעשה שימוש בתיאור מתמטי נוח יותר לפעולת הדגימה – דגימה ע"י מכפלה ב-"רכבת הלמים". נקבע כי האות נדגם ע"י המכפלה

$$\tilde{s}(t) = s(t) \cdot \text{III}_T(t) = s(t) \cdot \sum_n \delta(t - nT)$$

היתרון שבביצוע הנ"ל הוא שהאות הדגום יכול להיות מטופל בכלי שבנינו עבור הרצף, כי הוא אות רציף לכל דבר.

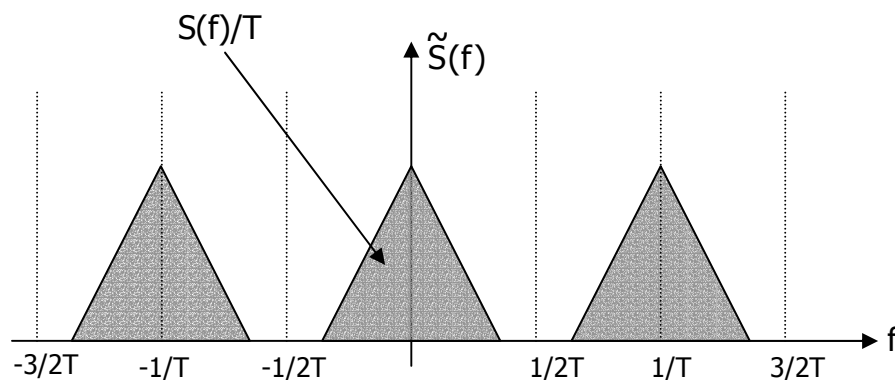
מעניין לציין כי אות זה נראה כמו פונקציות הלם הפזורות במרווחים אחידים על הציר הממשי עם גבהים משתנים. מבנה זה מוכר לנו מהדיון בטורי פורייה כהתמרת הפורייה של אות מחזורי. לכן, בשל הדואליות הקיימת בהתמרת הפורייה, נצפה שהתמרת הפורייה של אותות דגומים תהיה מחזורית, ואנו נראה שאמנם זה כך.

כיוון שהאות בנוי ממכפלה של שני אותות רציפים בעקרון, התמרת הפורייה שלו היא קונבולוציה של ההתמרות. לכן

$$\begin{aligned}\tilde{S}(f) &= F\{\tilde{s}(t)\} = F\left\{s(t) \cdot \sum_n \delta(t - nT)\right\} = F\{s(t)\} \otimes F\left\{\sum_n \delta(t - nT)\right\} = \\ &= \frac{1}{T} S(f) \otimes \sum_n \delta\left(f - \frac{n}{T}\right) = \frac{1}{T} \sum_n S\left(f - \frac{n}{T}\right)\end{aligned}$$

השתמשנו בעובדה שהתמרה על רכבת ההלמים נתונה כרכבת הלמים עם מרווחים $1/T$.

קיבלנו כי הפונקציה $S(f)$ משוכפלת שוב ושוב במרווחים קצובים $1/T$ ובהכפלה בקבוע $1/T$. הציור הבא ממחיש תופעה זו.



תופעת השכפול בדגימה של אות בזמן

ואמנם, כפי שציפינו, התמרת הפורייה של האות הדגום היא מחזורית. מציור זה גם ניכר כי כאשר $S(f)$ מתאפס מחוץ לתחום $[-1/2T, 1/2T]$, מתקבל כי השכפולים לא רומסים זה על זה. נניח כי אמנם $S(f)$ מקיים

$$S(f) = \begin{cases} S(f) & |f| \leq f_{\max} \\ 0 & \text{otherwise} \end{cases}$$

תכונה זו קרויה "אות חסום פס". אזי, על מנת שהשכפולים לא יעלו זה על זה נדרוש

$$f_{\max} < \frac{1}{2T} \Rightarrow T < \frac{1}{2f_{\max}}$$

הדרישה הזו קובעת שעלינו לדגום מספיק צפוף כדי לכלול כמות מספקת של דגימות לייצוג מלא של האות.

מדוע חשוב לנו למנוע מהשכפולים לעלות זה על זה? ובכן, בהינתן האות הדגום $\tilde{s}(t)$ (שכל דגימה בו היא פונקצית הלם), זהו אות רציף הניתן לסינון. ע"י הפעלת מסנן ריבועי מעביר נמוכים (LPF) מהצורה

$$H(f) = \begin{cases} T & |f| \leq \frac{1}{2T} \\ 0 & \text{otherwise} \end{cases}$$

על אות זה נקבל בדיוק את אות המקור, כיוון שהסינון הינו מכפלה בתדר, וע"י ההכפלה במסנן הנ"ל אנו מבטלים את כל השכפולים למעט זו הראשית, וכיוון שהיא בדיוק ההתמרה של אות המקור, הרי שביצענו בשלמות פעולה הפוכה לפעולת הדגימה וקיבלנו שחזור.

מבחינה מעשית, לא נתון לנו האות $\tilde{s}(t)$, אלא אות דגום $s[n]$. לכן השאלה המתבקשת היא כיצד בפועל מבצעים את הסינון המשחזר? נענה לשאלה זו ע"י כיתוב מפורש של פעולת השחזור

$$\begin{aligned} \hat{s}(t) &= F^{-1}\{H(f)F\{\tilde{s}(t)\}\} = \tilde{s}(t) \otimes h(t) = \\ &= \int \sum_n s[n] \delta(x - nT) h(t - x) dx = \sum_n s[n] h(t - nT) \end{aligned}$$

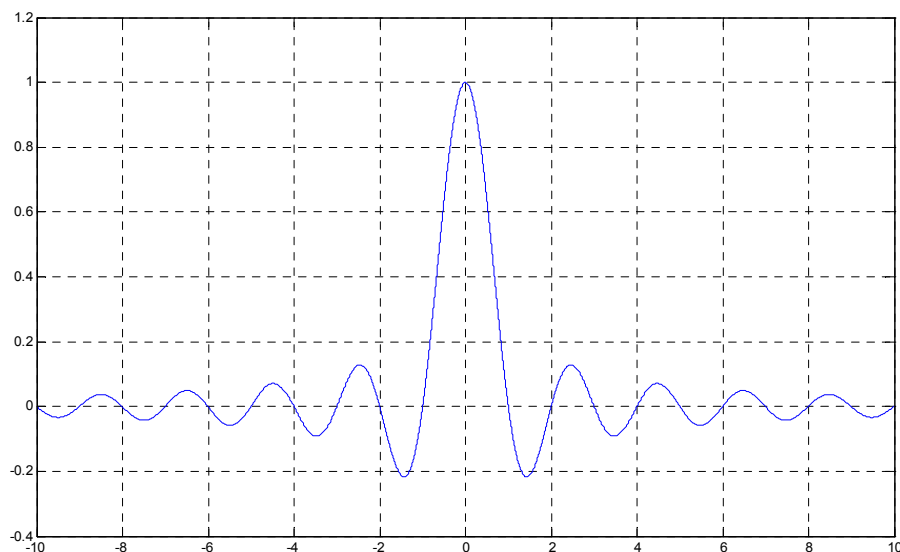
ההתמרה ההפוכה (מהתדר לזמן) של המסנן $H(f)$ תהיה (ראה פרק קודם)

$$\text{sinc}\{nt/T\} = \frac{\sin\{nt/T\}}{nt/T}$$

ולכן השחזור ייעשה ע"י הנוסחה

$$\hat{s}(t) = \sum_n s[n] h(t - nT) = \sum_n s[n] \text{sinc}\left(\frac{n(t - nT)}{T}\right)$$

מה משמעות נוסחה זו? ננסה להבינה ע"י הצגת פונקציית ה-sinc תחילה. פונקציה זו מתוארת בציור הבא.



פונקציית ה-sinc עבור $T=1$

ציור זה נבנה ע"י פקודות ה-Matlab :

```
T=1;t=-10.0005:0.001:10;ss=sin(pi*t/T)./(pi*t/T);
plot(t,ss); axis([-10 10 -0.4 1.2]); grid
```

נשים לב כי פונקציה זו דועכת באופן אוסצילטורי כשמרחקים מהראשית, והיא מתאפסת זהותית בנקודות $t = nT$ חוץ מאשר עבור $n=0$, שם ערכה הוא 1. משמעות הדבר היא שכאשר נרצה להשתמש בפונקציה זו לשחזור של ערך הפונקציה בנקודת זמן שהינה מכפלה שלמה של מרווח הדגימה, נקבל $t = k_0 T$

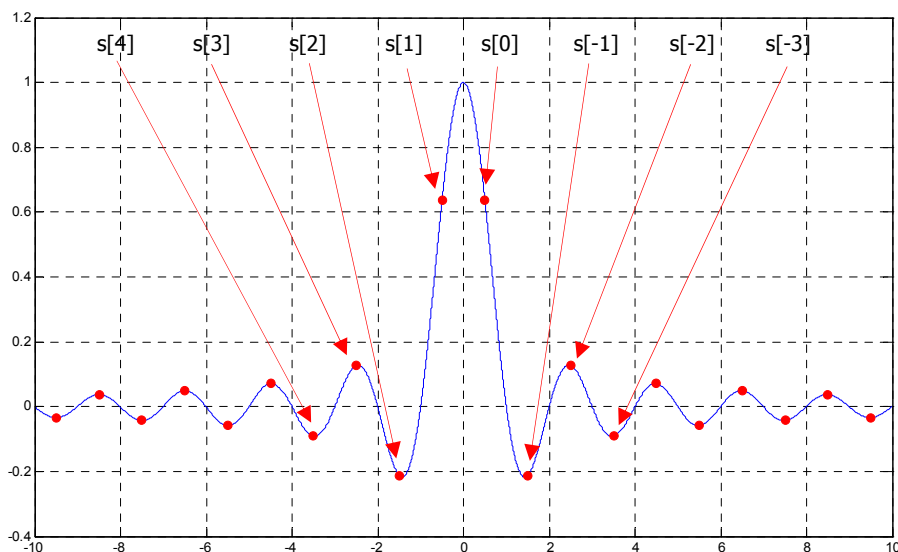
$$\hat{s}(k_0 T) = \sum_n s[n] \operatorname{sinc}(n(k_0 - n)T / T) = s(k_0 T)$$

תוצאה זו רצויה וצפויה. ערכי הדגימה ידועים במדויק ולכן טבעי שיתקבלו גם כערכי הפונקציה הרציפה בנקודות זמן אלה.

נבחן כעת כיצד מתנהגת משוואת השחזור כאשר נרצה לשחזר את ערך הפונקציה בנקודת הזמן $t = T/2$. נקבל

$$\begin{aligned} \hat{s}(T/2) &= \sum_n s[n] \operatorname{sinc}\left(n \frac{(T/2 - nT)}{T}\right) = \\ &= \dots + s[-1] \operatorname{sinc}(3\pi/2) + s[0] \operatorname{sinc}(\pi/2) + s[1] \operatorname{sinc}(-3\pi/2) + \dots \end{aligned}$$

הציור הבא ממחיש את משמעות המשוואה (עבור $T=1$).

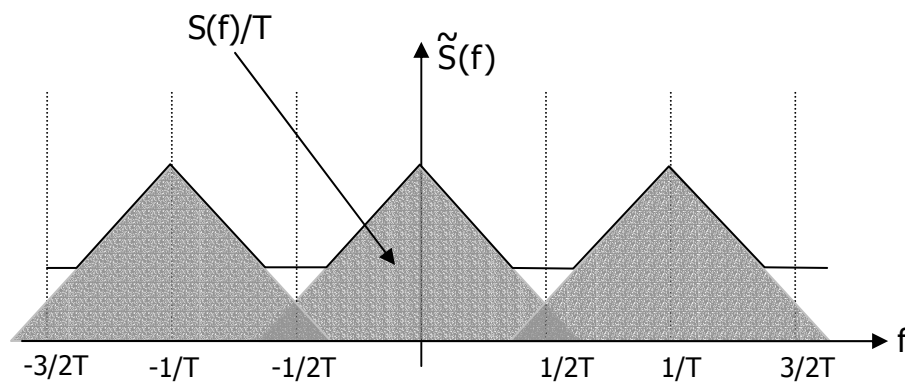


תיאור האינטרפולציה המתקבלת מנוסחת השחזור

אנו רואים כי הדומיננטיים בחישוב הערך הם באופן טבעי $s[0]$ ו- $s[1]$, אך לשם חישוב מלא יש להשתמש בכל ערכי הדגימות מ- $-\infty$ ועד ∞ , כשלכל דגימה יש השפעה פרופורציונית הפוכה למרחקה מנקודת היעד.

תופעת הקיפול ומשמעותה

דרשנו כי תדר הדגימה (אם T הוא מרווח הדגימה, תדרה הוא $1/T$ כדי לומר כי יש $1/T$ דגימות בשנייה) יהיה גדול מפעמיים התדר המירבי באות, וזה יבטיח יכולת תיאורתית של שחזור. השאלה בה נעסוק כעת היא – מה קורה אם מרווח הדגימה T אינו מקיים דרישה זו? במקרה זה, הרפליקות המופיעות באופן מחזורי תדרכנה זו על זו ותסתכמנה, כמתואר בציור הבא.



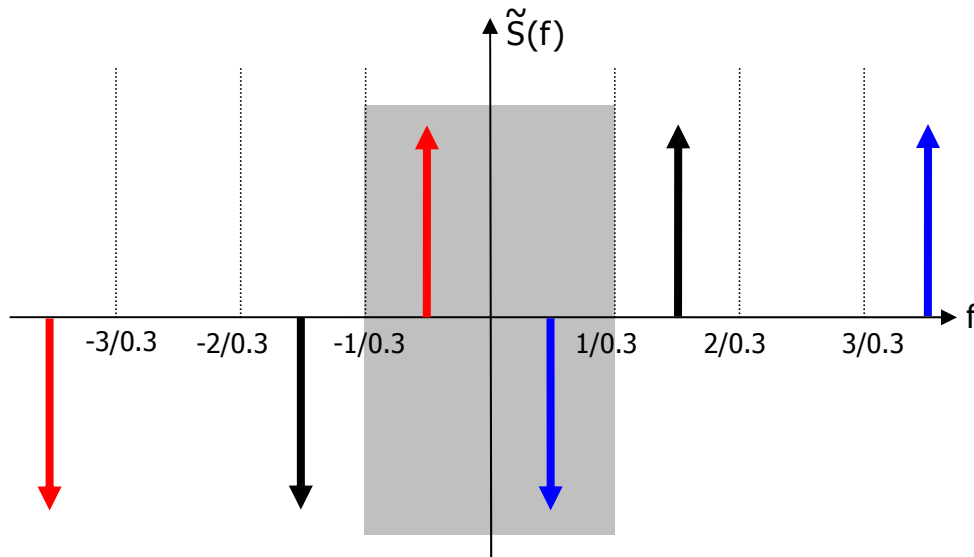
תופעת הקיפול בדגימה של אות בזמן – ההתנהגות בתדר

מציור זה ברור כי לא ניתן לשחזר את האות המקורי – זנבות המשולשים נפגשים ומסתכמים ולא נוכל לדעת את ערך הזנב האמיתי בתדרים אלו. רואים גם כי לא רק שנגרם נזק למידע שפולש מתחום התדרים $[-1/2T, 1/2T]$ אלא נזק גם נגרם לתדרים סמוכים להם בתוך האינטרוול, בשל אותה חפיפה. בשל תופעה זו ונזקה, מקובל לבצע סינון מגביל סרט לאותות קודם לדגימתן. אם את אותו שתואר בציור לעיל היינו מעבירים דרך מסנן

$$H(f) = \begin{cases} T & |f| < \frac{1}{2T} \\ 0 & \text{otherwise} \end{cases}$$

היינו מונעים את תופעת הקיפול וממזערים את הנזק רק לתדרים שפולשים מטווח התדרים שהדגימה מאפשרת.

מה משמעות תופעת הקיפול (Aliasing)? השם aliasing אינו קיפול כי אם התחזות, ואמנם ישנה כאן תופעה של התחזות. נראה זאת בדרך של דוגמה. נניח כי בידינו האות $s(t) = \sin(2\pi \cdot 5t)$. התדר המרבי בו הוא כמובן 5 ולכן מרווח הדגימה צריך להיות 0.1 שניות ומטה. נניח כי "פשענו" כנגד אות זה ודגמנו אותו במרווחים של 0.15 שניות (גדול מהמותר). הייצוג בתדר יראה כך

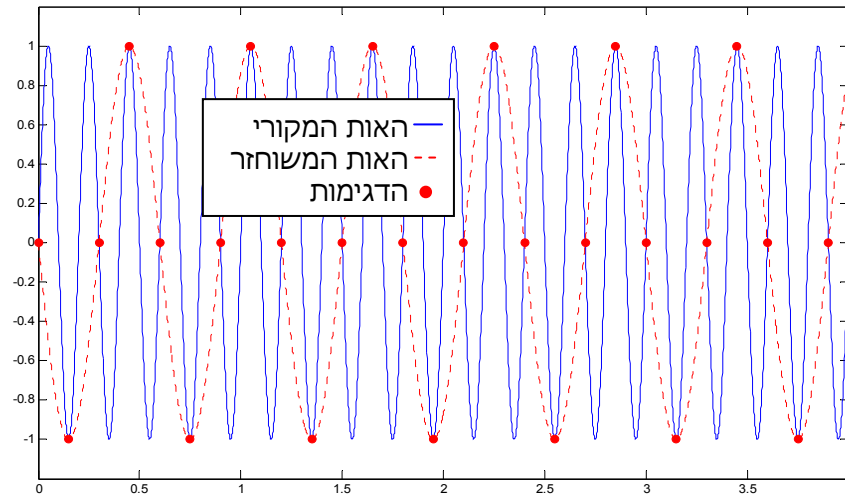


תיאור האות $s(t) = \sin(2\pi \cdot 5t)$ בתדר לאחר דגימה במרווח $T=0.15$

כל פונקצית הלים בציר הנ"ל בעלת גובה $1/2j$. הרפליקה המקורית מתוארת בשחור, זו שמימינה בכחול, וזאת שמשמאלה באדום. בבואנו לשחזר את האות, אם נבחר את התוכן בתחום $[-1/2T, 1/2T]$ ונאפס את היתר כפי ששחזור צריך להיעשות, נקבל שתי פונקציות הלים בתדרים $\pm 5/3$, ובהתמרה לזמן יהיה זה האות הסינוסי $s(t) = \sin(2\pi \cdot 5t/3)$ - קיבלנו אות מתחזה עם תדר שלא היה במקור. תיאור האות המקורי הרציף, הדגום והמשוחזר מתוארים בציר הבא. למעשה, התדר שהתקבל הוא תמונת המראה של התדר המקורי, כשציר הסימטריה הוא תדר הסף הנובע מהדגימה. תדר המקור הוא 5, תדר הסף הוא 3.333, והתדר שהתקבל הוא $3.333 - (5 - 3.333) = 1.667$.

קוד ה-Matlab שנתן גרף זה הוא

```
clf; t=0:0.001:4; s=sin(2*pi*5*t); plot(t,s); axis([0 4 -1.2 1.2]); hold on; t1=0:0.15:4; s1=sin(2*pi*5*t1); h=plot(t1,s1,'ro'); set(h,'MarkerFaceColor','r'); s2=-sin(2*pi*5*t/3); plot(t,s2,'r:');
```

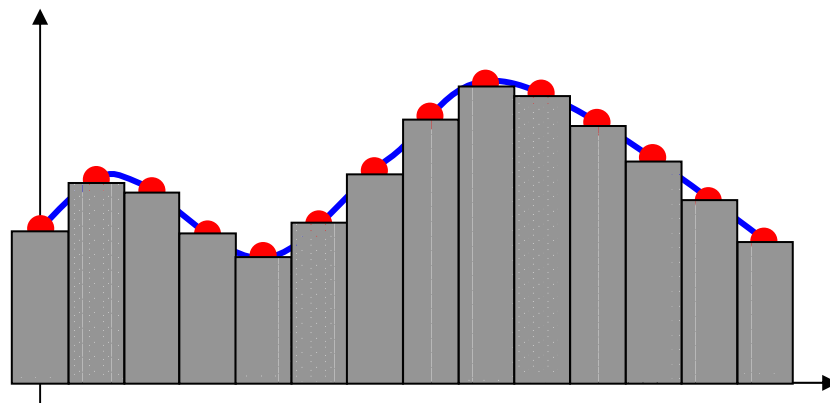


תופעת הקיפול כפי שתבטא לאות סינוסי (במקור יש 5 מחזורים בשניה, והאות המשוחזר כולל 1.666 מחזורים בשניה)

שחזור מעשי מדגימה אחידה

פעולת השחזור נועדה לקחת אות דגום ולחזור אל הרצף. שחזור מדגימה בנקודה כלשהי עפ"י נוסחת השחזור שהתקבלה קודם לכן מחייב ביצוע צירוף ליניארי המערב את כל אינסוף הדגימות הנתונות לנו. בפועל, כשמשחזרים חייבים לקרב פעולה זו ע"י שימוש בצירוף ליניארי בתמך מוגבל וסופי. קיימות מספר אפשרויות מעשיות.

השחזור הפשוט ביותר מציע החלפת כל דגימה בפונקציה מלבנית סימטרית ברוחב T ובגובה הדגימה. גישה זו קרויה Zero Order Hold, והיא מתוארת בציור הבא.



שחזור מעשי מדגימה אחידה בגישה Zero Order Hold

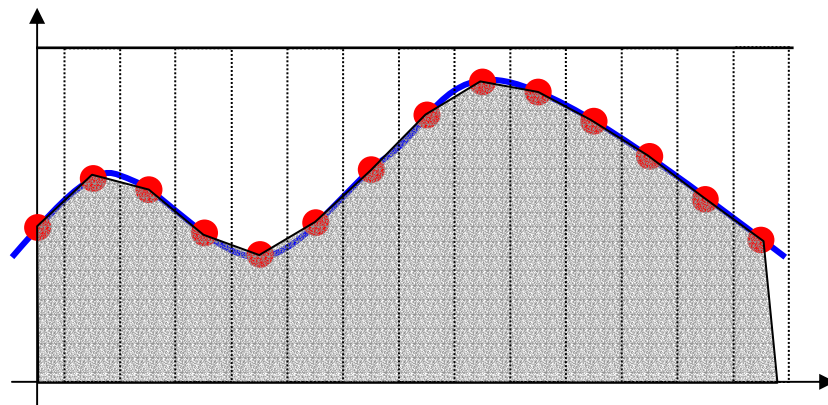
בפועל, התהליך הנ"ל אינו אלא קונבולוציה של האות הדיסקרטי (כשהוא מיוצג כרכבת הלמים בגבהים שונים - $\tilde{s}(t)$) עם פונקציה מלבנית מהצורה

$$h_{\text{ZOH}}(t) = \begin{cases} 1 & |t| \leq T/2 \\ 0 & \text{אחרת} \end{cases}$$

כלומר,

$$\hat{s}_{\text{ZOH}}(t) = \tilde{s}(t) \otimes h_{\text{ZOH}}(t)$$

שחזור מבוסס ZOH נחשב גרוע כיוון שהוא מציג מדרגות באות שלא היו במקור. אנו רואים גם כמה סוטה הפונקציה המשוחזרת בהשוואה לפונקצית המקור. אפשרות טובה יותר של שחזור היא מתיחת קווים ישרים בין הדגימות כמתואר בציור הבא. גישה זו קרויה First Order Hold.

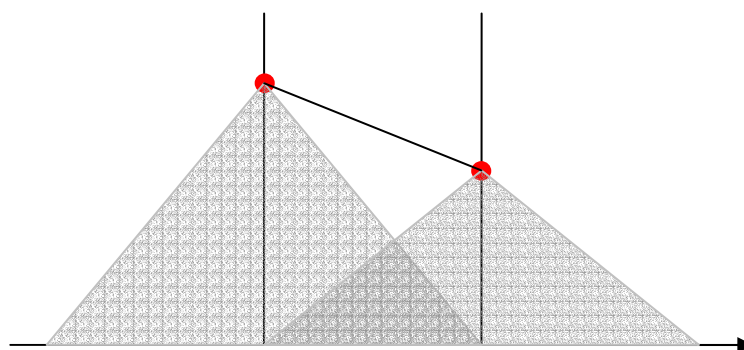


שחזור מעשי מדגימה אחידה בגישה ה-First Order Hold

גם כאן ישנה אינטרפטציה של קונבולוציה לתהליך השחזור, כשפעם פונקצית האינטרפולציה נתונה ע"י הפונקציה המשולשת

$$h_{\text{FOH}}(t) = \begin{cases} 1 - t/T & 0 \leq t \leq T \\ 1 + t/T & -T \leq t < 0 \\ 0 & \text{אחרת} \end{cases}$$

והציור הבא ממחיש כיצד קונבולוציה זו יוצרת את הקווים הנמתחים בין הדגימות



קונבולוציה עם FOH כדרך למתוח קוים בין דגימות באות

מעניין לציין כי ניתן להציג את הפונקציה המשולשת כקונבולוציה של שני פונקציות מלבן,

$$h_{\text{FOH}}(t) = \frac{1}{T} h_{\text{ZOH}}(t) \otimes h_{\text{ZOH}}(t)$$

ואז,

$$\hat{s}_{\text{FOH}}(t) = \tilde{s}(t) \otimes h_{\text{FOH}}(t) = \frac{1}{T} h_{\text{ZOH}}(t) \otimes h_{\text{ZOH}}(t) \otimes \tilde{s}(t) = \frac{1}{T} h_{\text{ZOH}}(t) \otimes \hat{s}_{\text{ZOH}}(t)$$

משוואה זו אומרת כי את אותו אפקט של אינטרפולציית FOH היינו מקבלים לו לקחנו את תוצאת ה-ZOH וסיננו אותה ע"י מסנן מלבני נוסף זהה אך עם שטח יחידה. למסנן נוסף זה יש אפקט של החלקת האות וביטול אותן קפיצות שהזכרנו.

ניתן להמשיך ולתאר עקומים מסדר גבוה יותר לאינטרפולציה, אך אנו נעצור את הדיון בשלב זה. אינטרפולציות מצליחות יפה במקום בו האות מתנהג חלק, ונכשלות כאשר האות מבצע מעברים חריפים. נושא זה יחזור להטרידנו כשנעסוק בתמונות. המשך ישיר של השיטות שהזכרו כאן יובא בהמשך כשנשאל כיצד להכליל לדו-מימד, ולהגדיל תמונה.

עיבוד אותות דגומים – מערכות ליניאריות וקבועות בזמן

דגמנו את האות שלנו $s(t)$ וקיבלנו $s[n]$. לרוב אף לא נצטרך לטרוח ונקבל את האות כשהוא כבר דגום. נוכל להתחיל לחשוב על עיבוד דיגיטלי של אותות, כשכלי היסוד הם התמרת פורייה של האות הדגום, וסינונו ע"י קונבולוציה. נתחיל את הדיון בתורת המערכות לאותות כאלה.

ממש כשם שביצענו סינון על אותות ברצף, נרצה לעשות זאת על אותות בזמן בדיד (דיסקרטיים). יהיה עלינו להראות כיצד מערכת מתנהגת, וכיצד סינון מומר לקונבולוציה. למעשה, אנו אמורים לחזור על כל הסיפור של מהי מערכת הפועלת על אותות כאלה, מהי ליניאריות, מהי קביעות בזמן, ועוד. נעשה זאת בתמציתיות משום שכל רעיונות אלו חוזרים על עצמם, וניגש מיד למערכות ליניאריות וקבועות בזמן.

מערכת מוגדרת כקבועה בזמן אם לכל אות כניסה מתקבל כי השהיית הכניסה יוצרת השהייה דומה ביציאה:

$$\forall s[n], d: g[n] = T\{s[n]\} \Rightarrow g[n-d] = T\{s[n-d]\}$$

מערכת מוגדרת כליניארית אם היא מקיימת את התכונה הבאה - לכל שני אותות $s_1[n]$ ו- $s_2[n]$ ולכל שני סקלרים a ו- b , יתקיים כי

$$T\{a \cdot s_1[n] + b \cdot s_2[n]\} = a \cdot T\{s_1[n]\} + b \cdot T\{s_2[n]\}$$

והציור שהראינו קודם בהקשר לכך ביחס למערכות בזמן רציף תקף באותה מידה. ניתן מספר דוגמאות להמחיש זאת:

- המערכת $g[n] = s[n] + 3s[n-1]$ היא מערכת ליניארית וקבועה בזמן. למעשה זו מערכת סיבתית, ועם זיכרון.

- המערכת $g[n] = s[n] + \sqrt{n} \cdot s[n-1]$ תלויה בזמן, אך היא עדיין ליניארית, כי היא בונה את מוצאה כצירוף ליניארי של אות הכניסה.

- המערכת $g[n] = 2s[n] + 3$ אינה ליניארית, אך היא קבועה בזמן. קל לראות זאת לפי ההגדרה.

- המערכת $g(t) = \sum_{j=0}^6 w_j \cdot s[n-j]$ היא מערכת ליניארית וקבועה בזמן. זוהי

מערכת שמבצעת ממוצע משוקלל של 7 הדגימות האחרונות באות.

- המערכת $g[n] = s[n] + 3s[n-0.5]$ לא מוגדרת! ברגע שעברנו לאות דגום, אין זמן שברי.

- המערכת $g[n] = s[n]s[n-1]$ אינה ליניארית אך היא קבועה בזמן.

ממש כשם שעשינו זאת לרצף, יש להגדיר פונקצית הלים דרכה נוכל לנתח התנהגות מערכות ליניאריות וקבועות בזמן. פונקצית ההלים הדיסקרטית היא

$$\delta_d[n] = \begin{cases} 1 & n = 0 \\ 0 & \text{אחרת} \end{cases}$$

בדומה להתנהגות פונקצית ההלם ברצף, קיימת התכונה הבאה: עבור אות $s[n]$ כלשהו ופונקצית ההלם הנ"ל, יתקבל

$$\sum_{k=-\infty}^{\infty} \delta_d[k] s[n-k] = \sum_{k=-\infty}^{\infty} \delta_d[n-k] s[k] = s[n]$$

וקל לראות נכונות נוסחה זו, כי פונקצית ההלם קיימת רק באחד האינדקסים ומתאפסת לכל האחרים. משמעות משוואה זו שהאות $s[n]$ יכול להיות מתואר כצירוף ליניארי של אותות הלם מוזות $\delta[n-k]$ כשהן משוקללות ע"י ערכי האות עצמו $s[k]$.

אם נזין למערכת זו הלם $\delta_d[n]$, נקבל מוצא אשר ייקרא התגובה להלם ויסומן $h[n]$, כלומר $T\{\delta_d[n]\} = h[n]$. נניח כעת כי הזנו אות כלשהו אחר $s[n]$ למערכת T . לכן, כיוון שהאות יכול להיות מתואר כסכום הלמים נקבל

$$\begin{aligned} g[n] = T\{s[n]\} &= T\left\{\sum_{k=-\infty}^{\infty} \delta_d[n-k] s[k]\right\} = \\ &= \sum_{k=-\infty}^{\infty} T\{\delta_d[n-k] s[k]\} = \sum_{k=-\infty}^{\infty} h[n-k] s[k] \end{aligned}$$

כאן השתמשנו בליניאריות של המערכת כשהחלפנו סדר פעולת המערכת עם הסכימה. השתמשנו בקביעות בזמן כשהנחנו שפעולת המערכת על הלם מוז היא תגובת הלם מוזת.

קיבלנו כי כל מערכת ליניארית ניתנת לאפיון ע"י תגובתה להלם, ובהינתן אות כניסה כלשהו, תגובתה ההלם מאפשרת לנו חישוב מוצאה. המשוואה שקיבלנו אינה אלא פעולת הקונבולוציה הדיסקרטית, אותה נסמן כ:

$$g[n] = T\{s[n]\} = \sum_{k=-\infty}^{\infty} h[n-k] s[k] = \sum_{k=-\infty}^{\infty} h[k] s[n-k] = h[n] \otimes s[n]$$

חישוב קונבולוציה זו דומה לחישוב זו הרציפה, אלא שסכימה דיסקרטית מחליפה את האינטגרל. נמחיש ביצוע חישוב זה דרך דוגמה:

$$g[n] = s[n-2] + s[n-1] + s[n] \Rightarrow h[n] = \delta_d[n-2] + \delta_d[n-1] + \delta_d[n]$$

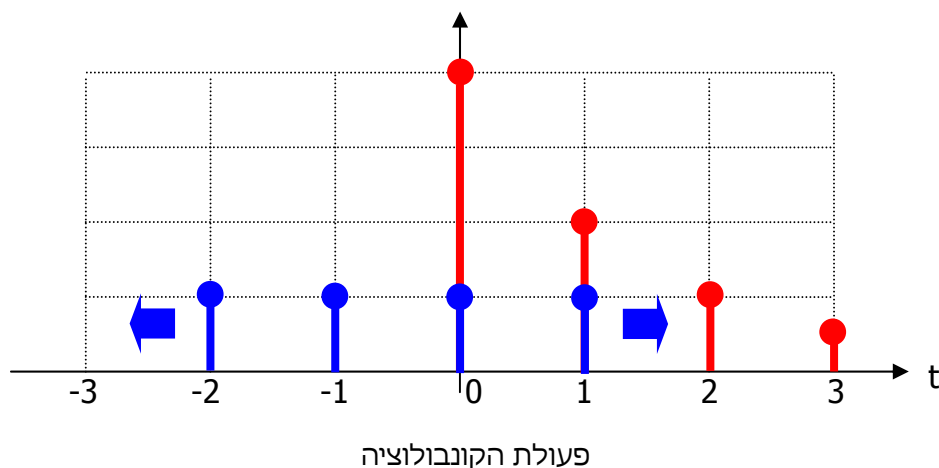
נניח כי אות הכניסה שלנו הוא

$$s[n] = \begin{cases} 4 \cdot 2^{-n} & n \geq 0 \\ 0 & \text{אחרת} \end{cases}$$

הציור הבא ממחיש כיצד חישוב זה יתבצע – אנו מקפוימים את ציור האות $s[n]$ (באדום), יוצרים תמונת מראה של $h[n]$, ומסיעים אותו לאורך ציר הזמן תוך כדי חישוב סכום המכפלות בחפיפה שבין שני האותות.

עבור $n=-1$ ומטה נקבל כי המוצא הוא אפס כי אין חפיפה בין השניים.
 נשים לב לכך שכיוון שהמערכת סיבתית (התגובה להלם קיימת רק לזמנים מאפס ומעלה), וכיוון שהאות סיבתי, ברור כי גם המוצא יהיה סיבתי.
 עבור $n=0$ נקבל חפיפה רק בנקודת זמן אחת והמוצא יהיה 4. עבור $n=1$ נקבל כי המוצא הוא $4 + 2 = 6$. באופן דומה, עבור $n=2$ המוצא הוא $4+2+1=7$, עבור $n=3$ המוצא הוא $4+2+1+0.5=7.5$. מנקודת זמן זו נקבל כי החפיפה מלאה תמיד, ועם כל תנועה בנקודת זמן אחת המוצא הוא חצי מקודמו. לכן התוצאה היא

$$g[n] = \begin{cases} 0 & n \leq -1 \\ 4 & n = 0 \\ 6 & n = 1 \\ 7 & n = 2 \\ 7.5 \cdot 2^{n-3} & n \geq 3 \end{cases}$$



כל תכונות הקונבולוציה שתוארו למקרה הרציף תקפות גם כעת:

- קומוטטיביות - $s[n] \otimes g[n] = g[n] \otimes s[n]$.
- אסוציאטיביות - $s[n] \otimes (g[n] \otimes f[n]) = (s[n] \otimes g[n]) \otimes f[n]$.
- דיסטריביוטיות - $s[n] \otimes [g[n] + f[n]] = s[n] \otimes g[n] + s[n] \otimes f[n]$.
- מכפלה בסקלר - $a \cdot [s[n] \otimes g[n]] = [as[n]] \otimes g[n] = [ag[n]] \otimes s[n]$.
- הזזה - $s[n] \otimes g[n - k] = s[n - k] \otimes g[n]$.

כשדנו במקרה הרציף ציינו שלוש תכונות של מערכות ליניאריות וקבועות בזמן – אלה תקפות גם כאן:

1. מערכות כאלה משתחלפות – זה נובע ישירות מהתכונות של הקונבולוציה.

2. מערכת סיבתית אם תגובתה להלם סיבתית –

$$g[n] = T\{s[n]\} = \sum_{k=-\infty}^{\infty} h[k]s[n-k] \quad \stackrel{\uparrow}{=} \sum_{k=0}^{\infty} h[k]s[n-k] \quad \forall n < 0, h[n] = 0$$

3. אם נזין למערכת אות חסום $\forall n, |s[n]| \leq M$, יתקבל כי מערכת ליניארית וקבועה בזמן תהיה יציבה BIBO אם סכום הערכים המוחלטים של תגובת ההלם הינו מספר סופי.

עיבוד אותות דגומים – התמרת פורייה

נדבר כעת על התמרת פורייה לאות בזמן בדיד – כיצד זו תתבצע? בדומה לנעשה עבור הרצף, נגדיר את האותות שחולפים דרך המערכת ללא שינוי. אותות אלו יהיו $s[n] = \exp\{j2\pi n\theta\}$. הזנת אות כזה למערכת ליניארית וקבועה בזמן המאופיינת ע"י תגובתה להלם $h[n]$ תיתן

$$g[n] = \sum_{k=-\infty}^{\infty} h[k]s[n-k] = \exp\{j2\pi n\theta\} \sum_{k=-\infty}^{\infty} h[k] \exp\{-j2\pi k\theta\} = H(\theta) \cdot \exp\{j2\pi n\theta\}$$

וכך קיבלנו הגדרה של התמרת פורייה לאות בזמן בדיד, ע"י

$$S(\theta) = F\{s[n]\} = \sum_n s[n] \cdot \exp\{-j2\pi n\theta\}$$

נשים לב כי $\forall k \in Z, S(\theta) = S(\theta + k)$, כלומר התמרת הפורייה הזו מחזורית, ואורך מחזור הוא 1. לכן נתעניין בתחום $-0.5 \leq \theta \leq 0.5$ בלבד.

כדוגמה, עבור האות הדיסקרטי $s[n] = \exp\{-a|n|\}$ עבור $a > 0$

התמרת הפורייה היא

$$\begin{aligned} S(\theta) = F\{s[n]\} &= \sum_{n=-\infty}^{\infty} s[n] \cdot \exp\{-j2\pi n\theta\} = \sum_{n=-\infty}^{\infty} \exp\{-j2\pi n\theta - a|n|\} = \\ &= \sum_{n=0}^{\infty} \exp\{-j2\pi n\theta - a|n|\} + \sum_{n=-\infty}^{-1} \exp\{-j2\pi n\theta - a|n|\} = \\ &= \sum_{n=0}^{\infty} \exp\{-(j2\pi\theta + a)n\} + \sum_{n=1}^{\infty} \exp\{(j2\pi\theta - a)n\} \end{aligned}$$

נשתמש בנוסחה של טור גיאומטרי

$$\forall |x| < 1, \sum_{n=0}^{\infty} x^n = \frac{1}{1-x}$$

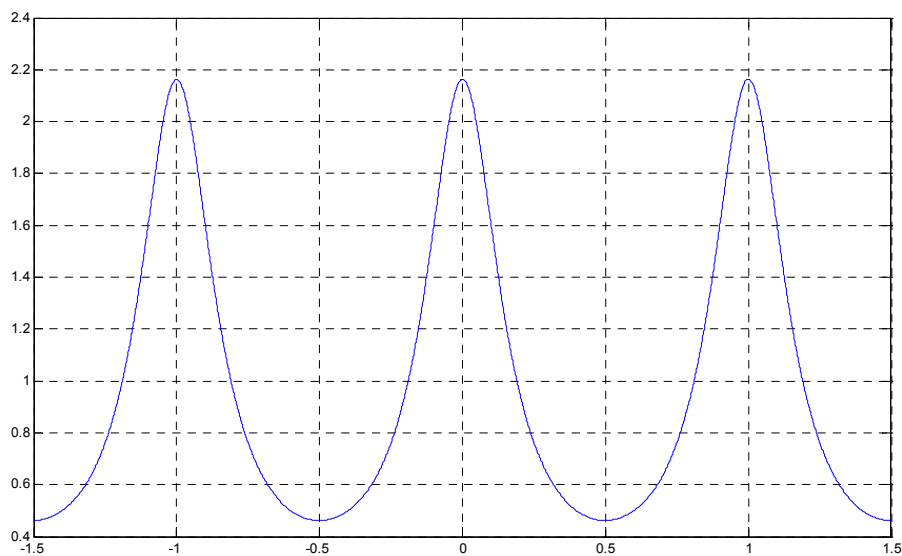
קל להוכיח קשר זה כיוון שע"י הכפלה ב- $(1-x)$ נקבל זהות:

$$(1-x) \sum_{n=0}^{\infty} x^n = \sum_{n=0}^{\infty} x^n - \sum_{n=1}^{\infty} x^n = 1$$

לכן,

$$\begin{aligned} S(\theta) &= \sum_{n=0}^{\infty} \exp\{-(j2\pi\theta + a)n\} + \sum_{n=1}^{\infty} \exp\{(j2\pi\theta - a)n\} = \\ &= \frac{1}{1 - \exp\{-j2\pi\theta - a\}} + \frac{1}{1 - \exp\{j2\pi\theta - a\}} - 1 = \\ &= \frac{1 - \exp\{j2\pi\theta - a\} + 1 - \exp\{-j2\pi\theta - a\}}{1 - \exp\{-j2\pi\theta - a\} - \exp\{j2\pi\theta - a\} + \exp\{-2a\}} - 1 = \\ &= \frac{2 - 2\exp\{-a\}\cos(2\pi\theta)}{1 - 2\exp\{-a\}\cos(2\pi\theta) + \exp\{-2a\}} - 1 = \\ &= \frac{1 - \exp\{-2a\}}{1 - 2\exp\{-a\}\cos(2\pi\theta) + \exp\{-2a\}} \end{aligned}$$

זהו הגרף של פונקציה זו, כשהוא מתואר בכוונה בתחום $[-1.5, 1.5]$ על מנת להראות 3 מחזורים שלמים שלו. נשים לב כי הפונקציה ממשית בשל היות האות המקורי סימטרי. אנו לא נטרח לחזור על תכונות התמרת הפורייה הדיסקרטית כיוון שהם מאוד דומים לתכונות התמרת הפורייה המקורית.



התמרת הפורייה הדיסקרטית של האות האקספוננציאלי הדו-צדדי

גרף זה התקבל ע"י פקודות ה-Matlab הבאות:

```
a=1; theta=-1.5:0.001:1.5;
F=(1-exp(-2))./(1-2*exp(-1)*cos(2*pi*theta)+exp(-2));
plot(theta,F);
```

מה באשר להתמרה ההפוכה? בהינתן $S(\theta)$, נרצה לחזור לאות המקור $s[n]$. זה ייעשה ע"י

$$s[n] = \int_{-1/2}^{1/2} S(\theta) \exp\{j2\pi n\theta\} d\theta$$

נוכיח את נכונות פעולה זו ע"י הצבת ההתמרה קדימה בתוך זו לקבלת אותו אות כניסה. נקבל:

$$\begin{aligned} s[n] &= \int_{-1/2}^{1/2} S(\theta) \exp\{j2\pi n\theta\} d\theta = \int_{-1/2}^{1/2} \left(\sum_{k=-\infty}^{\infty} s[k] \exp\{-j2\pi k\theta\} \right) \exp\{j2\pi n\theta\} d\theta = \\ &= \sum_{k=-\infty}^{\infty} s[k] \int_{-1/2}^{1/2} \exp\{j2\pi \theta(n-k)\} d\theta = \\ &= \sum_{k=-\infty}^{\infty} s[k] \left[\frac{\exp\{j\pi(n-k)\} - \exp\{-j\pi(n-k)\}}{j2\pi(n-k)} \right] = \\ &= \sum_{k=-\infty}^{\infty} s[k] \text{sinc}(\pi(n-k)) = \sum_{k=-\infty}^{\infty} s[k] \delta[n-k] = s[n] \end{aligned}$$

השתמשנו כאן בתכונה שפונקציית ה-sinc מקבלת ערכי אפס למעט בראשית, שם היא 1.

תכונות ההתמרה

להתמרה שהוגדרה יש את מרבית התכונות שהצגנו להתמרה הרציפה. חשוב להבהיר כי אין זה במקרה – ההתמרה לאות הדגום אינה אלא מקרה פרטי של ההתמרה הכללית הרציפה. במסגרת זו נציג מספר תכונות חשובות:

- הגדרה - עבור האות $g[n]$ התמרת הפורייה נתונה ע"י

$$G(\theta) = F\{g[n]\} = \sum_{n=-\infty}^{\infty} g[n] \exp\{-j2\pi n\theta\}$$

$$g[n] = F^{-1}\{G(\theta)\} = \int_{-0.5}^{0.5} G(\theta) \exp\{j2\pi n\theta\} d\theta$$

וההתמרה חזרה ע"י

- ליניאריות – התמרת הפורייה ליניארית, כלומר, עבור כל שני אותות $s[n]$ ו- $g[n]$, ושני סקלרים a ו- b , מתקיים כי

$$F\{a \cdot s[n] + b \cdot g[n]\} = a \cdot F\{s[n]\} + b \cdot F\{g[n]\}$$

- ממוצע – ממוצע האות $g[n]$ אינו אלא $G(0)$:

$$G(0) = F\{g[n]\}_{\theta=0} = \sum_{n=-\infty}^{\infty} g[n]$$

- ממשיות וסימטריות – בהינתן לנו אות $g[n]$ ממשי טהור (כלומר ללא מרכיב מדומה), להתמרת הפורייה שלו מבנה סימטרי צמוד,

$$G(\theta) = \sum_{n=-\infty}^{\infty} g[n] \exp\{-j2\pi n\theta\}$$

$$\overline{G(\theta)} = \sum_{n=-\infty}^{\infty} g[n] \exp\{-j2\pi n\theta\} = \sum_{n=-\infty}^{\infty} g[n] \exp\{j2\pi n\theta\} = G(-\theta)$$

אם האות הנתון לנו $g[n]$ גם ממשי וגם סימטרי יתקבל כי $G(\theta)$ ממשי וסימטרי אף הוא.

- שימור מכפלה פנימית – עבור שני אותות $g[n]$ ו- $s[n]$, והתמרתם $G(\theta)$ ו- $S(\theta)$. מכפלה פנימית בין שתי פונקציות בזמן מוגדרת ע"י

$$\langle g[n], s[n] \rangle_n = \sum_{n=-\infty}^{\infty} g[n] \cdot \overline{s[n]}$$

מכפלה פנימית בתדר של פונקציות ההתמרה תוביל לאותה תוצאה:

$$\begin{aligned} \langle G(\theta), S(\theta) \rangle_{\theta} &= \int_{-0.5}^{0.5} G(\theta) \overline{S(\theta)} d\theta = \\ &= \int_{-0.5}^{0.5} \left(\sum_{n=-\infty}^{\infty} g[n] \exp\{-j2\pi n\theta\} \right) \left(\sum_{k=-\infty}^{\infty} \overline{s[k]} \exp\{j2\pi k\theta\} \right) d\theta \\ &= \sum_{n=-\infty}^{\infty} g[n] \sum_{k=-\infty}^{\infty} \overline{s[k]} \int_{-0.5}^{0.5} \exp\{-j2\pi(n-k)\theta\} d\theta \end{aligned}$$

כבר ראינו כי האינטגרל הפנימי אינו אלא פונקציית הלם דיסקרטית (זוהי פונקציית sinc הדגומה בנקודות ההתאפסות של פונקציה זו).

$$\begin{aligned} \langle G(\theta), S(\theta) \rangle_{\theta} &= \sum_{n=-\infty}^{\infty} g[n] \sum_{k=-\infty}^{\infty} \overline{s[k]} \int_{-0.5}^{0.5} \exp\{-j2\pi(n-k)\theta\} d\theta = \\ &= \sum_{n=-\infty}^{\infty} g[n] \sum_{k=-\infty}^{\infty} \overline{s[k]} \delta_D[n-k] = \sum_{n=-\infty}^{\infty} g[n] \overline{s[n]} \end{aligned}$$

כמקרה פרטי תיתן את תוצאת פרסבל:

$$\int_{-0.5}^{0.5} |G(\theta)|^2 d\theta = \sum_{n=-\infty}^{\infty} |g[n]|^2$$

- תכונת הקונבולוציה – עבור שני אותות $g[n]$ ו- $s[n]$ עם התמרתם $G(\theta)$ ו- $S(\theta)$. הגדרנו את הקונבולוציה בין שתי הפונקציות בזמן ע"י

$$g[n] = \sum_{k=-\infty}^{\infty} h[n-k] s[k]$$

נראה כי בתדר חישוב הקונבולוציה הופך למכפלה רגילה:

$$\begin{aligned}
F\{g[n] \otimes s[n]\} &= F\left\{\sum_{k=-\infty}^{\infty} h[n-k]s[k]\right\} = \\
&= \sum_{n=-\infty}^{\infty} \left(\sum_{k=-\infty}^{\infty} h[n-k]s[k]\right) \cdot \exp\{-j2\pi n\theta\} = \\
&= \sum_{k=-\infty}^{\infty} s[k] \exp\{-j2\pi k\theta\} \cdot \sum_{n=-\infty}^{\infty} h[n-k] \exp\{-j2\pi(n-k)\theta\} \\
&= G(\theta)S(\theta)
\end{aligned}$$

משמעות תוצאה זו היא שבמקום לחשב קונבולוציה נוכל לעבור להתמרות הפורייה של שני האותות המעורבים, להכפילם זה בזה סקלרית, ואז לבצע התמרה הפוכה חזרה לזמן.

טיפול באותות דגומים בעלי תמך סופי

נניח כי בידנו אות $s[n]$ שקיים רק עבור $n \in [0, N-1]$. הרי מרבית האותות בהם נעסוק יהיו כאלה, מכיוון שלא ניתן לטפל בקבצי אותות בגודל אינסופי. מה נוכל לומר על אותות אלו ביחס לסינונם, התמרתם, ועוד? קל יהיה פשוט להניח כי נטפל באות זה כאילו היה בעל תמך אינסופי, כשנניח התאפסות מחוץ לתמך $n \in [0, N-1]$. טכניקה זו קרויה המשכת האות באפסים. בדרך זו נוכל להציע קונבולוציה בין מערכת שתגובתה להלם $h[n]$ ובין אות זה ע"י

$$h[n] \otimes s[n] = \sum_{k=0}^{N-1} h[n-k]s[k] = \sum_{k=0}^{N-1} h[k]s[n-k]$$

כל שהשתנה כאן הוא תחום הסכימה הלוקח בחשבון את התאפסות האות מחוץ לתחום מסוים. מכיוון שכך, נוכל לעשות שימוש בתכונת הקונבולוציה של התמרת הפורייה, להמיר את האות כאילו היה קיים בכל זמן, להמיר את תגובת ההלם, להכפיל, ולבצע התמרה הפוכה.

פגם מטרד בביצוע התמרת הפורייה על $s[n]$ הוא העובדה שאנו מתחילים עם N מספרים, ומוצאים עצמנו מאפיינים אותם בתדר באמצעות ערכים ברצף בתחום $[-0.5, 0.5]$, לפי:

$$S(\theta) = F\{s[n]\} = \sum_{n=-\infty}^{\infty} s[n] \exp\{-j2\pi n\theta\} = \sum_{n=0}^{N-1} s[n] \exp\{-j2\pi n\theta\}$$

זה נראה כבזבז.

נוכל להקל זאת ולהציע לדגום את ציר התדר ב-N הנקודות $\theta = k/N$ עבור $k = -N/2, -N/2 + 1, \dots, -1, 0, 1, \dots, N/2 - 1$, או, כפי שיותר מקובל, להשתמש בערכי k בתחום $[0, N-1]$. בצורה זו נקבל את התמרת הפורייה הדיסקרטית:

$$S[k] = \text{DFT}\{s[n]\} = \sum_{n=0}^{N-1} s[n] \exp\left\{-\frac{j2\pi nk}{N}\right\} \quad k = 0, 1, 2, \dots, N-1$$

פעולת ה-Discrete Fourier Transform (DFT) מקבלת N מספרים המתארים אות כזמן, ומחזירה N מספרים המתארים את התנהגות אות זה בתדר, כשהרעיון הוא לדגום את הפונקציה $S(\theta)$.

כיצד נצדיק את הדגימה? מדוע שלא ייווצר מצב של אובדן כתוצאה ממנה? התשובה לכך, מסתבר, פשוטה. אם כתחליף להנחה שהאות הינו אפס מחוץ לתחום $[0, N-1]$ נניח כי הוא הומשך מחזורית, נקבל כי התמרת הפורייה הרגילה הייתה מובילה ל- $S(\theta) = F\{s[n]\}$ שאינה קיימת אלא בנקודות הדגימה, וערכיה שם היו בדיוק הערכים של ההתמרה תחת הנחת ההמשכה באפסים! לכן, בדגימה המוצעת אנו בסך הכל משנים את כלל ההמשכה מחוץ לתחום הקיום של האות.

האם ה-DFT הפיכה? מסתבר שכן! מסתבר כי ההתמרה ההפוכה נתונה

ע"י

$$s[n] = \text{IDFT}\{S[k]\} = \frac{1}{N} \cdot \sum_{k=0}^{N-1} S[k] \exp\left\{\frac{j2\pi nk}{N}\right\} \quad n = 0, 1, 2, \dots, N-1$$

קל לראות נכונות התמרה הפוכה זו לפי ההגדרה:

$$\begin{aligned} \text{IDFT}\{\text{DFT}\{s[m]\}\} &= \frac{1}{N} \cdot \sum_{k=0}^{N-1} S[k] \exp\left\{\frac{j2\pi nk}{N}\right\} = \\ &= \frac{1}{N} \cdot \sum_{k=0}^{N-1} \left(\sum_{n=0}^{N-1} s[n] \exp\left\{-\frac{j2\pi nk}{N}\right\} \right) \exp\left\{\frac{j2\pi mk}{N}\right\} = \\ &= \frac{1}{N} \cdot \sum_{n=0}^{N-1} s[n] \left(\sum_{k=0}^{N-1} \exp\left\{\frac{j2\pi (m-n)k}{N}\right\} \right) \end{aligned}$$

הסכימה הפנימית ניתנת לחישוב לפי נוסחת טור גיאומטרי סופי,

$$\sum_{k=0}^{N-1} q^k = \frac{1 - q^N}{1 - q}$$

ולכן נקבל

$$\sum_{k=0}^{N-1} \exp\left\{\frac{j2\pi(m-n)k}{N}\right\} = \frac{1 - \exp\{j2\pi(m-n)\}}{1 - \exp\left\{\frac{j2\pi(m-n)}{N}\right\}} = N \cdot \delta_D[m-n]$$

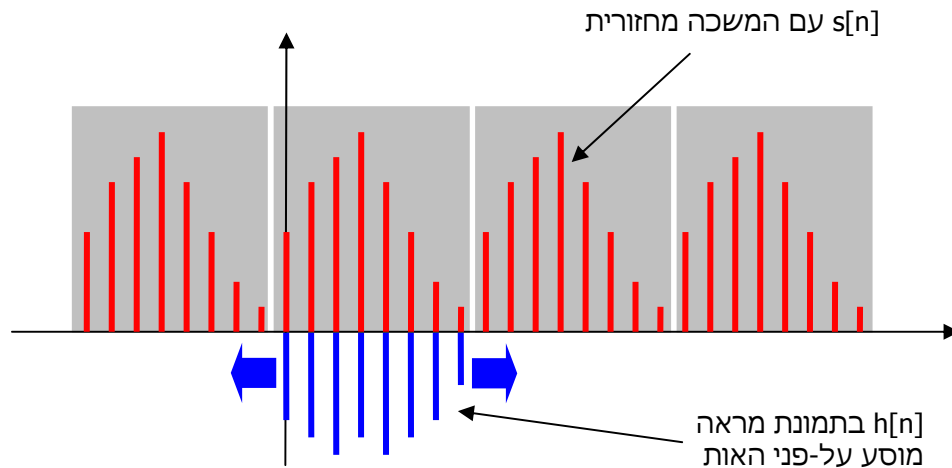
לכל בחירה $n \neq m$ (וכן $1-N \leq m-n \leq N-1$ כפי שאמנם קורה) המנה הנ"ל תתאפס כי המונה מתאפס. עבור $n=m$ מנה זו היא אפס חלקי אפס, ולפי פיתוח של פונקצית האקספוננט לפי טור טיילור ($\exp\{x\} \cong 1 + x + 0.5x^2$ עבור $x \rightarrow 0$) ערך המנה הוא N . לכן,

$$\text{IDFT}\{\text{DFT}\{s[m]\}\} = \frac{1}{N} \cdot \sum_{n=0}^{N-1} s[n] \left(\sum_{k=0}^{N-1} \exp\left\{\frac{j2\pi(m-n)k}{N}\right\} \right) = \sum_{n=0}^{N-1} s[n] \delta_D[m-n] = s[m]$$

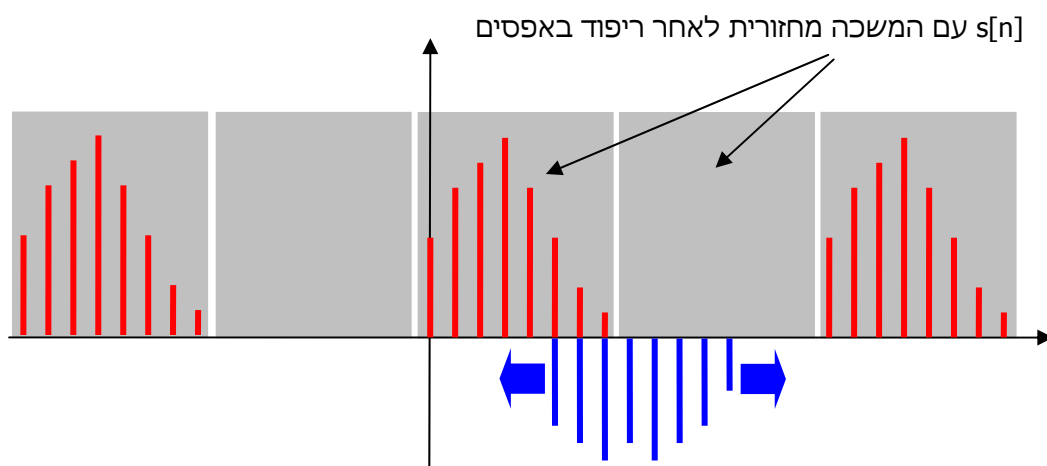
מסתבר כי ניתן לחזור ולהציג את כל אותן תכונות שהצגנו קודם (ליניאריות, הזזה ואפנון, ממוצע, ממשיות, פרסבל, קונבולוציה) גם להתמרה זו. ישנו הבדל אחד מהותי ואנחנו נבהיר אותו בקצרה – כבר ראינו בעבר כי דגימה של אות בתדר פירושה מחזוריות בזמן (ראינו זאת בהקשר לטורי פורייה). ראינו גם כי דגימה בזמן מתבטאת במחזוריות בתדר. בשל כך, בהגדרת ה-DFT דגמנו את פונקצית התדר, ולכן צפוי כי תתקבל מחזוריות בזמן. ובאמת, אם נתבונן בהגדרת ה-IDFT ונציב לתוכה ערכי n מחוץ ל- $[0, N-1]$, האות המקורי יחזור על עצמו במחזוריות של N .

תופעה זו פירושה שתכונת הקונבולוציה או הקורלציה עם DFT לא מתייחסות להמשכה של האות באפסים, אלא להמשכה מחזורית. לכן, אם בידינו שני אותות $s[n]$ ו- $h[n]$ – שניהם מוגדרים לתחום $[0, N-1]$, כשנבצע DFT על כל אחד מהם, נכפיל איבר-איבר (N מכפלות) ונבצע IDFT, האפקט יהיה של קונבולוציה של $h[n]$ עם אות מחזורי הנבנה מ- $s[n]$ (או להיפך). קונבולוציה זו קרויה קונבולוציה ציקלית, והיא מתוארת בציור הבא.

ומה אם נרצה קונבולוציה רגילה (ליניארית) בין שני אותות אלו? עלינו להאריך את שני האותות באפסים לאורך כפול מאורכם המקורי, ואז לעשות את הקונבולוציה בעזרת DFT. למרות היותה ציקלית, ה-"ריפוד" באפסים ייתן את התוצאה הנכונה, כפי שמתואר בציור הבא.



קונבולוציה ציקלית



קבלת קונבולוציה ליניארית ע"י DFT בשל ריפוד באפסים

בנושא אחרון בדיון זה, מזכיר כי פעולת ה- DFT יכולה להיות מיוצגת בצורה מטריצית, כמטריצה ריבועית בגודל N -על- N המקבלת וקטור באורך N ומכפילה אותו לקבלת וקטור מוצא באותו אורך. המספרים המאכלסים את המטריצה הזו הם $\exp\{j2\pi nk/N\}$. מסתבר כי מטריצה זו בעלת תכונות מיוחדות – היא יוניטרית עד כדי קבוע (כלומר, היפוכה מתקבל ע"י שחלוף צמוד), ותכונה זו מסבירה את נכונות תכונת הפרסבל, שהרי ידוע שמטריצות יוניטריות לא משנות אורך של וקטורים.

כשכופלים מטריצה בגודל N -על- N בוקטור באורך N יש N^2 מכפלות וסיכומים. תכונה חשובה של ה-DFT היא היכולת לבצע ההתמרה במורכבות של $N \log N$, והאלגוריתם המבצע זאת קרוי Fast Fourier Transform (FFT). אלגוריתם זה נפוץ והשימוש בו רווח. במסגרת זו לא נלמד כיצד מושג הרווח החישובי, ואנו נעצור את הדיון כאן.

עיבוד תמונות בעזרת מחשב (236327)

הרצאה מס' 5

אותות ומערכות 4 – אותות בדו-מימד – רצף

אותות דו-ממדיים

ממש כשם שיכולנו להגדיר אותות כפונקציה של הזמן, נוכל להגדירם לתחום הגדרה רב-מימדי. לשם דוגמה, אם רצוננו לעסוק במפות גבהים, נגדיר כפונקציות דו-ממדיות $s(x,y)$ המתארות גובה מעל פני הים בנקודות (x,y) במישור. כל שנאמר על הטיפול באותות בחד-מימד יעבור הכללה לדו-מימד – נדון בתכונות של אותות אלו, המערכות הפועלות עליהן, התמרתם לתדר ע"י התמרת פורייה, דגימתם, ועוד. הכללה זו תהיה טריוויאלית במקרים רבים, אך תעורר תהיות וסוגיות חדשות פה ושם.

אנו לא נחזור על כל פרטי הניתוח שבוצע לאותות חד-ממדיים, ונתרכז בעיקרי הדברים. מעבר לכך, נשתדל להגיע מהר ככל האפשר לאותות דגומים, שהרי אלה הם המעניינים אותנו בדיון בעיבוד ספרתי. לכן, כדי להיפטר אחת ולתמיד מאותות ברצף, נציג את משפט הדגימה לדו-מימד ואיתו נעבור לעולם האותות הדגומים. עם זאת, על מנת להציג משפט זה, חובה עלינו להגדיר כיצד מבוצעת קונבולוציה והתמרת פורייה בדו-מימד. לשם בניית משפט הדגימה עלינו גם לדעת כיצד נראית פונקציית ההלם הדו-מימדית, ופונקצית שה. נתחיל את הדיון בנושאים אלו.

יסודות בעיבוד אותות בדו-מימד

אנו עוסקים באות $f(x,y)$ המקבל ערך ממשי (או מרוכב) בכל נקודה במישור (x,y) . נסמן ב-T מערכת הפועלת על אות זה ומפיקה אות דו-מימדי אחר $g(x,y)$. לכן

$$T\{f(x,y)\} = g(x,y)$$

נגדיר בקצרה מספר תכונות יסודיות של מערכות מסוג זה:

- ליניאריות – בדומה לניסוח בחד-מימד, נדרוש שפעולת המערכת על סכום משוקלל של כניסות מביא לאותה תוצאה כמו סכום משוקלל של המוצאים:

$$T\{a \cdot f_1(x, y) + b \cdot f_2(x, y)\} = a \cdot T\{f_1(x, y)\} + b \cdot T\{f_2(x, y)\}$$
- זיכרון – אם $g(x, y) = T\{f(x, y)\}$ תלוי רק בערכה של הכניסה במקום (x, y) ולא בשום מקום אחר.
- קביעות במקום (זמן מוחלף במקום) – אם $g(x, y) = T\{f(x, y)\}$ אזי לכל הזזה במקום (dx, dy) יתקבל כי הפעלת המערכת על הכניסה המוזזת כמוה כמו הזזת המוצא המקורי, $g(x + dx, y + dy) = T\{f(x + dx, y + dy)\}$.
- יציבות – אם $\forall x, y \quad |f(x, y)| \leq M$ אזי גם המוצא חסום בכל מקום

$$|T\{f(x, y)\}| = |g(x, y)| < c \cdot M < \infty$$

ניתן מספר דוגמאות להמחיש תכונות אלו:

- המערכת $g(x, y) = \log(|f(x, y)|)$ אינה ליניארית, היא חסרת זיכרון, היא קבועה במקום, והיא אינה יציבה (כי ערכים קרובים לאפס מקבלים מוצא אינסופי).
- המערכת $g(x, y) = f(x, y) + \sqrt{x} \cdot f(x - 1, y)$ ליניארית, אינה חסרת זיכרון, אינה קבועה במקום ואינה יציבה (עבור $f(x, y) = 1$ יתקבל $g(\infty, y) = \infty$).
- המערכת $g(x, y) = \partial f(x, y) / \partial x + \partial^2 f(x, y) / \partial y^2$ ליניארית, קבועה במקום, בעלת זיכרון, ולא יציבה.

ממש בדומה לנעשה עבור אותות חד-ממדיים, נגדיר את פונקציית ההלם ונראה כי במערכות ליניאריות וקבועות במקום תגובת המערכת לאות זה מספרת הרבה על המערכת. פונקציית ההלם הדו-ממדית היא

$$\delta(x, y) = \begin{cases} 0 & x \neq y \\ \infty & x = y = 0 \end{cases} = \delta(x)\delta(y)$$

ברור שהגדרה זו לא מספקת כי עלינו לומר משהו על מידת האינסוף הנתונה. זאת נקבעת לפי

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \delta(x, y) dx dy = 1$$

גם את פונקציה זו יכולנו להציג כגבול של סידרת פונקציות בעלות תמך ריבועי סביב הראשית עם נפח יחידה, ולהקטין את התמך הריבועי בעוד הגובה צומח על מנת לשמר את הנפח.

התכונה המרכזית שעליה נתבסס היא התכונה

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \delta(x, y) f(x + a, y + b) dx dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \delta(\hat{x} - a, \hat{y} - b) f(\hat{x}, \hat{y}) d\hat{x} d\hat{y} = f(a, b)$$

ביטוי זה אומר כי כל אות דו-מימדי ניתן לתיאור כאוסף של הרבה פונקציות הלם דו-ממדיות מוזות וממושקלות.

בהינתן פונקצית ההלם הנ"ל, הזנתה למערכת ליניארית וקבועה במקום תיתן את המוצא $h(x, y)$. עבור אות כלשהו $f(x, y)$ מוצא המערכת יהיה

$$\begin{aligned} g(x, y) = T\{f(x, y)\} &= T\left\{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\hat{x}, \hat{y}) \delta(x - \hat{x}, y - \hat{y}) d\hat{x} d\hat{y}\right\} = \\ &\stackrel{\uparrow \text{ליניאריות}}{=} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\hat{x}, \hat{y}) T\{\delta(x - \hat{x}, y - \hat{y})\} d\hat{x} d\hat{y} \\ &\stackrel{\uparrow \text{קביעות במקום}}{=} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\hat{x}, \hat{y}) h(x - \hat{x}, y - \hat{y}) d\hat{x} d\hat{y} \end{aligned}$$

ניצלנו כאן את הליניאריות והקביעות במקום וקיבלנו את הקונבולוציה בדו-מימד כדרך לחישוב מוצא המערכת לכניסה כלשהי כתולדה של התגובה להלם. כלומר, נסמן מעתה:

$$g(x, y) = T\{f(x, y)\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\hat{x}, \hat{y}) h(x - \hat{x}, y - \hat{y}) d\hat{x} d\hat{y} = h(x, y) \otimes f(x, y)$$

כל תכונות הקונבולוציה שהצגנו קודם לחד-מימד (קומוטטיביות, אסוציאטיביות, דיסטריליביות, מכפלה בסקלר, הזזה, גזירה) תקפות כאן באותה מידה. גם המשמעות הגיאומטרית של פעולה זו דומה – עלינו לקחת אות אחד ולבצע היפוך מראה שלו בצירים x ו- y , להניחו על האות האחר, ולבצע סכימה על המכפלה שלהם.

נראה דוגמה להמחשת פעולה זו: נניח כי בידינו מערכת שתגובתה להלם היא $h(x, y) = (2\pi)^{-1} \exp\{-(x^2 + y^2)/2\}$, ורצוננו לחשב את מוצאה לאות המדרגה הדו-מימדי

$$s(x, y) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}$$

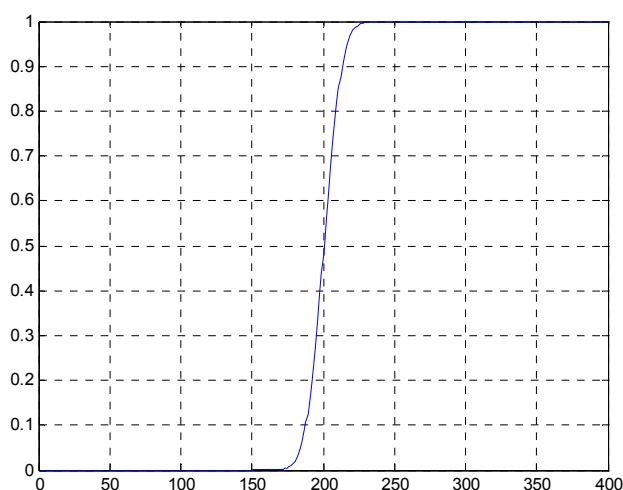
לפני שנפתור הבעיה, נשים לב כי פונקצית הלם חדה הופכת במערכת לתגובה להלם שהינה נקודה מטושטשת בעלת מבנה של פעמון גאוס. באופן דומה, הכנסת מדרגה תגרום לטשטוש המדרגה והחלפת המעבר החד במעבר מתון מ-0 ל-1. נראה זאת אנליטית:

$$\begin{aligned} g(x, y) &= T\{s(x, y)\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} s(\hat{x}, \hat{y}) h(x - \hat{x}, y - \hat{y}) d\hat{x} d\hat{y} = \\ &\stackrel{\substack{\uparrow \\ s(x, y) = 0 \\ \text{for } x < 0}}{=} (2\pi)^{-1} \int_{-\infty}^{\infty} \int_0^{\infty} \exp\left\{-\left([x - \hat{x}]^2 + [y - \hat{y}]^2\right)/2\right\} d\hat{x} d\hat{y} = \\ &\stackrel{\substack{\uparrow \\ \text{Gaussian} \\ \text{integral}}}{=} (2\pi)^{-1} \int_{-\infty}^{\infty} \exp\left\{-\frac{(y - \hat{y})^2}{2}\right\} d\hat{y} \cdot \int_0^{\infty} \exp\left\{-\frac{(x - \hat{x})^2}{2}\right\} d\hat{x} = \\ &= \frac{1}{\sqrt{2\pi}} \int_0^{\infty} \exp\left\{-\frac{(x - \hat{x})^2}{2}\right\} d\hat{x} \stackrel{\substack{\uparrow \\ z = \hat{x} - x}}{=} \frac{1}{\sqrt{2\pi}} \int_{-x}^{\infty} \exp\left\{-\frac{z^2}{2}\right\} dz = \\ &= \frac{1}{\sqrt{2\pi}} \int_{-x}^0 \exp\left\{-\frac{z^2}{2}\right\} dz + \frac{1}{\sqrt{2\pi}} \int_0^{\infty} \exp\left\{-\frac{z^2}{2}\right\} dz = \\ &\stackrel{\substack{\uparrow \\ \text{Half Gaussian} \\ \text{integral}}}{=} \frac{1}{\sqrt{2\pi}} \int_0^x \exp\left\{-\frac{z^2}{2}\right\} dz + 0.5 = \text{erf}\{x\} - 0.5 \end{aligned}$$

כשעשינו שימוש בקשרים

$$\frac{1}{\sqrt{2\pi}} \int_0^{\infty} \exp\left\{-\frac{z^2}{2}\right\} dz = 0.5 \quad \text{ו-} \quad \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left\{-\frac{z^2}{2}\right\} dz = 1$$

הפונקציה $\text{erf}(x)$ מקבלת ערך 0.5 עבור $x \rightarrow \infty$, ערך 0 עבור $x=0$, וערך -0.5 עבור $x \rightarrow -\infty$. לכן, עבור y קבוע (חתך בתמונה), הערכים כפונקציה של x יהיו 0.5 ב- $x=0$, משם דעיכה לאפס כש- x קטן לערכים שליליים, ועליה ל-1 עבור ערכים חיוביים הולכים וגדלים, ממש כפי שצפינו. הציור הבא מראה את אות המקור, התגובה להלם, התוצאה וחתך אופייני שלה.



תמונת המקור (למעלה משמאל), תוצאת הקונבולוציה (למעלה באמצע), התגובה להלם (למעלה מימין). למטה – חתך אופקי של תמונת התוצאה.

ציור זה נעשה ע"י פקודות ה-Matlab

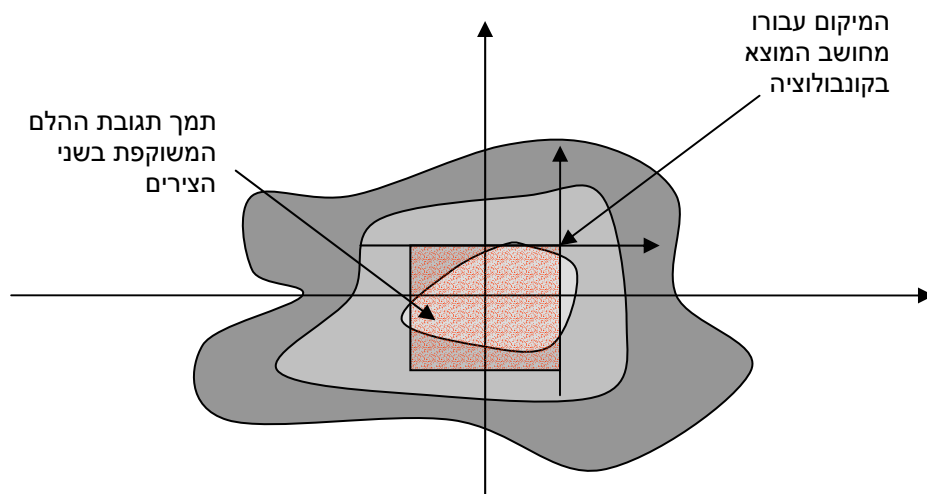
```
s=[zeros(500,250),ones(500,250)]; figure(1); imagesc(s);
colormap(gray(256)); axis off; axis image;
[x,y]=meshgrid(-5:0.1:5);
h=1/(2*pi)*exp(-(x.^2+y.^2)/2)/100; g=conv2(s,h,'valid');
figure(2); imagesc(g); axis off; colormap(gray(256));
axis image; figure(3); plot(g(200,:)); grid on;
figure(4); imagesc(h); axis off; colormap(gray(256));
axis image;
```

כנושא אחרון בדיון זה נדון בתכונות של מערכות ליניאריות וקבועות במקום, כשאנו מרחיבים את אותן תכונות שהוצגו למערכות הפועלות על אותות בחד-מימד:

- שחלוף – מערכות אלו משתחלפות, כלומר סדר הפעלתן לא משנה:

$$\begin{aligned}
 g(x, y) &= h_2(x, y) \otimes (h_1(x, y) \otimes s(x, y)) = \\
 &= (h_1(x, y) \otimes h_2(x, y)) \otimes s(x, y) = \\
 &= (h_2(x, y) \otimes h_1(x, y)) \otimes s(x, y)
 \end{aligned}$$

- סיבתיות – ניתן לגלות סיבתיות של מערכת זו ע"י בחינת תגובתה להלם אשר חייבת להיות סיבתית גם כן (שונה מאפס רק ברביע החיובי). קל לקרות זאת גיאומטרית כיוון שבתמונת המראה בשני הצירים, מוצא המערכת יהיה תלוי רק בערכי עבר, כמתואר בציור.



סיבתיות בקונבולוציה דו-ממדית – אנו רואים כי רק ערכי עבר משמשים לחישוב.

- חשוב להבהיר כי סיבתיות הרבה פחות נחוצה בדו-מימד כיוון שאין זה ציר זמן. בטיפול באותות דו-ממדיים מקובל לוותר על תכונה זו.
- יציבות – מערכת יציבה BIBO אם יתקיים

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |h(x, y)| dx dy < \infty$$

- פרידות (ספרביליות) – זוהי תכונה חדשה שקשורה לעובדה שהעיבוד נעשה כעת בדו-מימד. מערכת תיחשב פרידה אם הפעלתה יכולה להיעשות ע"י שרשרת של שתי מערכות – האחת פועלת בציר x והאחרת בציר y. לדוגמה, המערכת

$$g(x, y) = f(x, y) + f(x, y + 1) + f(x + 1, y) + f(x + 1, y + 1)$$

היא מערכת פרידה כיוון שניתן להפעילה ע"י שתי מערכות מהצורה

$$\begin{aligned}
 g_1(x, y) &= f(x, y) + f(x, y + 1) \\
 g_2(x, y) &= g_1(x, y) + g_1(x + 1, y)
 \end{aligned}$$

הפעולה הראשונה מערבת מידע רק בציר ה- y , ומוצאה מזון למערכת שנייה שמערבת מידע רק בציר x . בהינתן תגובת ההלם של מערכת, אם פונקציה זו פרידה, כלומר אם מתקיים $h(x, y) = h_1(x)h_2(y)$, נקבל כי המערכת פרידה כיוון ש:

$$\begin{aligned} g(x, y) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\hat{x}, \hat{y}) h(x - \hat{x}, y - \hat{y}) d\hat{x} d\hat{y} = \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\hat{x}, \hat{y}) h_1(x - \hat{x}) h_2(y - \hat{y}) d\hat{x} d\hat{y} = \\ &= \int_{-\infty}^{\infty} h_1(x - \hat{x}) \left[\int_{-\infty}^{\infty} f(\hat{x}, \hat{y}) h_2(y - \hat{y}) d\hat{y} \right] d\hat{x} = \int_{-\infty}^{\infty} h_1(x - \hat{x}) g_1(\hat{x}) d\hat{x} \end{aligned}$$

אנו רואים כי הפעלת המערכת כללה פעולה על ציר y לבד, ואחריה הפעלת מערכת שנייה על ציר x .

התמרת הפורייה לאותות בדו-מימד

נניח כעת כי בידינו האות הספרבילי ההרמוני

$$s(x, y) = \exp\{j2\pi(f_x x + f_y y)\} = \exp\{j2\pi f_x x\} \cdot \exp\{j2\pi f_y y\}$$

זהו אות מרוכב הרמוני (מחזורי) בעל תדרים f_x ו- f_y – כלומר ביחידת מרחק אחת בצירים x ו- y הוא משלים f_x ו- f_y מחזורים שלמים בהתאמה. נשים לב לכך שזהו אות פריד (מוצג כמכפלה של אות ב- x ואות ב- y). נזין אות זה למערכת ליניארית וקבועה במקום המאופיינת ע"י תגובתה להלם $h(x, y)$ ונקבל

$$\begin{aligned} g(x, y) &= h(x, y) \otimes s(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(\hat{x}, \hat{y}) s(x - \hat{x}, y - \hat{y}) d\hat{x} d\hat{y} = \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(\hat{x}, \hat{y}) \exp\{j2\pi[f_x(x - \hat{x}) + f_y(y - \hat{y})]\} d\hat{x} d\hat{y} = \\ &= \exp\{j2\pi[f_x x + f_y y]\} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(\hat{x}, \hat{y}) \exp\{-j2\pi[f_x \hat{x} + f_y \hat{y}]\} d\hat{x} d\hat{y} \end{aligned}$$

קיבלנו כי אות הכניסה מועבר למוצא כמות שהוא כשהוא מוכפל בגורם

$$H(f_x, f_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y) \exp\{-j2\pi[f_x x + f_y y]\} dx dy = F_{2D}\{h(x, y)\}$$

וזהו התמרת הפורייה הדו-ממדית על תגובת ההלם של המערכת. התמרה זו הינה פעולה פרידה הניתנת לרישום כ:

$$H(f_x, f_y) = \int_{-\infty}^{\infty} \exp\{-j2\pi f_y y\} \left[\int_{-\infty}^{\infty} h(x, y) \exp\{-j2\pi f_x x\} dx \right] dy =$$

$$= \int_{-\infty}^{\infty} \exp\{-j2\pi f_y y\} g(f_x, y) dy$$

ברישום הנ"ל אנו מתייחסים ל-x מסויים (כפרמטר) ובכך מתייחסים לחתך של הפונקציה. עבור חתך זה אנו מבצעים באינטגרל הפנימי התמרת פורייה חד-ממדית. כך אנו עושים לכל חתך (לכל ערכי x). בשלב שני אנו מבצעים על תוצאת התהליך הקודם (שסומנה כפונקציה g של המשתנים y והתדר ב-x) פעולת התמרת פורייה חד-ממדית הפעם על ציר y.

מהפירוק הנ"ל גם ניתן להראות (ודי בקלות) כי ההתמרה ההפוכה נתונה כשתי התמרות חד-ממדיות הפוכות הפועלות גם הן כמכונה פרידה,

$$h(x, y) = F_{2D}^{-1}\{H(f_x, f_y)\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H(f_x, f_y) \exp\{j2\pi[f_x x + f_y y]\} df_x df_y$$

משמעות התמרת הפורייה כאן דומה למשמעותה במקרה החד-מימדי. כל אות $s(x, y)$ ניתן לפירוק פורייה של אותות הרמוניים דו-ממדיים בתדרים (f_x, f_y) . מסנן $h(x, y)$ הפועל על אות זה לוקח כל הרמוניה כזו ומכפיל אותה במספר $H(f_x, f_y)$, וכך מתקבל המוצא כצירוף כל הרמוניות אלו עם הגברן/ניחותן. אנו לא נחזור על התכונות של התמרת הפורייה – מרביתן המכריע דומות, וענייננו הרי אינו באותות ברצף. נעבור לפיכך למשפט הדגימה.

דגימה ושחזור בדו-מימד

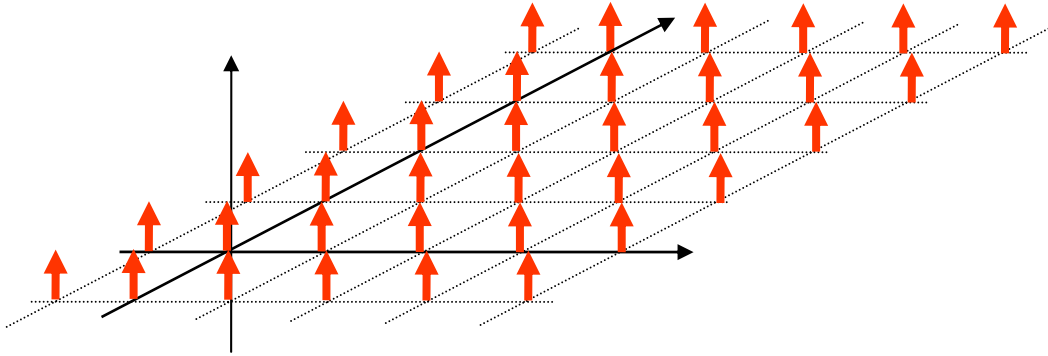
נתון לנו אות $s(x, y)$, ורצוננו לדגום אותו במרווחים d_x, d_y אחידים בכל ציר. על מנת להמירו לאות דיסקרטי דו-מימדי, פעולת הדגימה שנציע היא

$$s[n, m] = s(x, y) \Big|_{x=nd_x, y=md_y} = s(nd_x, md_y)$$

גם כאן ניתן לשאול: באילו תנאים ניתן יהיה לשחזר את $s(x, y)$ מתוך דגימותיו $s[n, m]$ ללא אובדן? נעשה שימוש בתיאור מתמטי נוח יותר לפעולת הדגימה – דגימה ע"י מכפלה ב-"מיטת פאקיר" שתהווה הכללה של פונקצית השה. נגדיר את הפונקציה

$$\text{III}_{d_x, d_y}(x, y) = \sum_{n, m} \delta(x - nd_x, y - md_y) = \sum_n \delta(x - nd_x) \cdot \sum_m \delta(y - md_y)$$

פונקציה זו מתוארת בציור הבא. למעשה, פונקציה זו ניתנת לתיאור כפרידה עם שתי פונקציות שהן חד-ממדיות, כמתואר.



פונקציה שהיא דו-ממדית.

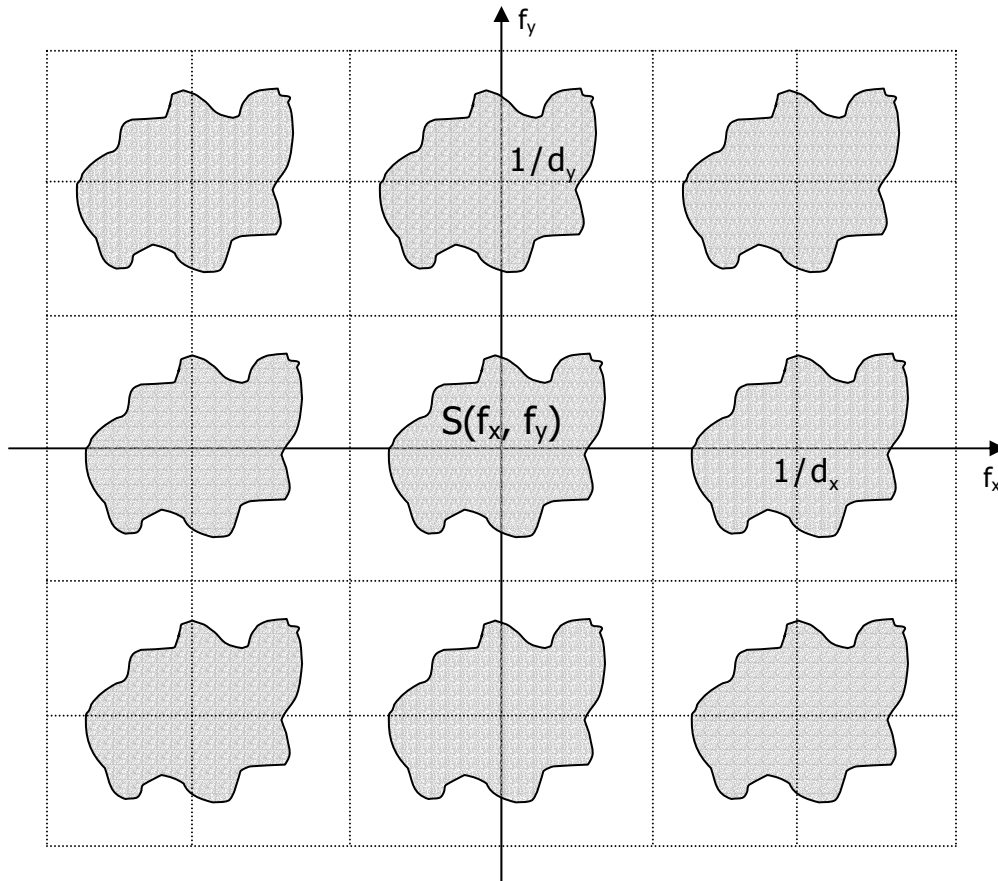
נקבע כי האות נדגם ע"י המכפלה

$$\tilde{S}(x, y) = s(x, y) \cdot \text{III}_{d_x, d_y}(t) = s(x, y) \cdot \sum_{n, m} \delta(x - nd_x, y - md_y)$$

כיוון שהאות בנוי ממכפלה של שני אותות רציפים בעקרון, התמרת הפורייה שלו היא קונבולוציה של ההתמרות. לכן

$$\begin{aligned} \tilde{S}(f_x, f_y) &= F\{\tilde{S}(x, y)\} = F\left\{s(x, y) \cdot \sum_{n, m} \delta(x - nd_x, y - md_y)\right\} = \\ &= F\{s(x, y)\} \otimes F\left\{\sum_{n, m} \delta(x - nd_x, y - md_y)\right\} = \\ &= \frac{1}{d_x d_y} S(f_x, f_y) \otimes \sum_{n, m} \delta\left(f_x - \frac{n}{d_x}, f_y - \frac{m}{d_y}\right) = \\ &= \frac{1}{d_x d_y} \sum_{n, m} S\left(f_x - \frac{n}{d_x}, f_y - \frac{m}{d_y}\right) \end{aligned}$$

השתמשנו בעובדה שהתמרה על פונקציה הדו-ממדית נתונה כפונקציה דומה עם מרווחים שהינם $1/d_x$ ו- $1/d_y$. לכן $S(f_x, f_y)$ משוכפלת שוב ושוב במרווחים קצובים. הציור הבא ממחיש תופעה זו.



תופעת שכפול בתדר כתוצאה מדגימה בדו-מימד.

כאשר $S(f_x, f_y)$ מתאפס מחוץ לתחום מלבני נאות, נקבל כי השכפולים לא רומסים זה על זה, ואז שחזור מדויק אפשרי. אם מתקיים

$$S(f_x, f_y) = \begin{cases} S(f_x, f_y) & |f_x| \leq F_x^{\text{Max}} \text{ \& } |f_y| \leq F_y^{\text{Max}} \\ 0 & \text{otherwise} \end{cases}$$

האות חסום פס. אזי, על מנת שהשכפולים לא יעלו זה על זה נדרוש

$$F_x^{\text{Max}} < \frac{1}{2d_x} \Rightarrow d_x < \frac{1}{2F_x^{\text{Max}}}$$

$$F_y^{\text{Max}} < \frac{1}{2d_y} \Rightarrow d_y < \frac{1}{2F_y^{\text{Max}}}$$

הדרישה הזו קובעת שעלינו לדגום מספיק צפוף כדי לכלול כמות מספקת של דגימות לייצוג מלא של האות. עוד רואים כי הדרישה על מרווחי הדגימה עשויה להיות שונה מציר x לציר y .

באשר לשחזור, בהינתן האות הדגום $\tilde{s}(x, y)$ (שכל דגימה בו היא פונקצית הלים), זהו אות במקום רציף הניתן לסינון. ע"י הפעלת מסנן מעביר נמוכים (LPF) מהצורה

$$H(f_x, f_y) = \begin{cases} d_x d_y & |f_x| \leq F_x^{\text{Max}} \text{ \& } |f_y| \leq F_y^{\text{Max}} \\ 0 & \text{otherwise} \end{cases}$$

על אות זה נקבל בדיוק את אות המקור, כיוון שהסינון הינו מכפלה בתדר, וע"י ההכפלה במסנן הנ"ל אנו מבטלים את כל השכפולים למעט זו הראשית, וכיוון שהיא בדיוק ההתמרה של אות המקור, הרי שביצענו בשלמות פעולה הפוכה לפעולת הדגימה וקיבלנו שחזור. מבחינה מעשית, סינון זה ייעשה ע"י

$$\begin{aligned} \hat{s}(x, y) &= F^{-1} \{ H(f_x, f_y) F \{ \tilde{s}(x, y) \} \} = \tilde{s}(x, y) \otimes h(x, y) = \\ &= \int \int \sum_{n,m} s[n, m] \delta(\hat{x} - nd_x, \hat{y} - md_y) h(x - \hat{x}, y - \hat{y}) dx = \\ &= \sum_{n,m} s[n, m] h(x - nd_x, y - md_y) \end{aligned}$$

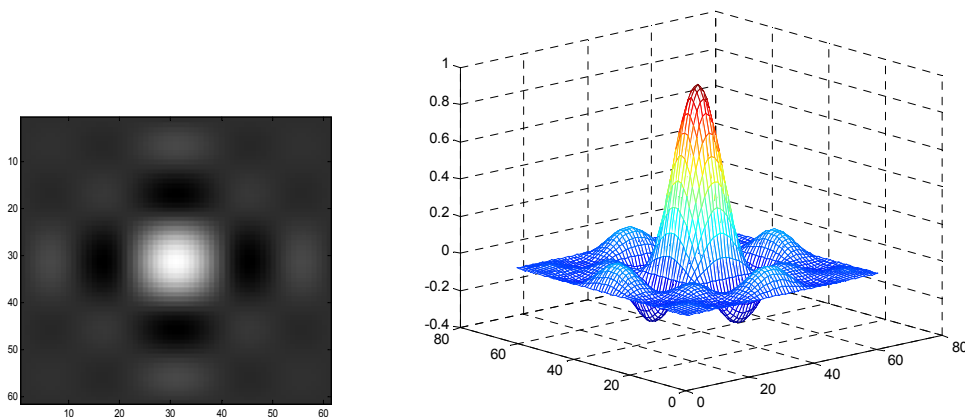
ההתמרה ההפוכה (מהתדר למקום) של המסנן $H(f_x, f_y)$ תהיה

$$\text{sinc}\{nx/d_x\} \text{sinc}\{ny/d_y\} = \frac{\sin\{nx/d_x\}}{nx/d_x} \cdot \frac{\sin\{ny/d_y\}}{ny/d_y}$$

ולכן השחזור ייעשה ע"י הנוסחה

$$\hat{s}(x, y) = \sum_{n,m} s[n, m] h(x - nd_x, y - md_y) = \sum_{n,m} s[n, m] \text{sinc}\left(n \frac{x - nd_x}{d_x}, n \frac{y - md_y}{d_y}\right)$$

גם כאן יוצא כי עלינו לבצע אינטרפולציה עם פונקציה שדורשת שימוש בשכנים רחוקים מאוד (עד אינסוף) על מנת לבצע אינטרפולציה של נקודת ערך לא שלמה. פונקצית ההשפעה הדו-ממדית (sinc דו-ממדי) מתוארת בציור הבא:



פונקצית ה-sinc איתה יש לבצע את פעולת האינטרפולציה בשחזור.

ציור זה נעשה ע"י פקודות ה-Matlab

```
t=-3.00001:0.1:3; SS=sin(pi*t)./(pi*t); figure(1);  
imagesc(SS'*SS); colormap(gray(256));  
figure(2); mesh(SS'*SS);
```

הערה: אנו רואים כאן התלבטות באשר לדרך הצגתו של אות דו-מימדי – ניתן לתארו כמשטח בתלת-מימד או כתמונה.

חשוב להבהיר כי לעתים רחוקות אנו נדרשים לשחזר רצף בדו-מימד מאת דגום. באות דיבור למשל, השחזור מדגימות חיוני כי אם נרצה להקשיב לאות, חובה להזין לרמקול אות רציף. לעומת זאת, בתמונות הצגת תמונה על מסך נעשית במבנה דגום. מדוע אם כך נושא השחזור חשוב? מכיוון שבכל פעולה גיאומטרית על תמונה אנו נדרשים לחשב את ערך התמונה במקומות (x,y) שאינם על הדגימות הנתונות. לדוגמה, אם נתונה תמונה ורצוננו להגדילה פי 4 בכל ציר לשם הדפסתה, עלינו להכניס בין כל רביעיית דגימות (בשני שורות עוקבות) עוד 18 ערכים במרווחים קצובים, ואלו ייבנו ע"י אינטרפולציה. כדוגמה שנייה, אם רצוננו לסובב תמונה ב- 10° , כל הפיקסלים בתמונה שתיווצר ייפלו במקומות שבין הפיקסלים הקיימים, ושוב נדרשת אינטרפולציה.

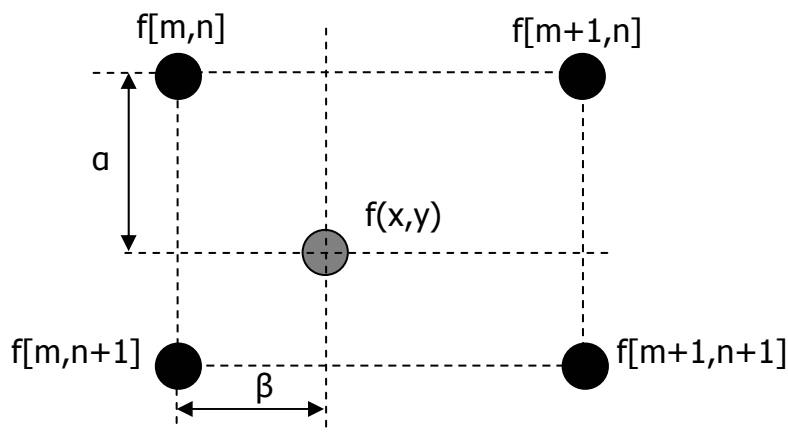
שחזור מעשי לא יכול להיעשות ע"י פונקציה זו. האלטרנטיבות המוכרות הן ה-Zero Order Hold ו-First Order Hold כפי שהוצגו בדיון בחד-מימד, כשהן מיושמות באופן ספרבילי. במקרה של ה-ZOH גישה זו מובילה לשיטת השכן הקרוב, לפיה הערך בנקודה (x,y) באות הדו-מימדי המשוחזר ייקבע כערך הדגימה הקרובה ביותר הקיימת. זהו השחזור הפשוט ביותר והוא מציע החלפת כל דגימה בפונקציה מלבנית בגודל $d_x \times d_y$ ממורכזת סביב מרכז, בעלת גובה קבוע ששווה לערך הדגימה. אם לדוגמה תהליך זה מבוצע לאינטרפולציה לשם הגדלת האות (התמונה) פי 2 בכל ציר, נקבל כי פשוט עלינו לשכפל כל שורה וכל עמודה פעמיים – תהליך הקרוי Pixel Replication, או Nearest Neighbor, כיוון שהערכים החסרים ממולאים ע"י העתקת השכן הקרוב ביותר.

במקרה של ה-FOH יישום זה יוביל לאינטרפולציה בי-ליניארית הקובעת את הערך בנקודה (x,y) כממוצע משוקלל של ארבעת השכנים סביבה, וזוהי

אפשרות טובה יותר. ממש כמו שראינו במקרה החד-מימדי, מבוצעת כאן קונבולוציה עם פונקציה דו-מימדית הנבנית ממכפלת שתי פונקציות משולשות חד-מימדיות סימטריות סביב הראשית, הכופלות זו את זו ליצירת פונקציה דו-מימדית ספרבילית. הפונקציה המשולשת עבור הציר האופקי מקבלת גובה 1 בראשית ו-0 בנקודות $\{-d_x, +d_x\}$. שימוש בפונקציה זו לאינטרפולציה קרוי Bilinear Interpolation, והשימוש בו מאוד פופולרי. עבור 4 דגימות סמוכות $\{f[m,n], f[m+1,n], f[m,n+1], f[m+1,n+1]\}$ ניתן להראות כי ערך המצוי בתוך הריבוע נתון ע"י

$$f(x,y) = \begin{bmatrix} 1-\alpha & \alpha \end{bmatrix} \begin{bmatrix} f[m,n] & f[m,n+1] \\ f[m+1,n] & f[m+1,n+1] \end{bmatrix} \begin{bmatrix} 1-\beta \\ \beta \end{bmatrix}$$

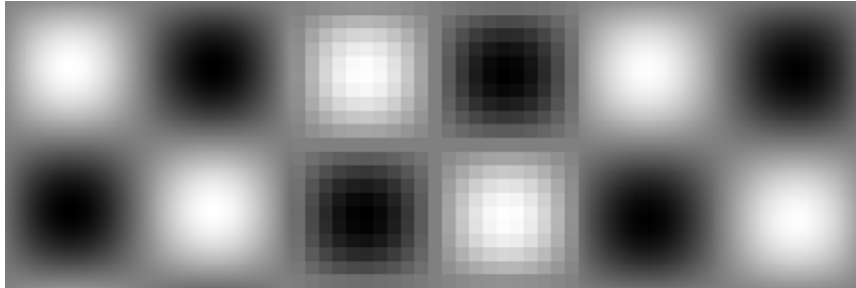
כאשר α, β הם ערכים בין 0 ל-1 הנותנים את מיקום ה- x וה- y יחסית לאורך הדגימה, כפי שמתואר בציור הבא.



פעולת האינטרפולציה הביליניארית.

בציור הבא מתוארת תמונת סינוס דו-מימדי כפי שהוא במקור. פונקציה זו היא $f(x,y) = \sin(x)\sin(y)$ והיא מוצגת עבור הערכים $-3 \leq x,y \leq 3$. פונקציה זו נדגמה במרווח של 0.3. מחזור אחד של הפונקציה בכל ציר הינו באורך $2\pi = 6.28$, ולכן יש לנו יותר מ-20 דגימות לכל מחזור – הרבה מעבר לדרוש לשם שחזור מדויק (הכלל הוא – יותר משתי דגימות למחזור). בציור מתוארים השחזור מבוסס ה-NN והאינטרפולציה הביליניארית. תוצאת הגדלת תמונה ע"י אינטרפולציה ביליניארית מביאה לשיפור עצום. התוצאה עבור דוגמה זו מצוינת

כיוון שהתמונה המקורית חלקה מאוד, ואינטרפולציה זו ממלאת את החסר בקווים אלכסוניים הנותנים תחושה של רצף.



פונקציית סינוס דו-ממדית מקורית (משמאל), ואינטרפולציה שלה ע"י שיטת השכן הקרוב (אמצע) ואינטרפולציה ביליניארית (ימין).

ציור זה נעשה ע"י פקודות ה-Matlab

```
[x,y]=meshgrid(-3:0.03:3.29);z_true=sin(x).*sin(y);  
z_sample=z_true(1:10:end,1:10:end);  
z_inc1=imresize(z_sample,10,'nearest');  
z_inc2=imresize(z_sample,10,'bilinear');  
imagesc([z_true,z_inc1,z_inc2]); axis image;  
colormap(gray(256));
```

תופעת הקיפול ומשמעותה

תופעת הקיפול – Aliasing - הינה פלישה של השכפולים בתדר כתוצאה מדגימה, זה לתוך זה, וברור כי במקרה כזה לא ניתן לשחזר את האות המקורי ללא מידע נוסף. תופעת הקיפול מתרחשת כאשר מרווח הדגימה גדול מדי, ואינו מקיים את דרישת הסף שהזכרה בסעיף קודם. גם בדו-מימד, בדומה לחד-מימד, תהיה תופעה של התחזות בה תדרים גבוהים יראו כנמוכים.

נוכל להימנע מתופעת הקיפול אם האות שלנו חסום פס והדגימה נאותה. שאלה מעניינת היא האם תמונות (אותות דו-ממדיים מיוחדים בהם יש לנו עניין) אמנם חסומות פס? מסתבר שלא! האם פירוש הדבר שנסבול מתופעת הקיפול? ואם כן, כיצד תתבטא אותה תופעה בתמונות אלו אחר דגימה? כמה גדול הנזק שייגרם (אם בכלל)? ובכן, לאותות מעשיים לא ניתן להניח התנהגות חסומת תדר. אולם, בהחלט ניתן להניח כי העוצמה של התמרת הפורייה דועכת ואפילו מהר כשמרחקים מראשית הצירים (ה-DC). לכן, גם אם חלה פלישה של

השכפולים אין פירוש הדבר נזק גדול בהכרח, כיוון שפלישתו של זנב ההתמרה תשנה רק במקצת. אם בכל זאת הזנב אינו כה חלש, השפעתו תתבטא בתמונה הדגומה, ותתקבלנה תופעות המוכרות שהיבטים שונים של תופעת הקיפול.

בציור הבא נתונה תמונה דיסקרטית אשר נדגמה בצפיפות מספקת (כלומר, אין בה תופעת קיפול), ולצידה תת-דגימות שלה ע"י דילול פשוט ביחסים שלמים. לדוגמה, לקיחת כל פיקסל שני מכל שורה שנייה יוצר תמונה קטנה פי ארבע המתייחס למרווח דגימה כפול. התמונות השונות הינן בגודל שונה, אך מוצגות בגודל אחד ע"י מתיחה פשוטה הגורמת לשכפול פיקסלים. כפי שניתן לראות על פניו של הבית בצד שמאל של התמונה, נוצר תדר שונה לכל מרווח – אפקט קיפול קלאסי. אפקט נוסף ניתן לאבחנה בגדר, שם ניכר כי אזורים יוצאים במרווח מסוים מטושטשים ובמרווחים גדולים יותר חוזרים להיות פעילים – גם אפקט של קיפול. מלבד שתי תופעות לאה ניכרת גם תופעת שינון של שפות חדות המצויות בזוית אלכסונית כמו במגדלור ובגגות.

תופעת הקיפול היא תופעה מטרידה כשהיא בעוצמה חזקה. על מנת להימנע ממנה בתהליך הדגימה, נדרש להפעיל מסנן מעביר נמוכים (LPF) אשר יהפוך את התמונה הנתונה לחסומת תדר באופן אפקטיבי, ורק לאחר מכן לבצע את הדילול. הציור הבא ממחיש אפקט זה. הפעם חזרנו על הניסוי הקודם כשלפני פעולת הדילול ביצענו סינון. במסגרת זו לא נאמר מהו המסנן שהופעל – דיון בנושא תכנון מסנן נאות יינתן במסגרת פרק אחר. נאמר רק כי ברור כי על מסנן זה להעביר טוב ככל האפשר את התדרים הנמוכים עד סף מסוים, ולחסום טוב ככל האפשר מתדר סף זה. כמו כן ברור כי בשל אופייה של הדגימה, המסנן הנדרש ניתן לתכנון כמסנן ספרבילי, כשתדר הסף בכל ציר עשוי להיות שונה. כפי שניתן לראות כי כל התופעות שהוזכרו נעלמו או דוכאו בצורה משמעותית.



המחשת תופעת הקיפול: למעלה בצד ימין – תמונת מקור.
למעלה בצד שמאל – דגימה במרווח כפול. למטה בצד שמאל – דגימה במרווח גדול פי 3.
למטה בצד ימין – דגימה במרווח גדול פי ארבע.



חזרה על הציור הקודם כשלפני פעולת הדילול מבוצע מסנן LPF נאות לקיצוץ התדרים הגבוהים

עיבוד תמונות בעזרת מחשב (236327)

הרצאה מס' 6

אותות ומערכות 5 – אותות בדידים בדו-מימד

עיבוד אותות דו-ממדיים דגומים

מהאות ברצף $s(x,y)$ התקבל האות הדגום $s[n,m]$. לרוב אף לא נצטרך לטרוח ונקבל את האות (תמונה!) כשהוא כבר דגום. בדיון שיובא להלן נתאר מערכות על אותות כאלה, נראה כיצד מופעלת קונבולוציה, וכיצד מיושמת התמרת הפורייה.

מערכת מוגדרת כקבועה במקום אם לכל אות כניסה מתקבל כי הזזת הכניסה יוצרת ההזזה דומה ביציאה:

$$\forall s[n,m], n_0, m_0 : g[n,m] = T\{s[n,m]\} \\ \Rightarrow g[n-n_0, m-m_0] = T\{s[n-n_0, m-m_0]\}$$

מערכת מוגדרת כליניארית אם היא מקיימת את התכונה הבאה - לכל שני אותות $s_1[n,m]$ ו- $s_2[n,m]$ ולכל שני סקלרים a ו- b , יתקיים כי

$$T\{a \cdot s_1[n,m] + b \cdot s_2[n,m]\} = a \cdot T\{s_1[n,m]\} + b \cdot T\{s_2[n,m]\}$$

ניתן מספר דוגמאות להמחיש זאת:

- המערכת $g[n,m] = s[n,m] + 3s[n-1, m+2]$ היא מערכת ליניארית וקבועה במקום. למעשה זו עם זיכרון כיוון שהיא עושה שימוש בערכי הכניסה במקומות שונים ממקום המוצא.
- המערכת $g[n,m] = s[n,m] + \sqrt{nm} \cdot s[n-1, m+1]$ תלויה במקום, אך היא עדיין ליניארית, כי היא בונה את מוצאה כצירוף ליניארי של אות הכניסה.
- המערכת $g[n,m] = 2s[n,m] + 3$ אינה ליניארית, אך היא קבועה במקום. קל לראות זאת לפי ההגדרה.

- המערכת $g[n,m] = \sum_{j=-3}^3 \sum_{k=-3}^3 w_{j,k} \cdot s[n-j, m-k]$ היא מערכת ליניארית וקבועה במקום. זוהי מערכת שמבצעת ממוצע משוקלל של 49 הדגימות סביב נקודת האמצע בה מחושב המוצא.
- המערכת $g[n,m] = s[n,m] + 3s[n-0.5, m]$ לא מוגדרת! ברגע שעברנו לאות דגום, אין מקום שאינו אינדקס שלם.
- המערכת $g[n,m] = s[n,m]s[n-1, m]$ אינה ליניארית אך היא קבועה במקום.

ממש כשם שעשינו זאת לרצף, יש להגדיר פונקצית הלים דרכה נוכל לנתח התנהגות מערכות ליניאריות וקבועות במקום. פונקצית ההלים הדיסקרטית הדו-ממדית היא

$$\delta_{2D}[n,m] = \begin{cases} 1 & n = m = 0 \\ 0 & \text{אחרת} \end{cases}$$

בדומה להתנהגות פונקצית ההלים ברצף, קיימת התכונה הבאה: עבור אות $s[n,m]$ כלשהו ופונקצית ההלים הנ"ל, יתקבל

$$\sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \delta_{2D}[k,j] s[n+k, m+j] = \sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \delta_{2D}[n-k, m-j] s[k,j] = s[n,m]$$

וקל לראות נכונות נוסחה זו, כי פונקצית ההלים קיימת רק באחד האינדקסים ומתאפסת לכל האחרים. משמעות משוואה זו שהאות $s[n,m]$ יכול להיות מתואר כצירוף ליניארי של אותות הלים מוזות $\delta_{2D}[n-k, m-j]$ כשהן משוקללות ע"י ערכי האות עצמו $s[k,j]$.

אם נזין למערכת זו הלים $\delta_{2D}[n,m]$, נקבל מוצא אשר ייקרא התגובה להלים ויסומן $h[n,m]$, כלומר $T\{\delta_{2D}[n,m]\} = h[n,m]$. נניח כעת כי הזנו אות כלשהו אחר $s[n,m]$ למערכת T . לכן, כיוון שהאות יכול להיות מתואר כסכום הלים נקבל

$$\begin{aligned} g[n,m] &= T\{s[n,m]\} = T\left\{\sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \delta_{2D}[n-k, m-j] s[k,j]\right\} = \\ &= \sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} T\{\delta_{2D}[n-k, m-j]\} s[k,j] = \\ &= \sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} h[n-k, m-j] s[k,j] \end{aligned}$$

כאן השתמשנו בליניאריות של המערכת כשהחלפנו סדר פעולת המערכת עם הסכימה. השתמשנו בקביעות במקום כשהנחנו שפעולת המערכת על הלים מוזז היא תגובת הלים מוזזת. קיבלנו שוב כי כל מערכת ליניארית וקבועה במקום ניתנת לאפיון ע"י תגובתה להלים, ובהינתן אות כניסה כלשהו, תגובתה ההלים מאפשרת לנו חישוב מוצאה. המשוואה שקיבלנו אינה אלא פעולת הקונבולוציה הדיסקרטית, אותה נסמן כ:

$$\begin{aligned} g[n,m] &= T\{s[n,m]\} = \sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} h[n-k, m-j] s[k,j] = \\ &= \sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} h[k,j] s[n-k, m-j] = h[n,m] \otimes s[n,m] \end{aligned}$$

חישוב קונבולוציה זו דומה לחישוב זו הרציפה, אלא שסכימה דיסקרטית מחליפה את האינטגרל, וכל התכונות שהוצגו בעבר על פעולת הקונבולוציה תקפות גם כאן.

התמרת פורייה דו-ממדית על אותות דגומים

בדומה לנעשה עבור הרצף, נגדיר את האותות שחולפים דרך המערכת ללא שינוי. אותות אלו יהיו $s[n,m] = \exp\{j2\pi n\theta_x\} \exp\{j2\pi m\theta_y\}$. הזנת אות כזה למערכת ליניארית וקבועה בזמן המאופיינת ע"י תגובתה להלים $h[n]$ תיתן

$$\begin{aligned} g[n,m] &= \sum_{k=-\infty}^{\infty} \sum_{p=-\infty}^{\infty} h[k,p] s[n-k, m-p] = \\ &= \exp\{j2\pi n\theta_x\} \exp\{j2\pi m\theta_y\} \sum_{k=-\infty}^{\infty} \sum_{p=-\infty}^{\infty} h[k,p] \exp\{-j2\pi(k\theta_x + p\theta_y)\} = \\ &= H(\theta_x, \theta_y) \cdot \exp\{j2\pi(n\theta_x + m\theta_y)\} \end{aligned}$$

וכך קיבלנו הגדרה של התמרת פורייה לאות בזמן בדיד, ע"י

$$\begin{aligned} S(\theta_x, \theta_y) &= F\{s[n,m]\} = \sum_{n,m} s[n,m] \cdot \exp\{-j2\pi(n\theta_x + m\theta_y)\} = \\ &= \sum_n \exp\{-j2\pi n\theta_x\} \sum_m s[n,m] \cdot \exp\{-j2\pi m\theta_y\} \end{aligned}$$

אנו רואים התמרת הפורייה הזו מחזורית, ואורך מחזוריה הוא 1 בשני ציריה. לכן נתעניין בתחום $-0.5 \leq \theta_x, \theta_y \leq 0.5$ בלבד. מעבר לכך אנו רואים כי התמרה זו פרידה וניתנת ליישום כהתמרת פורייה חד-ממדית על כל שורה (או עמודה) במערך הדגימות, וביצוע התמרת פורייה חד-ממדית על כל עמודה (או שורה)

של תוצאת התהליך הקודם. לכן, קל לבנות את ההתמרה ההפוכה מתוך דרך ביצועה בחד-מימד. בהינתן $S(\theta_x, \theta_y)$ נחזור לאות המקור $s[n, m]$ ע"י

$$s[n, m] = \int_{-1/2}^{1/2} \int_{-1/2}^{1/2} S(\theta_x, \theta_y) \exp\{j2\pi(n\theta_x + m\theta_y)\} d\theta_x d\theta_y$$

להתמרה שהוגדרה יש את מרבית התכונות שהצגנו להתמרה הרציפה. חשוב להבהיר כי אין זה במקרה – ההתמרה לאות הדגום אינה אלא מקרה פרטי של ההתמרה הכללית הרציפה.

במסגרת זו נציג מספר תכונות חשובות:

- הגדרה - עבור האות $g[n, m]$ התמרת הפורייה נתונה ע"י

$$G(\theta_x, \theta_y) = F\{g[n, m]\} = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} g[n, m] \exp\{-j2\pi(n\theta_x + m\theta_y)\}$$

וההתמרה חזרה ע"י

$$g[n, m] = F^{-1}\{G(\theta_x, \theta_y)\} = \int_{-1/2}^{1/2} \int_{-1/2}^{1/2} G(\theta_x, \theta_y) \exp\{j2\pi(n\theta_x + m\theta_y)\} d\theta_x d\theta_y$$

- ליניאריות – התמרת הפורייה ליניארית, כלומר, עבור כל שני אותות $s[n, m]$ ו- $g[n, m]$, ושני סקלרים a ו- b , מתקיים כי

$$F\{a \cdot s[n, m] + b \cdot g[n, m]\} = a \cdot F\{s[n, m]\} + b \cdot F\{g[n, m]\}$$

- ממוצע – ממוצע האות $g[n, m]$ אינו אלא $G(0)$:

$$G(0, 0) = F\{g[n, m]\} \Big|_{\theta=0} = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} g[n, m]$$

- ממשיות וסימטריות – בהינתן לנו אות $g[n, m]$ ממשי טהור (כלומר ללא מרכיב מדומה), להתמרת הפורייה שלו מבנה סימטרי צמוד,

$$\begin{aligned} G(\theta_x, \theta_y) &= \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} g[n, m] \exp\{-j2\pi(n\theta_x + m\theta_y)\} \\ \overline{G(\theta_x, \theta_y)} &= \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} g[n, m] \exp\{-j2\pi(n\theta_x + m\theta_y)\} = \\ &= \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} g[n, m] \exp\{j2\pi(n\theta_x + m\theta_y)\} = G(-\theta_x, -\theta_y) \end{aligned}$$

אם האות הנתון לנו $g[n, m]$ גם ממשי וגם סימטרי יתקבל כי ההתמרה ממשית וסימטרית אף הוא.

- שימור מכפלה פנימית – עבור שני אותות $g[n,m]$ ו- $s[n,m]$, ותוצאות התמרתם $G(\theta_x, \theta_y)$ ו- $S(\theta_x, \theta_y)$, מכפלה פנימית בין שתי פונקציות בזמן מוגדרת ע"י

$$\langle g[n,m], s[n,m] \rangle_{n,m} = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} g[n,m] \cdot \overline{s[n,m]}$$

מכפלה פנימית בתדר של פונקציות ההתמרה תוביל לאותה תוצאה:

$$\begin{aligned} \langle G(\theta_x, \theta_y), S(\theta_x, \theta_y) \rangle_{\theta_x, \theta_y} &= \int_{-0.5}^{0.5} \int_{-0.5}^{0.5} G(\theta_x, \theta_y) \overline{S(\theta_x, \theta_y)} d\theta_x d\theta_y = \\ &= \int_{-0.5}^{0.5} \int_{-0.5}^{0.5} \left(\sum_{n_1=-\infty}^{\infty} \sum_{m_1=-\infty}^{\infty} g[n_1, m_1] \exp\{-j2\pi(n_1\theta_x + m_1\theta_y)\} \right) \\ &\quad \cdot \left(\sum_{n_2=-\infty}^{\infty} \sum_{m_2=-\infty}^{\infty} \overline{s[n_2, m_2]} \exp\{j2\pi(n_2\theta_x + m_2\theta_y)\} \right) d\theta_x d\theta_y \\ &= \sum_{n_1=-\infty}^{\infty} \sum_{m_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} \sum_{m_2=-\infty}^{\infty} g[n_1, m_1] \overline{s[n_2, m_2]} \cdot \\ &\quad \cdot \int_{-0.5}^{0.5} \int_{-0.5}^{0.5} \exp\{j2\pi([n_1 - n_2]\theta_x + [m_1 - m_2]\theta_y)\} \end{aligned}$$

האינטגרל הפנימי אינו אלא פונקצית הלם דיסקרטית

$$\begin{aligned} \langle G(\theta_x, \theta_y), S(\theta_x, \theta_y) \rangle_{\theta_x, \theta_y} &= \\ &= \sum_{n_1=-\infty}^{\infty} \sum_{m_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} \sum_{m_2=-\infty}^{\infty} g[n_1, m_1] \overline{s[n_2, m_2]} \cdot \delta_{2D}[n_1 - n_2, m_1 - m_2] = \\ &= \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} g[n, m] \overline{s[n, m]} \end{aligned}$$

כמקרה פרטי תיתן את תוצאת פרסבל:

$$\int_{-0.5}^{0.5} \int_{-0.5}^{0.5} |G(\theta_x, \theta_y)|^2 d\theta_x d\theta_y = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} |g[n, m]|^2$$

- תכונת הקונבולוציה – עבור שני אותות $g[n,m]$ ו- $s[n,m]$ עם התמרתם, הגדרנו את הקונבולוציה בין שתי הפונקציות בזמן ע"י

$$g[n, m] \otimes s[n, m] = \sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} g[n - k, m - j] s[k, j]$$

נראה כי בתדר חישוב הקונבולוציה הופך למכפלה רגילה:

$$\begin{aligned}
F\{g[n,m] \otimes s[n,m]\} &= F\left\{\sum_{k=-\infty}^{\infty} \sum_{p=-\infty}^{\infty} g[n-k, m-p] s[k,p]\right\} = \\
&= \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \left(\sum_{k=-\infty}^{\infty} \sum_{p=-\infty}^{\infty} g[n-k, m-p] s[k,p] \right) \cdot \exp\{-j2\pi(n\theta_x + m\theta_y)\} = \\
&= \sum_{k=-\infty}^{\infty} \sum_{p=-\infty}^{\infty} s[k,p] \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} g[n-k, m-p] \cdot \exp\{-j2\pi(n\theta_x + m\theta_y)\} = \\
&= \left(\sum_{k=-\infty}^{\infty} \sum_{p=-\infty}^{\infty} s[k,p] \exp\{-j2\pi(k\theta_x + p\theta_y)\} \right) \cdot \\
&\quad \cdot \left(\sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} g[n,n] \cdot \exp\{-j2\pi(n\theta_x + m\theta_y)\} \right) = G(\theta_x, \theta_y) S(\theta_x, \theta_y)
\end{aligned}$$

משמעות תוצאה זו היא שבמקום לחשב קונבולוציה נוכל לעבור להתמרות הפורייה של שני האותות המעורבים, להכפילם זה בזה סקלרית, ואז לבצע התמרה הפוכה חזרה לזמן.

אותות דגומים בעלי תמך סופי

לסיכום דיון זה, נטפל באותות בעלי תמך סופי – נניח כי בידנו אות $s[n,m]$ שקיים רק עבור $n, m \in [0, N-1]$. הרי מרבית האותות בהם נעסוק יהיו כאלה, מכיוון שלא ניתן לטפל בקבצי תמונות בגודל אינסופי. נטפל באות זה כאילו היה בעל תמך אינסופי, כשנניח התאפסות מחוץ לתמך $n \in [0, N-1]$. טכניקה זו קרויה המשכת האות באפסים. בדרך זו נוכל להציע קונבולוציה כמו קודם אך גבולות הקונבולוציה יוגבלו ל:

$$g[n,m] = T\{s[n,m]\} = \sum_{k=0}^{N-1} \sum_{j=0}^{N-1} h[n-k, m-j] s[k,j] = h[n,m] \otimes s[n,m]$$

אנו מניחים ש- s בעל תמך סופי. אם גם h בעל תמך דומה, נקבל כי המוצא יהיה שונה מאפס עבור $0 \leq n, m \leq 2N-1$. מכיוון שכך, נוכל לעשות שימוש בתכונת הקונבולוציה של התמרת הפורייה, להמיר את האות כאילו היה קיים בכל זמן, להמיר את תגובת ההלם, להכפיל, ולבצע התמרה הפוכה.

כמו בדיון בחד-מימד, נוכל להציע לדגום את צירי התדר ב- N נקודות שוות מרחק - $\theta_x = p/N$, $0 \leq p \leq N-1$ ו- $\theta_y = q/N$, $0 \leq q \leq N-1$. ונקבל את התמרת הפורייה הדיסקרטית:

$$S[p, q] = \text{DFT}\{s[n, m]\} = \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} s[n, m] \exp\left\{-\frac{j2\pi(np + mq)}{N}\right\} \quad p, q = 0, 1, 2, \dots, N-1$$

פעולת ה-Discrete Fourier Transform (DFT) מקבלת N^2 מספרים המתארים תמונה בתמך סופי, ומחזירה N^2 מספרים המתארים את התנהגות אות זה בתדר, כשהרעיון הוא לדגום את הפונקציה $S(\theta_x, \theta_y)$.

האם ה-DFT הפיכה? מסתבר שכן! מסתבר כי ההתמרה ההפוכה נתונה

ע"י

$$s[n, m] = \text{IDFT}\{S[p, q]\} = \frac{1}{N^2} \cdot \sum_{p=0}^{N-1} \sum_{q=0}^{N-1} S[p, q] \exp\left\{\frac{j2\pi(np + mq)}{N}\right\}$$

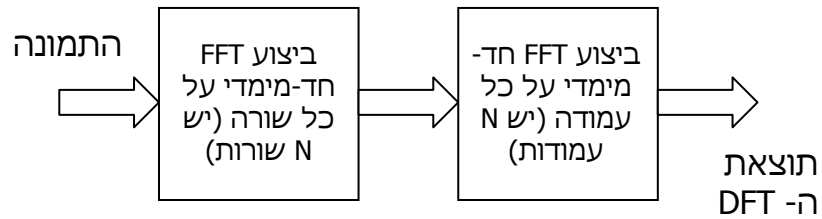
for $n, m = 0, 1, 2, \dots, N-1$

וקל לראות נכונות התמרה הפוכה זו לפי ההגדרה.

מסתבר כי ניתן לחזור ולהציג את כל אותן תכונות שהצגנו קודם (ליניאריות, הזזה ואפנון, ממוצע, ממשיות, פרסבל, קונבולוציה) גם להתמרה זו. גם כאן עולה עניין הריפוד באפסים על מנת להימנע מקונבולוציה ציקלית, וזו מתרחשת בשני הצירים.

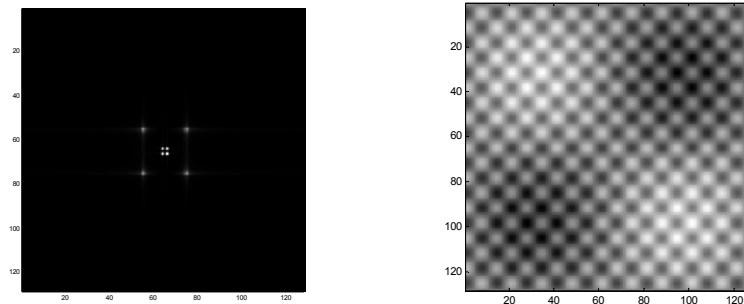
פעולת ה-2D-DFT יכולה להיות מיוצגת בצורה מטריצית, כמטריצה ריבועית בגודל N^2 -על- N^2 המקבלת וקטור באורך N^2 ומכפילה אותו לקבלת וקטור מוצא באותו אורך. מטריצה זו בעלת תכונות מיוחדות – היא יוניטרית עד כדי קבוע (כלומר, היפוכה מתקבל ע"י שחלוף צמוד), ותכונה זו קשורה הדוק לתכונת הפרסבל, שהרי ידוע שמטריצות יוניטריות לא משנות אורך של וקטורים.

בשל היות התמרה זו פרידה, ניתן ליישמה כסדרת התמרות חד-ממדיות ואז להשתמש ב-FFT. לכן מורכבות החישוב כאן תהיה $N^2 \log N$. לכן, ביצוע 2D-DFT ייעשה כמתואר בציור הבא:



ביצוע 2D-DFT על התמונה בדרך פרידה

נביא דוגמה להפעלת DFT בדו-מימד. עבור אות סינוסואידלי הכולל 2 הרמוניות ואשר הינו בגודל של 128 על 128 פיקסלים, תוצאת ה-DFT נתונה בערכה המוחלט.



תמונה דו-ממדית והתמרתה ב-DFT דו-מימדי.

ציור זה נעשה ע"י פקודות ה-Matlab

```

[x,y]=meshgrid(-3:6/127:3);
z=sin(x).*sin(y)+sin(x*10).*cos(y*10);
figure(1); imagesc(z); axis image; colormap(gray(256));
Zf=fft2(z); Zf=fftshift(Zf); figure(2);
imagesc(abs(Zf)); axis image; colormap(gray(256));
  
```

קוונטיזציה

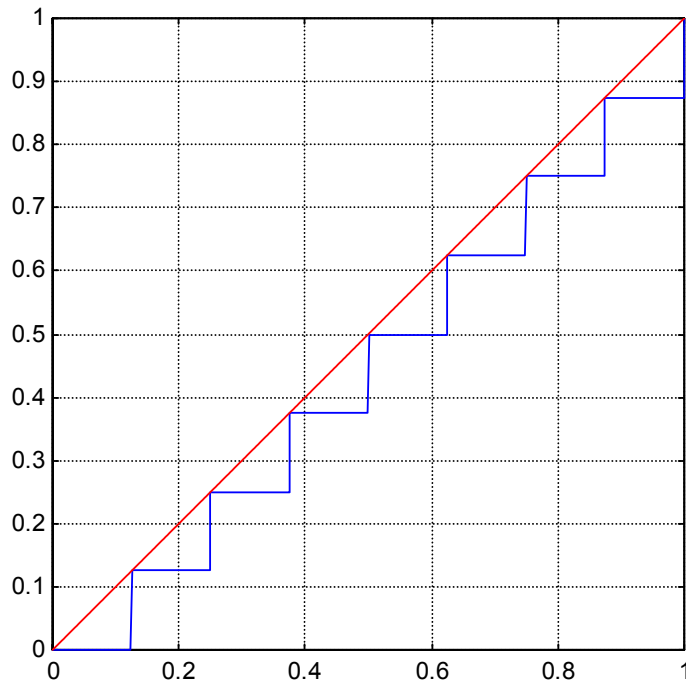
נניח כי נתונה לנו תמונה דגומה אשר תסומן כ- $f[m,n]$. ערכי הדגימות הינם מעל הממשיים וטווח הערכים מצוי באינטרוול $[0,1]$. כזכור, הערך 0 מתייחס לגוון שחור מוחלט, והערך 1 לגוון לבן מוחלט. ברור כי לשם ייצוג נאות של ערכי רמות האפור נדרש לאינסוף סיביות. מוצע אם כך להשתמש ב- L סיביות (L הוא ערך קבוע מראש וערכיו הסבירים יהיו בתחום 1-16). השאלה המתבקשת היא כמובן כיצד להמיר את הערכים הרציפים לערכים הדיסקרטיים. כיוון ש- L סיביות מייצגות 2^L ערכים שונים, אנו נדרשים לבחור 2^L רמות ייצוג שונות.

אחת הפעולות הפשוטות ביותר אותן ניתן להציע היא פעולת קיצוץ לסיביות הנמוכות. בגישה זו, עלינו לייצג את רמת האפור בקבוצת L הסיביות הגבוהות ביותר (Most Significant). מתמטית, פעולה זו נעשית ע"י המשוואה

$$Q(f) = \frac{\lfloor 2^L \cdot f \rfloor}{2^L}$$

לדוגמה, עבור $L=2$, נקבל כי הערכים בתחום $[0,0.25]$ ימופו לאפס, הערכים באינטרוול $(0.25,0.5]$ ימופו לגובה 0.25 וכו'. באופן דומה ניתן להראות כי במוצא הקוונטיזר ייתכנו רק 4 ערכים אפשריים – 0, 0.25, 0.5 ו- 0.75. לערכי הכניסה בהם חל שינוי במוצא נקרא רמות ההחלטה, ולערכי המוצא האפשריים נקרא רמות הייצוג. הציור הבא ממחיש את הקשר בין הכניסה למוצא של הקוונטיזר המוצע עבור $L=3$. רמות הייצוג הם מיקומי הקווים האופקיים בפונקצית המדרגות על הציר האנכי, ורמות ההחלטה הם מיקומי הקווים האנכיים על הציר האופקי. כפי שאנו רואים, הקשר מאופיין ע"י פונקצית מדרגות.

קוונטיזר מושלם היה קוונטיזר בעל מוצא זהה לכניסה, כיוון שאז אין שגיאה. ברור כי שימוש במספר סופי של סיביות לא מאפשר קוונטיזר כזה. כפי שניתן לראות, בגישה המוצעת תמיד נקבל כי מוצא הקוונטיזר נמוך מכניסתו. השגיאה בין הכניסה ליציאה תהיה אפס בנקודות המפגש בין הקו המדרג לקו האלכסוני (המייצג את מערכת היחידה), ומקסימלי בנקודות הברך שביניהן.



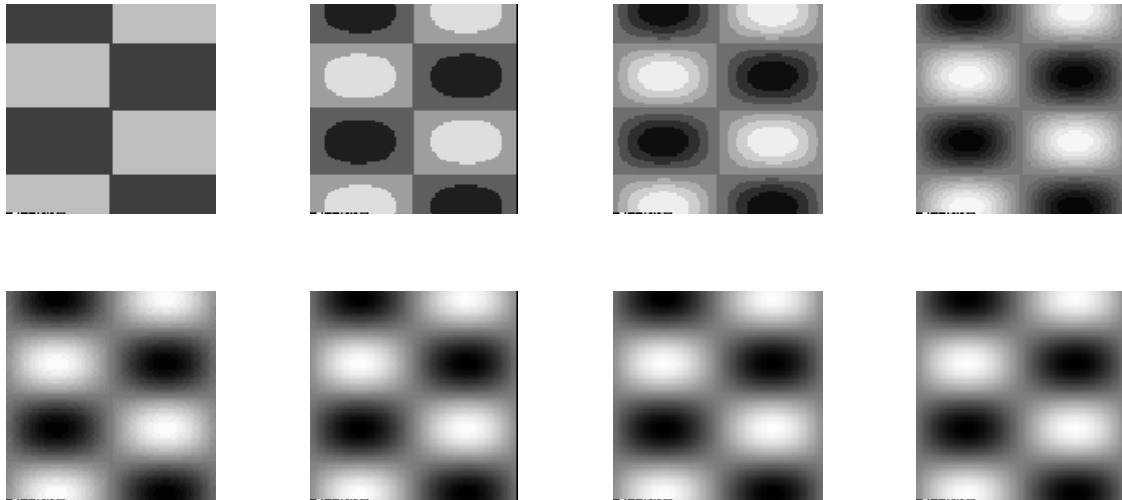
קוונטיזציה ע"י קיצוץ הסיביות התחתונות (Least Significant)
 L הוא 3 (8 רמות ייצוג במוצא).

ניתן לשפר את הפעולה הנ"ל ולהקטין את השגיאה ע"י הכנסת תוספת למוצא:

$$Q(f) = \frac{\lfloor 2^L \cdot f \rfloor}{2^L} + 2^{-(L+1)} = \frac{\lfloor 2^L \cdot f \rfloor + 0.5}{2^L}$$

התוספת תגרום לגרף המדרגות לנוע מעלה ולהתמקם סביב מערכת היחידה (הקו האלכסוני), ואמנם, השגיאה בין הכניסה ליציאה תהיה קטנה יותר.

הגישה שתוארה פשוטה למדי ולכן פופולארית. בתמונות אופייניות, ביצוע קוונטיזציה כזו ליותר מ- 5 סיביות (32 רמות ייצוג) מניבה על פי רוב תוצאה בה הקוונטיזציה לא מורגשת לעיין אנושית. כשירידים מ- 5 סיביות ניתן לראות בפועל את פעולת הקוונטיזציה, ונוצרות תופעות צורמות. אנו נדון בדרכים לתיקון תופעות אלו בהמשך. הציור הבא מראה תמונה פשוטה (תמונת הסינוס הדו-מימדי), עם קוונטיזציה למספר סיביות משתנה, וממחישה תופעה זו. הקוונטיזציה שבוצעה משתמשת בנוסחה עם העדכנית (הכוללת את התוספת להזזת הגרף כלפי מעלה).



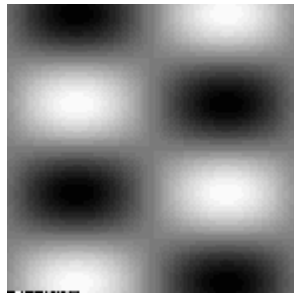
קוונטיזציה פשוטה של תמונה למספר סיביות משתנה

נציע כעת ניסוי מעניין. ניקח את אותה תמונה הממלאת ערכים בתחום $[0,1]$ ונכווץ את התחום הדינאמי שלה ע"י הקשר

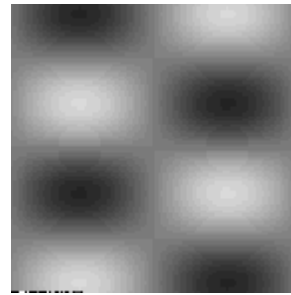
$$\hat{f}[m,n] = \alpha[f[m,n] - 0.5] + 0.5$$

כאשר α בתחום $[0,1]$. עבור $\alpha = 0.5$ יתקבל כי התמונה כעת מקבלת ערכי רמות אפור בתחום $[0.25, 0.75]$, כלומר התחום הדינאמי של התמונה קטן. ביצוע קוונטיזציה עם כמות סיביות נאותה (נניח 5 סיביות) בגישה שתוארה לעיל תיצור בעיה. קוונטיזציה על התמונה המקורית תיתן איכות נאותה, אך עם כיווץ התחום הדינאמי נקבל כי הקוונטיזציה מניבה תוצאה צורמת. הצירוף הבא מראה תוצאה זו עבור שימוש ב- 5 סיביות ושינוי α .

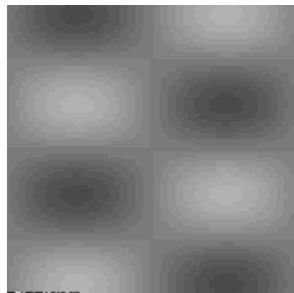
ניכר שוב כי עבור ערכי α קטנים, 5 סיביות אינן מספיקות. הסיבה לכך ברורה – עבור תחום דינאמי קטן, רק חלק מ- 32 רמות הייצוג האפשריות באות לידי ביטוי, וכל האחרות מתבזבזות! רעיון נאות יהיה, לפיכך, ביצוע קוונטיזציה שתלויה בתוכן התמונה, כך שבכל התמונות הנ"ל ינוצלו כל רמות הייצוג באופן מיטבי. ניתן להשיג זאת ע"י שימוש בקוונטיזר באופן אדפטיבי במקום, אך דרך היישום מחייבת סיביות. הרעיון יהיה הבא: על סמך פיקסלים שכנים שכבר עברו קוונטיזציה והראו על התכווצות התחום הדינאמי, נפעיל על שכניהם הסיבתיים קוונטיזר בעל פרמטרים אחרים – זה אפשרי אך מסוכן, כי יתכן שהפיקסל הבא לא יתאים לניבוי של שכניו ויקבל קוונטיזציה נוראית בשל כך.



$\alpha = 1, L=5$



$\alpha = 0.7, L=5$



$\alpha = 0.4, L=5$



$\alpha = 0.1, L=5$

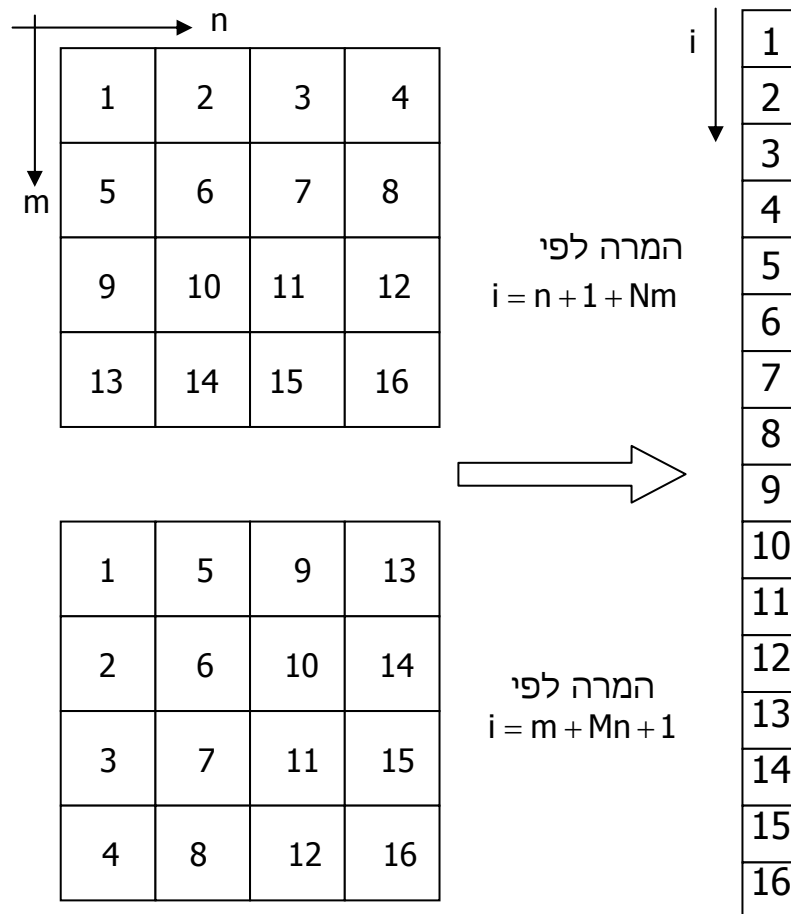
המחשת ההשפעה של תחום דינאמי על איכות הקוונטיזציה

הקוונטיזציה שתוארה לעיל (על שתי גרסאותיה) קרויה אחידה (יוניפורמית) כיוון שהמרווחים בין רמות הייצוג השונות או בין רמות ההחלטה השונות קבועים. נשאלת השאלה: מדוע לא לאמץ תפיסה בה הפונקציה המדורגת כללית ומאופיינת ע"י רמות החלטה וייצוג כלשהם? ואם כן, כיצד נחליט מה תהיינה הרמות הנאותות? במקרה הכללי, פעולת הקוונטיזציה תהיה פונקצית מדרגות הממפה את האינטרוול $[0,1]$ לסידרה של $N = 2^L$ גבהים שונים באופן שיתאים לאופי רמות האפור – נקודה זו מכוסה בדיון על קוונטיזר מקס-לויד.

המרת אותות מדו מימד לחד-מימד

כנקודה אחרונה במסגרת פרק זה נזכיר כי ניתן להתייחס לתמונה כאות חד-מימדי ע"י סידור כלשהו. סידור פופולרי מציע ארגון העמודות בזו מתחת זו (או השורות). סידור כזה קרוי Column-Lexicographic Ordering. במקרים רבים אלגוריתמים לעיבוד תמונה נתפרים לתמונות שהומרו לוקטורים חד-ממדיים.

ברור כי עם שינוי כזה נאבד קשר בין שכנים שהיו קרובים בדו-מימד והפכו רחוקים. אנו נחזור לנקודה זו כשנדבר על דחיסה.

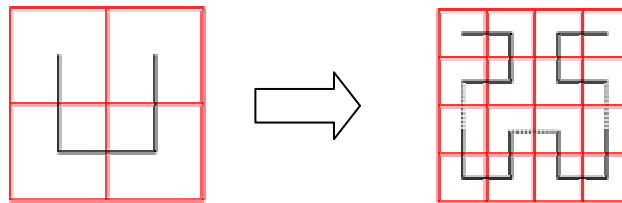


סידור לקסיקוגרפי מבוסס שורות (למעלה)
ועמודות (למטה) לתמך $\Omega = \{[m, n] \mid 0 \leq m \leq M-1, 0 \leq n \leq N-1\}$

סידור פופולרי אחר הוא סידור לפי עקומת Peano-Hilbert. עקומה זו היא אחת מיני גישות רבות קיימות לציור עקומים הממלאים את המישור ללא חציה עצמית (space filling curves) וזוהי הזירה במתמטיקה בה עוסקים בפרקטלים ובכאוס. הרעיון הבסיסי הוא ציור של תבנית בסיסית, לקבוע כלל החלפה שלה בתבניות כמוה קטנות יותר, ולחזור על התהליך עוד ועוד באופן איטרטיבי.

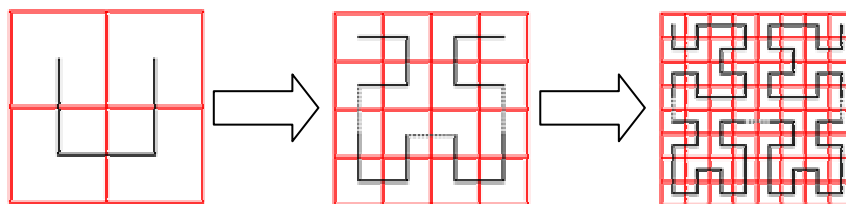
נמחיש תהליך זה לעקומת Peano שאותה יצר Peano (מדען איטלקי שחי לפני כמאה שנה) ב-1890. תבנית היסוד שלנו היא עקומה בצורת האות 'ח'.

כלל ההחלפה יהיה יצירת תבנית זהה בגודל חצי מהמקור, ובניית רביעיית תבניות כאלה – 2 למטה כמו 'ח', ולמעלה כשהן מסובבות ימינה ומשאלה כמתואר בצור.



תבנית היסוד ושיטת ההחלפה בבניית עקומת Peano

הקו המקווקו מראה כי את ארבעת התבניות שנוצרו יש לחבר, אך לקווים אלו נתייחס כקווי קישור ולא חלק מהבניה עצמה. כעת, כל תבנית מהארבע הקיימות תעבור את אותו תהליך שוב, בו היא מוחלפת בעקומה מפורטת יותר, כמתואר.



עקומת Peano-Hilbert להמרת תמונה לוקטור

כשנתייחס לקוביות כמרכזיהן של פיקסלים, הרי שקיבלנו עקומה רציפה העוברת דרך כל הפיקסלים, ומבקרת בכל אחד פעם אחת בלבד.

ההמרה לחד מימד נוחה בעיקר כי אז ניתן להשתמש בכל אותם כלים שנתפרו לאותות חד-ממדיים, להכניס משמעות של סיבתיות באופן טבעי ועוד. קיים קו כזה של עבודות בעיבוד תמונה, ויישומים אלו אופייניים למערכות מימוש בחומרה שם שיקולי זיכרון ומהירות חיוניים. עם זאת, ברור כי יש הפסד עצום במעבר לחד-מימד וביצועי מערכות אלו צפויים להיות ירודים.

עיבוד תמונות בעזרת מחשב (236327)

הרצאה מס' 7

שיפור תמונות - בסיס

הצגת הבעיה

בפרק זה נתחיל את הדיון בפעולות עיבוד על תמונות. העיבודים בהם נעסוק בפרק זה יתמקדו בניסיון לשפר את מראה של התמונה הנתונה. חשוב להבהיר כבר כעת כי בעיבוד תמונות קיימים עיבודים בעלי מטרה שונה, כגון דחיסה, ניתוח תוכן התמונה ועוד, אך אלה יידחקו לפרקים מאוחרים יותר.

מדוע לשפר תמונה? ובכן, במקרים רבים מתקבלת תמונה שאינה באיכות הרצויה, אם בשל דרך השגתה, או בשל קלקולים שעברה בתהליכי עיבוד קודמים. בעיות אופייניות הן תמונה מרוחה הדורשת חידוד, תמונה בה נדרש למתוח את הקונטרסט, תמונה בה העצמים החשובים אינם נראים טוב ונדרש להבליטם, תמונה בה יש רעש נלווה ונחוץ לנקותו, ועוד. כנגד כל משימה כזו ניתן להציע אלגוריתמים שונים ומגוונים. לכן, תחת הכותרת "שיפור תמונה" נכנס מגוון רחב מאוד של אלגוריתמים אשר אמנם משפרים את התמונה אך במובנים שונים, ולעיתים קרובות אין כל קשר ביניהם.

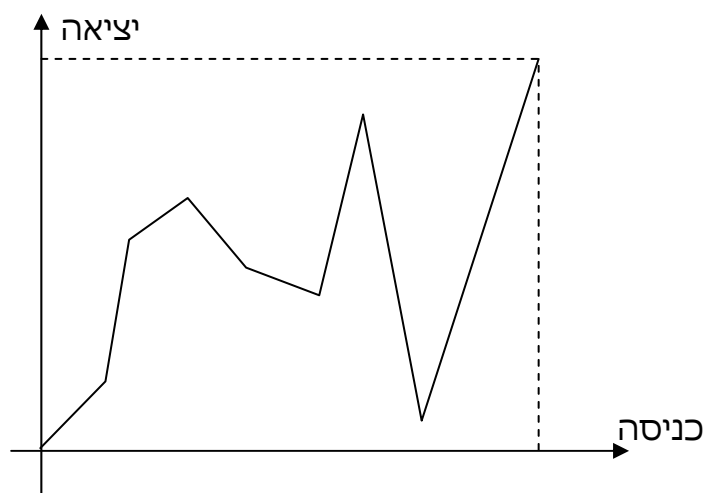
מבחינתנו היה נפלא לו יכולנו להציע אלגוריתם כללי המקבל תמונה ומוציא גרסה משופרת שלה, ללא מעורבותו של המשתמש. המכשלה העיקרית לקיומו של תהליך כזה היא חוסר בתיאור מתמטי של מהי תמונה טובה. כיוון שאין מדד כמותי כזה, שיפור תמונות הוא עניין של מומחיות וטעם אישי - צריך לזהות את הבעיה בתמונה ואז לתפור אלגוריתם ספציפי של פתרון. לכן, התחום הקרוי שיפור תמונות הוא תחום עם הרבה היורסטיקות, ובו פעולות תתבססנה על ניסוי וטעייה. במסגרת פרק זה ננסה להציג את היסודות וכלי העבודה הבסיסיים לשיפור תמונות.

כנקודה אחרונה במסגרת זו, חשוב להבהיר את ההבדל בין שיפור תמונה ושחזור תמונה. קיים תחום נפרד הקרוי "שחזור תמונה", ובו לוקחים תמונה

שעברה קלקול ידוע ומשחזרים את המקור. אנו נדון באלגוריתמים כאלו בהמשך. במסגרת פרק זה איננו מניחים כי הקלקול ידוע, ולכן דרך הפתרון משתנה באופן מהותי.

פעולות נקודה והיסטוגרמת תמונה

אחת הפעולות הפשוטות ביותר שניתן לבצע על תמונה היא פעולה חסרת זיכרון. כפי שכבר ראינו, פעולה זו בצורתה הכללית ביותר תהיה $g[m,n] = T\{f[m,n]\}$, כאשר המוצא בנקודה הדיסקרטית $[m,n]$ תלוי רק בערך הכניסה באותה נקודה. מערכת כזו ניתן לתאר ע"י פונקצית מעבר פשוטה הקובעת כיצד כל ערך כניסה עובר לערך יציאה, כמתואר בציור הבא. אם נניח למשל כי תמונות הכניסה והמוצא מיוצגות ע"י 8 סיביות לפיקסל, הרי שלפנינו מערכת המעבירה כל מספר בתחום $[0,255]$ לערך מוצא בתחום זה, ולכן המערכת ניתנת לתיאור ע"י טבלת התאמה פשוטה הקרויה Look Up Table (LUT).



הקשר כניסה ליציאה של מערכת חסרת זיכרון

השתכנענו כי אמנם מדובר במערכת פשוטה, והשאלה המתבקשת היא מה ניתן להשיג באמצעות מערכת כזו? מסתבר כי לעיתים תכופות יכולה מערכת LUT לספק שיפור ממשי לתמונת הכניסה. לפני שנמחיש את הפעולות

הללו נציג את מושג ההיסטוגרמה לתמונה, אשר יסייע בהבנתן של פעולות אלו.

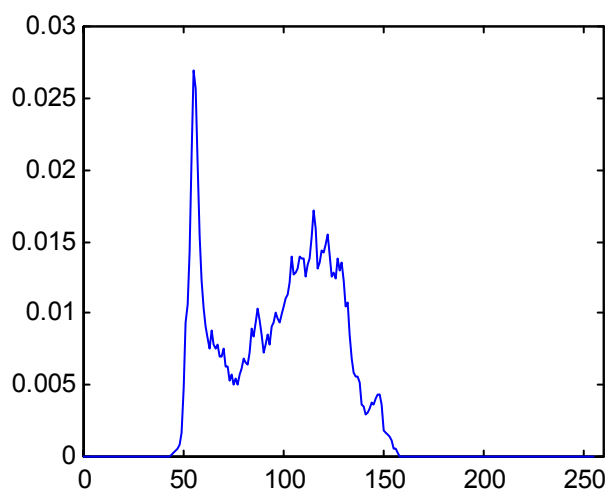
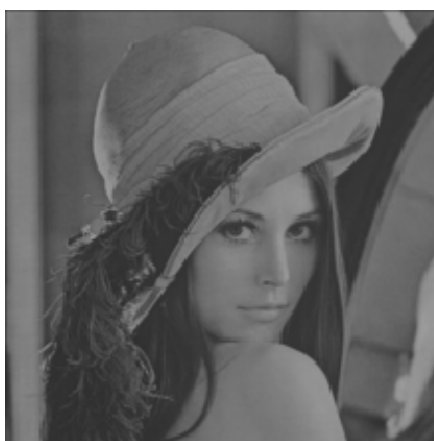
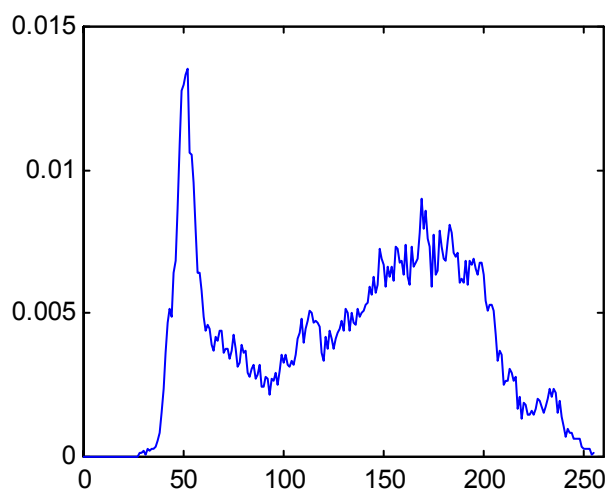
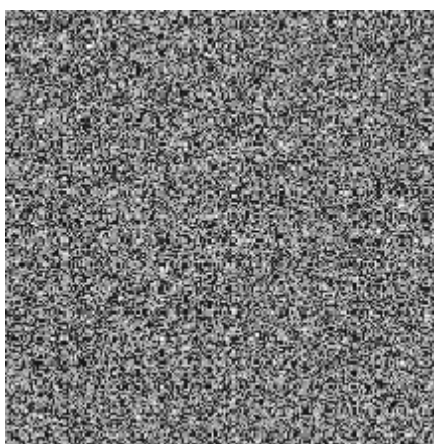
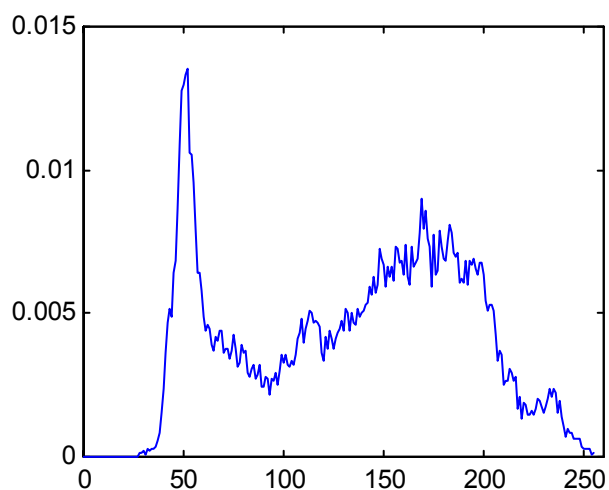
היסטוגרמה לתמונה היא פונקציה השכיחות של רמות האפור בתמונה, בדומה לפונקציות שתוארו במסגרת פרק קודם בדיון על קוונטיזר מקס-לוייד. בהינתן תמונה בה כל פיקסל מיוצג ע"י n סיביות, כל פיקסל יכול לקבל ערך רמת אפור שלם בתחום $[0, 2^n - 1]$. הערך $h[k]$ הוא כמות הפיקסלים בתמונה שערכם הוא k . לכן, הפונקציה $h[k]$ היא פונקציה חד-מימדית המקבלת ערך חיובי כנגד כל k בתחום $[0, 2^n - 1]$. בדרך כלל מקובל לחלק את ערכי הפונקציה בכמות הפיקסלים הכוללת בתמונה, לקבלת פונקציה שסכום איבריה הוא 1. במקרה כזה, הפונקציה $h[k]$ נראית כפונקציה הסתברות לכל דבר המקיימת:

$$\begin{aligned} 1. & \forall k \in \mathbf{Z}, 0 \leq k < 2^n, h[k] \geq 0 \\ 2. & \sum_{k=0}^{2^n-1} h[k] = 1 \end{aligned}$$

היסטוגרמה של תמונה מתארת תכונה מסוימת של התמונה – מידת אכלוס רמות האפור בתמונה. להיסטוגרמה אין מרכיב המתאר התנהגות מקומית, כלומר, מהיסטוגרמה לא ניתן לומר מהי רמת האפור במקום מסוים בתמונה, וברור כי מהיסטוגרמה לא ניתן לבנות מחדש את התמונה. בציור הבא מוצגות מספר תמונות וההיסטוגרמה שלהן. בכל המקרים מדובר בתמונות המיוצגות ב-8 סיביות לפיקסל. כפי שמודגם בציור זה (בשורה האמצעית), שינוי סדר הפיקסלים בתמונה אינו משפיע כלל על ההיסטוגרמה. ע"י שינוי סדר הפיקסלים בתמונה מתקבלת תמונה שונה לחלוטין. למרות זאת, ברור כי ההיסטוגרמה נותרת ללא שינוי כלל.

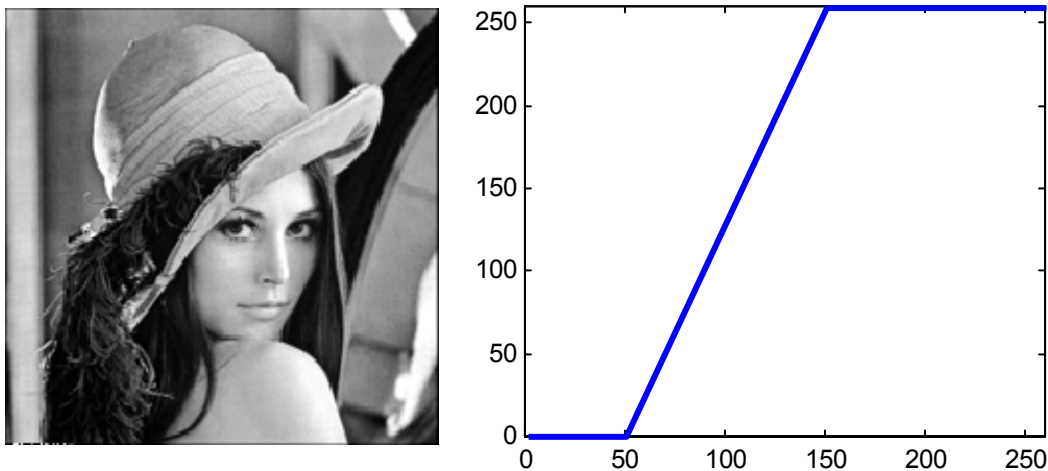
מדוע אנו מתעניינים בהיסטוגרמת תמונה? מסתבר כי עיון בהיסטוגרמת התמונה חושף מיד בעיות של ניצול תחום דינאמי. כך למשל, התמונה החשוכה מעט שמתוארת בתחתית הציור הבא נראית כך כיוון שערכי רמות האפור בה מאכלסים בקירוב את התחום $[50, 150]$. זוהי תמונה שמבזבזת תחום דינאמי. ע"י פעולת LUT על תמונה זו מהצורה

$$\text{Out} = 2.55 \cdot [\text{In} - 50]$$



מספר דוגמאות לתמונה וההיסטוגרמה שלה. בשורה העליונה נתונה התמונה Lena וההיסטוגרמה המתאימה לה. בשורה האמצעית נלקחה התמונה Lena ועורבלה (שינוי סדר אקראי של הפיקסלים בה), וכצפוי, ההיסטוגרמה שלה זהה. בשורה התחתונה התמונה המקורית הוכפלה ב-0.5 ונוסף לה הערך 30, ולכן, ההיסטוגרמה דומה במראה הכללי, אך צרה יותר (ולכן גם גבוהה יותר) ומוסטת.

מתקבלת תמונה אשר מאכלסת את כל רמות האפור מאפס ועד 255. תוצאה זו ופעולת ה- LUT התואמת לה מתוארים בציור הבא. פעולה זו תיקרא מתיחת קונטרסט (Contrast Stretching). נשים לב כי פעולה זו ממפה לערכים שליליים את רמות האפור הקטנות מ- 50, ולערכים גבוהים מ- 255 את ערכי הכניסה מעל 150. אבל בתמונה הנתונה אף פיקסל כמעט אינו בתחומי ערכים אלו, ולכן אין בעיה. על מנת להבטיח תוצאה נאותה, אנו מרתקים כל ערך שלילי לאפס, וכל ערך מעל 255 ל- 255.



שיפור התמונה בציור הקודם (שורה תחתונה) ע"י פעולת מתיחת קונטרסט המתוארת כ- LUT בצד ימין

במקרה הכללי, בהינתן תמונה בה רמת האפור המינימלית תסומן $f_{\min}$ ורמת האפור המקסימלית $f_{\max}$, פעולת המתיחה לשם ניצול כל הרמות בתחום $[0, 255]$ תהיה:

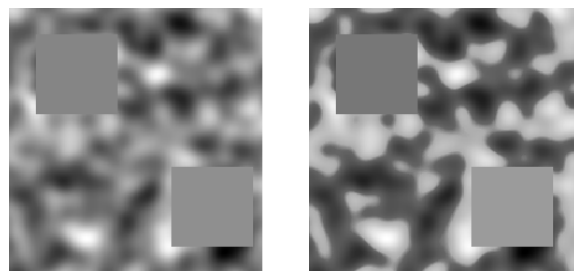
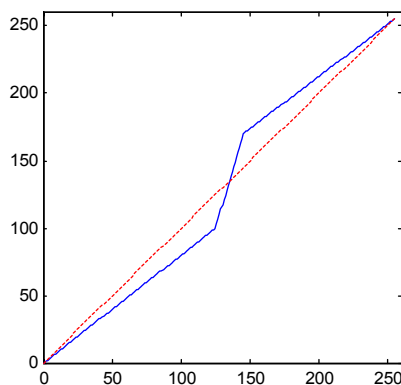
$$\text{Out} = 255 \cdot \frac{\text{In} - f_{\min}}{f_{\max} - f_{\min}}$$

לעיתים שימוש בערכים הקיצוניים ($f_{\max}$ ו- $f_{\min}$) אינה מוצלחת. לדוגמה, אם בתמונה כל רמות האפור למעט אחת מצויות בתחום $[0, 100]$, וישנו פיקסל בודד בעל ערך 255, הרי שלפי המתואר לעיל, מתיחת קונטרסט לא תשנה את התמונה, בשל אותו פיקסל חריג. כדי להתגבר על בעיה זו, ניתן להגדיר את $f_{\min}$ ו- $f_{\max}$ באופן שונה מעט. נזכור כי פונקצית ההיסטוגרמה היא פונקצית צפיפות פילוג, ולכן נוכל להגדיר פונקצית הצטברות הפילוג $H(x)$ ע"י:

$$H[k] = \sum_{j=0}^k h[j] \Rightarrow H[k] = 0; H[2^n - 1] = 1$$

(n – כמות הסיביות לפיקסל). אזי, אם קודם הגדרנו את $f_{\min}$ להיות הנקודה הגבוהה ביותר בה $H[k] = 0$, נגדיר אותה הפעם להיות הנקודה בה $H[k] = \varepsilon$. משמעות בחירה זו היא של- ε מהפיקסלים הכי נמוכים נתייחס כחריגים ונבחר את הגדול שביניהם בערך מינימלי. באופן דומה, בחירת $f_{\max}$ כנקודה הנמוכה ביותר בה $H[k] = 1$ תוחלף בנקודה בה $H[k] = 1 - \varepsilon$. מתיחה על-פי שני ערכים אלו תגרום לכמות יחסית של 2ε מהפיקסלים בתמונה להיכנס לרוויה (מעל 255) ולקיטעון (מתחת 0), אך זו קבוצה מצומצמת של נקודות, ולעומתה, עיקר התמונה יעבור שיפור מוצלח.

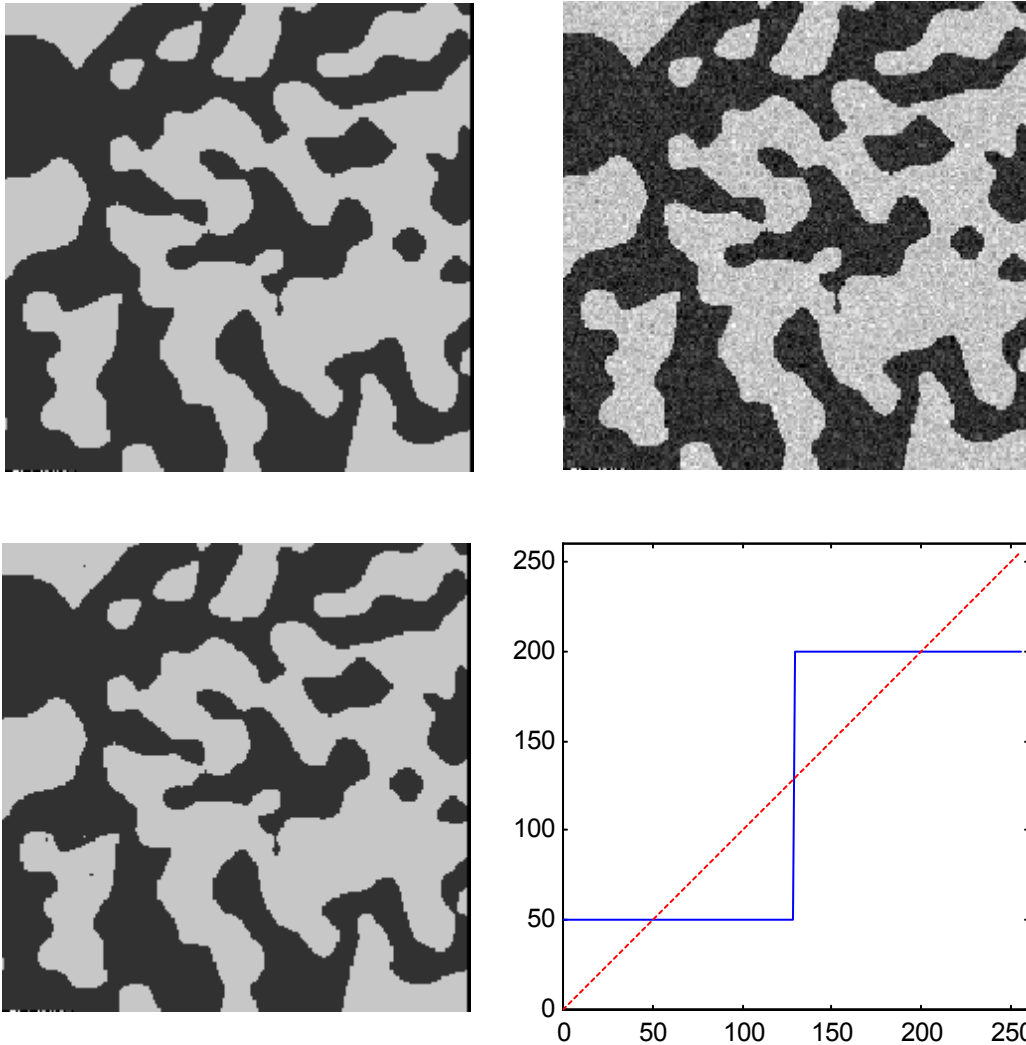
כדוגמה אחרת לפעולת מתיחת קונטרסט, בציור הבא מתוארת תמונה בה ישנם שני אובייקטים בעלי רמות אפור קרובות. נניח כי הם העיקר בתמונה ורצוננו להגדיל את יכולת האבחנה ביניהם. בציור זה מוצע גם LUT להגשמת יעד זה. רמת האפור של המטרות הן 130 ו-140. פעולת ה-LUT מרחיקה שתי רמות אלה לערכים 100 ו-170 בהתאמה, תוך פגיעה ביתר התמונה (אשר כאמור, נחשב לפחות חשוב במשימה שלפנינו). גם זוהי פעולת מתיחת קונטרסט, אך רק לתחום מסוים של רמות האפור. בתוצאה ניכר כי אמנם שני הריבועים הינם בעלי גוון שונה, כנדרש.



מתיחת קונטרסט לאזורים ספציפיים על חשבון אזורים אחרים. שמאל למעלה – תמונת מקור. ימין למעלה – תוצאת המתיחה, בה ניכר כי שני הריבועים בעלי גוון שונה. למטה – ה-LUT שיושם

פעולה אחרת הניתנת לתיאור כ-LUT היא פעולת סף. נניח כי נתונה תמונה אשר יודע כי במקורה היא תמונה בינארית (נניח ערכי רמת אפור 50 או

200 בלבד!). לתמונה זו צורף רעש בתהליך הרכשתה ורצוננו לנקותה. ע"י פעולת סף נוכל להעלים חלק גדול מהרעש. בציור הבא ממחישים תהליך זה ותוצאותיו כפי שניתן לראות אנו מצליחים לקבל תמונה כמעט זהה למקור לפני הרעש.

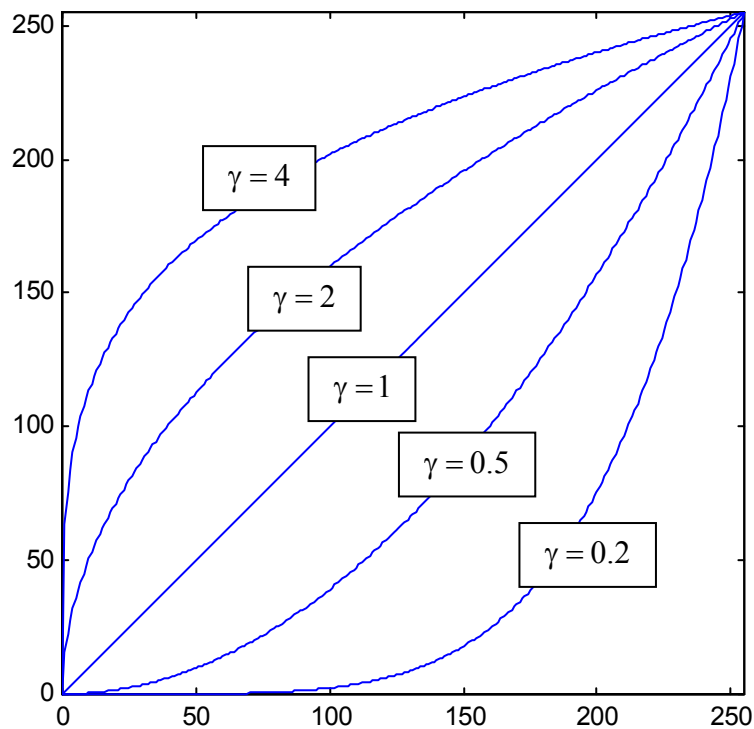


ניקוי תמונה בינארית מרעש נלווה ע"י פעולת סף. למעלה משמאל – תמונת מקור בינארית. למעלה מימין – התמונה לאחר זיהום ברעש. למטה משמאל – תוצאת הסף. למטה מימין – ה- LUT שיושם

הראינו כיצד אנו משנים קונטרסט – כעת נעבור להראות שינוי בהירות. בהירות היא בעצם רמת האפור הממוצעת בתמונה. לכן, בחירה אינטואיטיבית לפעולת שינוי בהירות תהיה הוספת ערך קבוע לרמת האפור. בחירה כזו אינה מוצלחת בדרך כלל כיוון שהיא גורמת להוצאת הערכים מהתחום המותר. כתחליף, מקובל להשתמש בתיקון גאמה (Gamma Correction), אשר נראית כך:

$$\text{Out} = \text{In}^{\frac{1}{\gamma}}$$

עבור ערכי גאמה גדולים מ-1 נקבל הבהרה של התמונה, ועבור ערכי גאמה קטנים מ-1 התמונה תוכה. הציור הבא מראה את צורת ה-LUT למגוון ערכי גאמה.



תיקון בהירות ע"י פעולת Gamma Correction

לפעולת תיקון הגאמה שורשים מוצקים בתהליכי הצילום וההדפסה. מסתבר כי בצילום תמונה (לסרט צילום או לגלאי CCD) רגישות החישה לאור אינה ליניארית וניתנת לתיאור כללי כפעולת חזקה על הכניסה, עם מעריך חיובי וגדול מ-1. לעומת זאת, תהליך ההדפסה (או ההצגה למסך) מאופיינים אף הם בעקומה לא ליניארית של הצגת התמונה, כשעבורם מקובל להניח כי הפעולה היא פעולת חזקה עם מעריך חיובי קטן מ-1. ברור כי עבור ערכים תואמים (יחס של היפוך בין השניים) יתקבל ביטול הדדי ואין צורך בתיקון נוסף. לעומת זאת, כאשר התקני הקלט\פלט אינם מתואמים, נדרשת פעולת תיקון ביניהם, וזוהי בדיוק פעולת הגאמה. הציור הבא מראה סכימת בלוקים של המערכת אשר תוארה.



סכימת בלוקים המתארת התנהגות אופיינית של מערכות קלט ופלט, והקשר לצורך בתיקון גאמה

פעולות עיצוב היסטוגרמה

עד כה עסקנו בפעולות LUT פשוטות למדי, שנובעות מתוך שיקולים בסיסיים הניכרים מההיסטוגרמה. במסגרת סעיף זה נציג דרכים מורכבות יותר לתכנון פעולת ה-LUT. היעד הראשון אותו נרצה להגשים הוא הבא – בהינתן תמונה כלשהי, נשאל מהי פעולת ה-LUT אשר תמיר את ההיסטוגרמה שלה להיות היסטוגרמה אחידה. מדוע אחידה? ובכן, אינטואיטיבית נראה כי תמונה שמנצלת כל רמת אפור באופן אחיד תהיה תמונה מאוזנת ונעימה לצפייה.

משימה שנייה בה נעסוק דומה בעיקרה - בהינתן תמונה כלשהי, נשאל מהי פעולת ה-LUT אשר תמיר את ההיסטוגרמה שלה להיות היסטוגרמה אחרת נתונה? בעיה זו חשובה ביישומים מסוימים כגון מציאת תנועה בין שתי תמונות, פסיפס של תמונות ועוד. הרעיון אם כך מכליל את הבעיה הקודמת, כך שהפעם אנו נדרשים לעצב את ההיסטוגרמה לפי צורה כלשהי.

האם פעולת LUT יכולה להגשים את היעדים הללו? מסתבר כי בכפוף לכמה דרישות סבירות, פעולות זו אפשריות. נתחיל את המענה לשאלות אלו בשאלת יישור היסטוגרמה, ונמחיש את התהליך דרך דוגמה, וממנה נגזור את הגישה הכללית.

דוגמה: נניח כי נתונה לנו תמונה בגודל 256×100 פיקסלים בה כל פיקסל מיוצג ב-8 סיביות, ולתמונה זו היסטוגרמה $h[k]$ ($k = 1, 2, 3, \dots, 255$). רצונו להעבירה דרך LUT לקבלת תמונה בה ההיסטוגרמה היא $\forall 0 \leq k \leq 255, g[k] = 1/256$, כלומר כל רמת אפור תאכלס ב-100 פיקסלים בדיוק. לשם נוחות נתייחס ל- $h[k]$ ו- $g[k]$ לפני נירמולן במספר הפיקסלים, כך

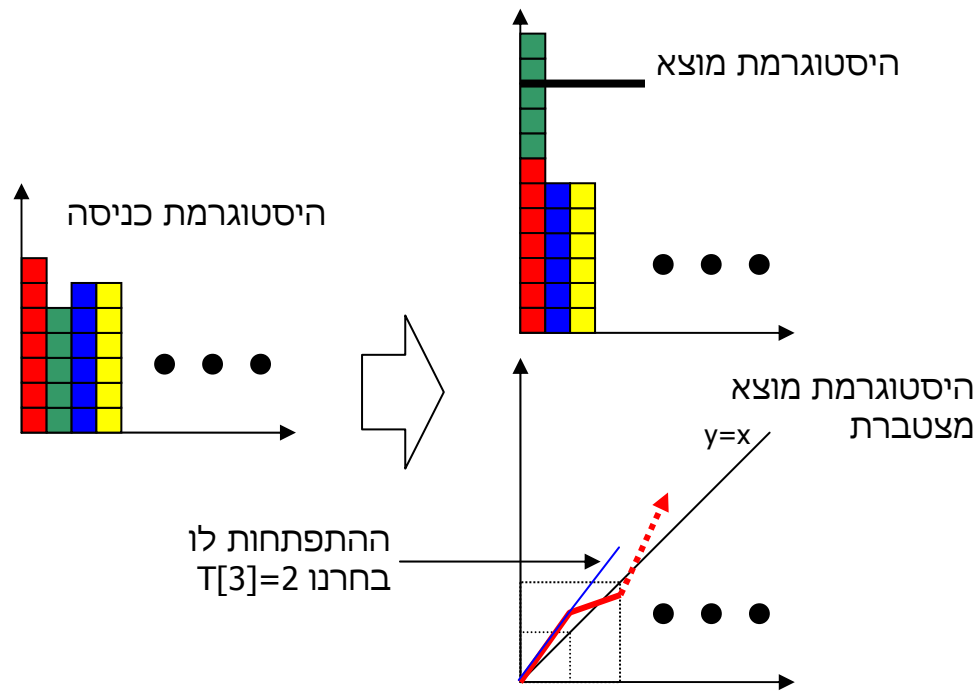
שהן למעשה סופרות כמויות של פיקסלים. במצב זה היעד הוא $g[k]=100$ לכל k . אנו נתייחס לתאים בהיסטוגרמה כדליים בהם אנו ממלאים פיקסלים.

את ה-LUT נתכנן בסדר עולה מ- $k=0$ ועד $k=255$. נניח כי יש 70 פיקסלים שהינם בעלי רמת אפור '0' בתמונת המקור. ברור שיש לקבוע $T[0]=0$ כך שפיקסלים אלו יועברו בתמונת היעד גם לערך '0'. נחוצים עוד 30 פיקסלים להשלמת הדלי של הפיקסלים להם ערך '0' בתמונת המוצא. נעבור לאלו להן ערך '1' במקור. אם יש בדיוק 30 כאלה, נקבע $T[1]=0$ וכך הושלמה בנייתו של הדלי הראשון בהיסטוגרמת היעד. אם לעומת זאת $h[1]=50$ ייחוסם ל-'0' יביא לכמות של 120 פיקסלים (במקום 100) בעלי ערך '0' במוצא. ייחוסם ל-'1' הבא (עם הערך '1' במוצא) יגרום ל-'1' הדלי הראשון להיות עם 70 פיקסלים בלבד. ברור שמה שייעשה לא יוכל להביא להיסטוגרמה מישרת (למעט גישה הסתברותית בה חלק פיקסלים בעלי אותו ערך במקור ממופים לתאים שונים). עם זאת, נוכל לפחות להתקרב לכך, ולבחור באופציה $T[1]=0$ כיוון ששגיאתה קטנה יותר (חריגה של +20 מול חריגה של -30). בדרך זו סיימנו את מילוי של הדלי הראשון $g[0]$ וכעת אנו עוברים למילוי $g[1]$.

נניח כי $h[2]=60$. נקבע $T[2]=1$ וכך מילאנו 60 פיקסלים לדלי $g[1]$. חסרים עוד 40 להשלמתו. נניח כי $h[3]=60$. לכאורה נראה כי רצוי ש- $T[3]=1$ גם הוא על מנת שיהיו 120 פיקסלים בתא זה, מצב הנראה עדיף על חוסר ב-40 פיקסלים. אולם, השגיאה שנעשתה בדלי הקודם צריכה להילקח בחשבון בהחלטה. צירוף 60 הפיקסלים הללו יביא לשני תאים ראשונים בני 120 פיקסלים כל אחד, וביחד חריגה של 40 מהנחוץ (240 לעומת הצורך ב-200). הקביעה $T[3]=2$, לעומת זאת, תביא לשני תאים ראשונים בהם יש 120 ו-60 פיקסלים, חריגה של -20 מהנחוץ (180 מול 200), וזה עדיף. כלומר – המסר כאן הוא שנכון יותר להתייחס להיסטוגרמה המצטברת בבניית T .

ניתן דוגמה קיצונית להמחשת המוטיב הנ"ל של שימוש בהיסטוגרמה מול היסטוגרמה מצטברת. נניח כי בתהליך המילוי נשתמש רק בערכי ההיסטוגרמה (ולא המצטברת), נמלא דליים ונקבל כי כל תא בתהליך עבור $k=1,2,3, \dots, 254$ חורג רק ב-10 פיקסלים (כלומר מכיל 90 במקום 100 פיקסלים). לכאורה הסטייה קטנה, אך לדלי האחרון יישארו עודף של 2540

פיקסלים שלא מופו ושעליהם להיות בהכרח בעלי ערך מוצא 255. אם נתבונן בהיסטוגרמה המצטברת של מקרה זה, זו תהייה בעלת שיפוע קבוע אך נמוך מהרצוי, באופן שיוצר שגיאה מצטברת גדולה מאוד. סטייה זו תתוקן באופן קיצוני ע"י קפיצה ברמת האפור האחרונה, כדי לתת סך נכון של כמות פיקסלים.



תיאור גרפי של הדוגמה שמתוארת בטקסט – בניית היסטוגרמת היעד האחידה כמילוי דליים תוך כדי התייחסות לשגיאות מצטברות

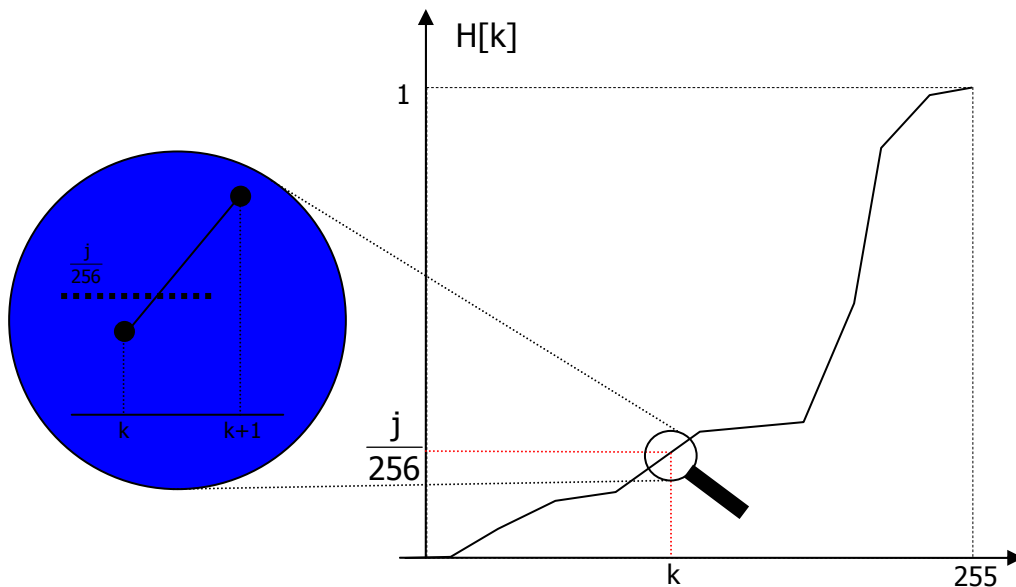
השתכנענו כי עלינו לעבוד עם היסטוגרמה מצטברת $H[k]$. כיצד נעשה זאת? מסתבר כי הדבר פשוט "יחסית". כאשר $H[k]=1/256$ פירוש הדבר הוא שרמות האפור $0,1,2,...,k$ יחדיו כוללת בדיוק $1/256$ מכלל הפיקסלים (100 במקרה שלנו). במקרה כזה יש למפות כל אינדקסים אלו לערך המוצא 0. אם, כמו קודם, אין אינדקס שלם הנותן את הקשר $H[k]=1/256$, קיימים שני אינדקסים עוקבים הנותנים

$$H[k] < 1/256 \text{ and } H[k+1] > 1/256$$

לכן, נבחר את k או $(k+1)$ – הקרוב מהשניים לערך $1/256$ – ונקבע את כל רמות האפור עד אליו להיות ממופים ל- '0' במוצא. נעבור לדלי השני

בהיסטוגרמת המוצא. הוא צריך להכיל את הנתח ה- $1/256$ הבא מכלל הפיקסלים בתמונה. נבדוק מהו k כך ש -
 $H[k] < 2/256$ and $H[k+1] > 2/256$

וכל הערכים הקטנים מאינדקס זה ושלא מופו עדיין יהיו משויכים ל-'1'. באופן זה נמשיך למלא את טבלת ה-LUT כשאנו למעשה סורקים את ערכי המוצא לפי סידורם 0, 1, 2, וכך הלאה. כנגד כל אחד מוצאים מיהם ערכי הכניסה שיוחסו אליהם. הציור הבא ממחיש רעיון זה.



תכנון ה- LUT ע"י היסטוגרמה מצטברת

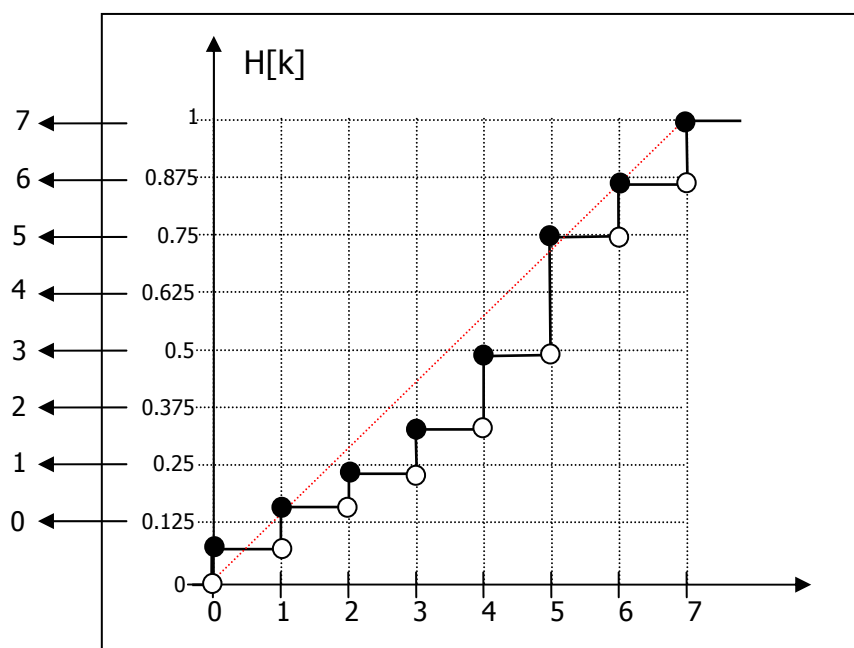
נקל על עצמנו לעת עתה ונניח כי המשוואה $H[k] = j/256$ מניבה לכל j פתרון k שלם אשר יסומן k_j - כלומר, $H[k_j] = j/256$. משמעות הדבר היא שרמות האפור $0 \leq k \leq k_1$ בכניסה ימופו ל-'0', רמות האפור $k_1 + 1 \leq k \leq k_2$ ימופו ל-'1' וכך הלאה. נשים לב כי פעולת ה- LUT אינה אלא הביטוי

$$T[k] = \lfloor 256 \cdot H[k] - \epsilon \rfloor$$

כלומר, כל שעלינו לעשות הוא לבנות את ההיסטוגרמה המצטברת $H[k]$ ולחשב ממנה ישירות את תוכן ה-LUT לפי הנוסחה הנ"ל. הצורך ב- ϵ נובע מאפקטי קצוות במעברים שבין הרמות. למעשה נוסחה זו נכונה גם למקרה בו אין פתרון שלם k למשוואה $H[k] = j/256$, כשכל ההבדל הוא שאנו נבחר בפתרון שיביא למינימום את ההפרש

$$k_j = \min_k |H[k] - j/256|$$

נמחיש את התהליך שתואר למקרה בו יש 8 רמות אפור אפשריות, לשם פשטות.



היסטוגרמה מצטברת והשימוש בה לבניית ה- LUT ליישור היסטוגרמה

K	0	1	2	3	4	5	6	7
T[k]	0	0	1	2	3	5	6	7

ה- LUT הנובע מהדוגמה הנ"ל

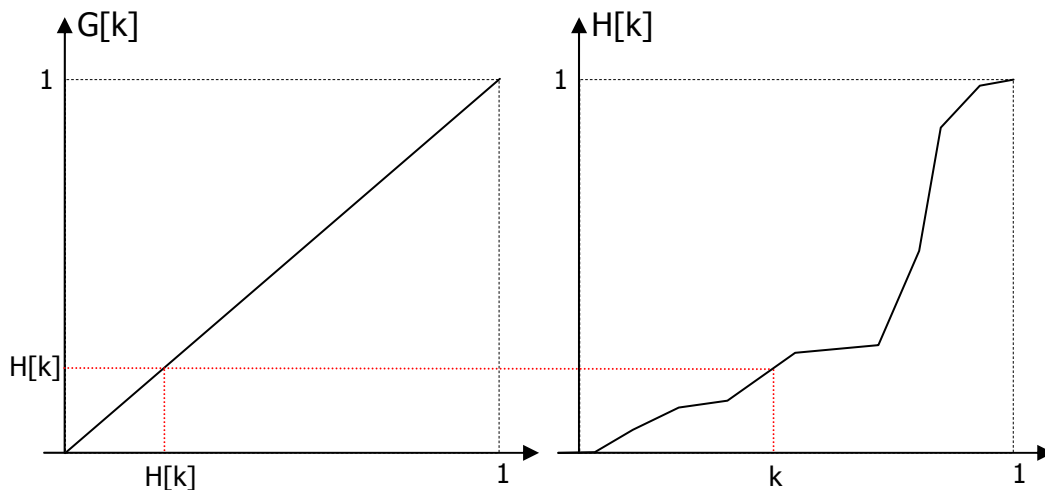
ישנה דרך אחרת להמחיש תוצאה זו, ודרך זו תסייע לנו בבניית התוצאה למקרה הכללי יותר של התאמת היסטוגרמות. לערך כניסה k פעולת ה- LUT תיתן את הערך $T[k]$. פירוש הדבר הוא שסך כל הפיקסלים בתמונה המקורית להם רמת אפור k ומטה שווה לכמות הפיקסלים בתמונת המוצא להם רמת אפור שווה או קטנה מ- $T[k]$. במשוואה נאמר כי מתקיים

$$H[k] = G[T[k]]$$

הפונקציה $G[k]$ היא היסטוגרמת המוצא המצטברת האמורה להיות פונקציית יחידה - $G[k] = k$. לכן מתקבלת המשוואה

$$H[k] = G[T[k]] = T[k]$$

וזו התוצאה שנבנתה קודם. הציור הבא ממחיש קשר זה באופן גרפי. נשים לב לכך שההכפלה ב- 255 חסרה כי בציור אנו מניחים רמות אפור בתחום $[0,1]$.



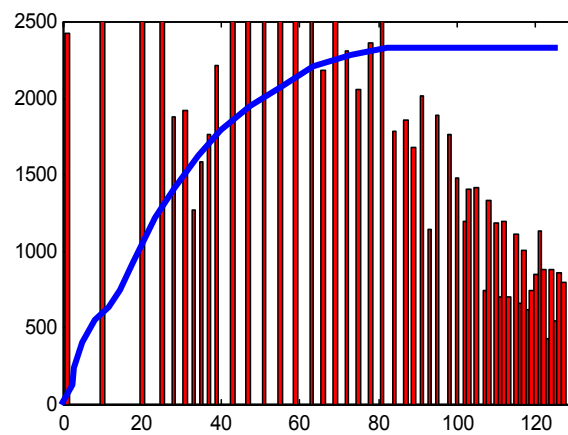
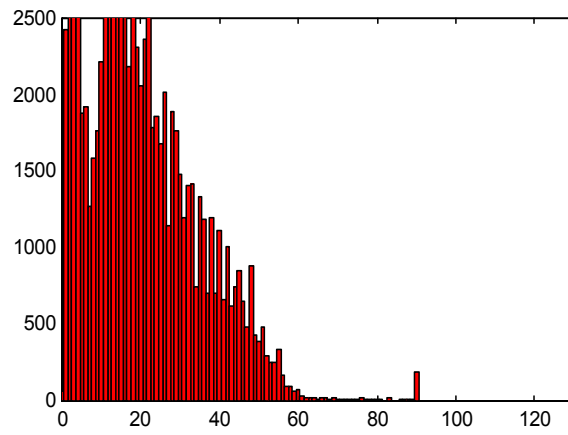
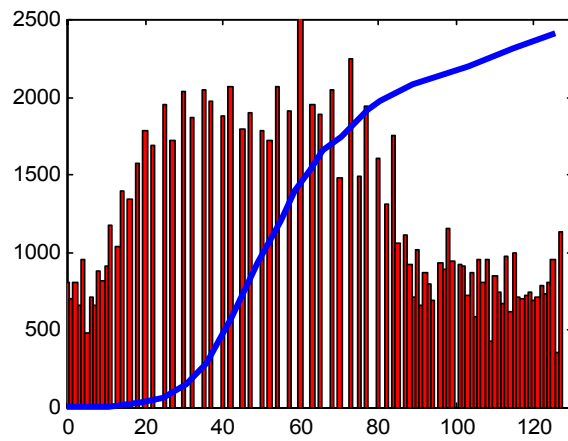
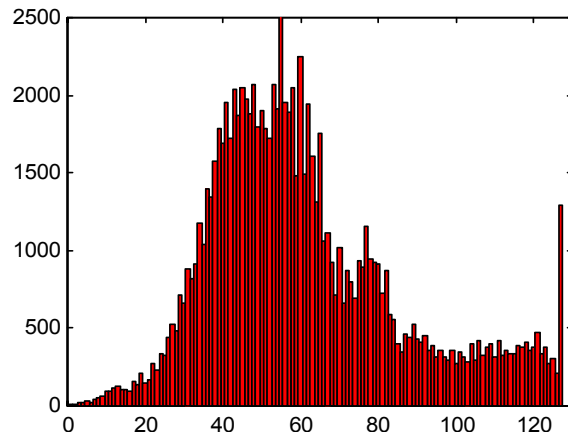
היסטוגרמות מצטברות – כניסה ויציאה, והדרך לבנות מתוכן את ה- LUT המיישר

למעשה, בהחלפה פשוטה של היסטוגרמת היעד (המצטברת) לכלשהי אחרת נאמר כי התאמת היסטוגרמות מ- $h[k]$ ל- $g[k]$ תיעשה ע"י פעולת LUT שתהיה

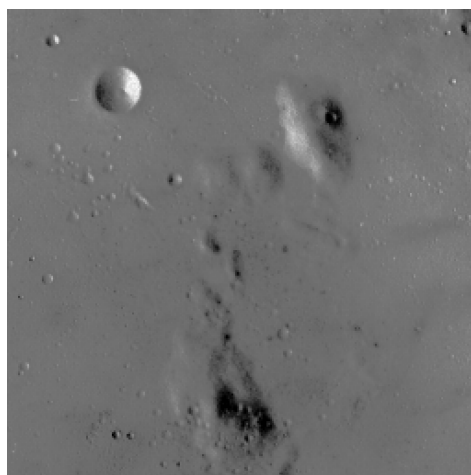
$$H[k] = G[T[k]] \Rightarrow G^{-1}\{H[k]\} = T[k]$$

כצעד אחרון במסגרת זו, נראה מספר דוגמאות להמחשת תוצאות תהליך יישור התאמת ההיסטוגרמה. בציורים נתונה תמונת מקור ותוצאת יישור היסטוגרמה. הגרף הכחול מתאר את ה- LUT הנדרש (בכל המקרים, פיקסל מיוצג ע"י 7 סיביות).

חשוב לומר כי יישור התאמת היסטוגרמה אינו בהכרח משפר את התמונה, וקורה לא אחת שהתמונה אף נזוקה. תמונות בהן לא צפוי פיזור אחיד של רמות אפור הן תמונות בהן תהליך יישור ההיסטוגרמה צפוי להזיק. זה נכון למשל לתמונות בהן צילום של גרפיקה/טקסט הנראות כמעט בינאריות, תמונות בהן גוון אחד דומיננטי מאוד, תמונות בהן מטרות קטנות יחסית אך חשובות (בשל גודלן, אין להן השפעה על היישור, אך הן מושפעות ממנו בפועל).

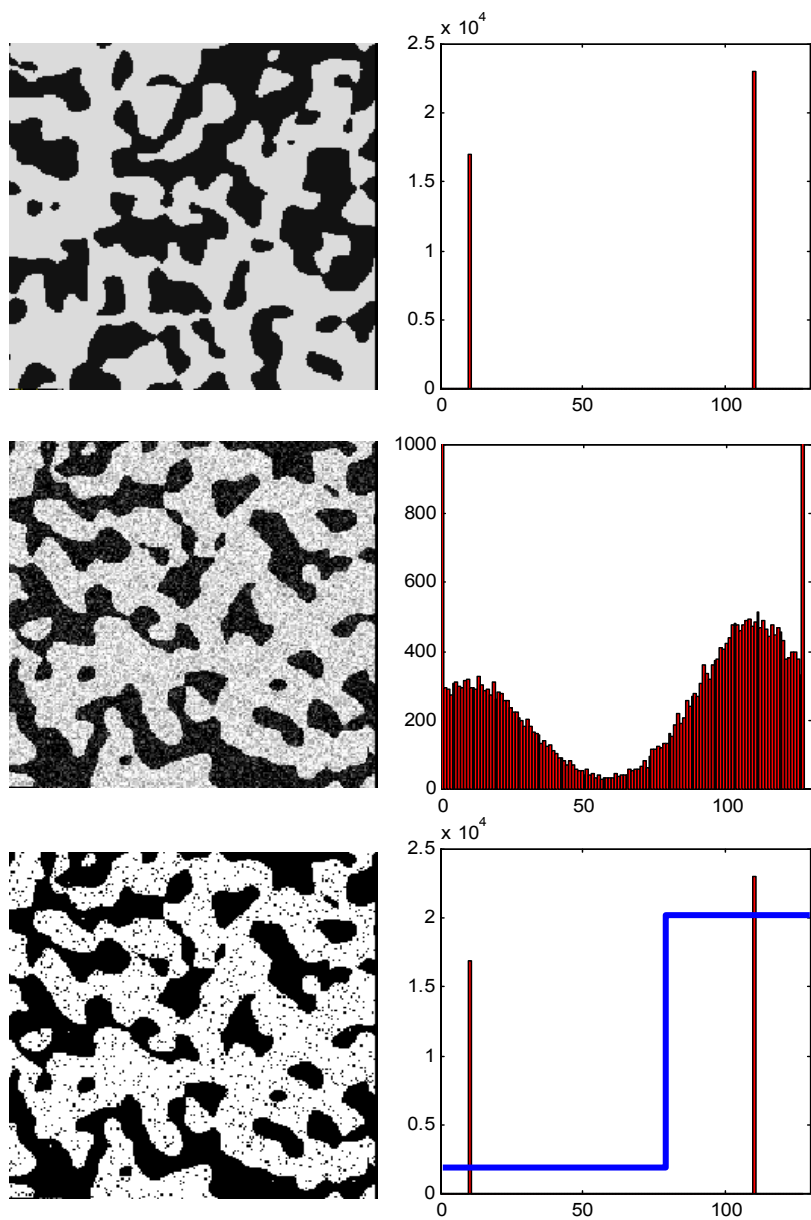


תמונת מבחן ותוצאת יישור היסטוגרמה עבורה (בכחול – ה- LUT הנדרש)



תוצאות יישור היסטוגרמה

הציור הבא מראה תהליך של התאמת היסטוגרמות. נתונות שתי תמונות בעלות אופי דומה (שורות עליונה ואמצעית בציור), כשאחת מהן מזוהמת, ונדרש לנקותה. לוקחים את התמונה המקולקלת ומבצעים עליה פעולת LUT לשם קבלת היסטוגרמה זהה לזו של התמונה הנקייה. התוצאה מוצגת בשורה התחתונה, ורואים כי אמנם ההיסטוגרמה במוצא התהליך זהה להיסטוגרמת הייחוס. התמונה נקייה יותר, ובעצם, הפעולה שנמצאה להיות היא השוואה לסף כלשהו, כמתואר בגרף הכחול בציור.



המחשת תהליך להתאמת היסטוגרמות לשם ניקוי תמונה. שורה עליונה – תמונת יחוס עם ההיסטוגרמה שלה. שורה אמצעית – תמונה הדורשת עיבוד. שורה תחתונה – תוצאת התאמת ההיסטוגרמה של התמונה העליונה לזו האמצעית

עיבוד תמונות בעזרת מחשב (236327)

הרצאה מס' 8

שיפור תמונות – גישות מרחביות

פעולות מרחביות בסיסיות

עד כה עסקנו בפעולות LUT אשר הינן פעולות חסרות זיכרון מרחבי - ערך מוצא המערכת תלוי בערך הסקלארי של ערך הכניסה באותו מקום. יתרונה של פעולת LUT בפשטותה, אך ברור כי מגוון התוצאות הניתנות להשגה באמצעותה מוגבל למדי. פעולה כללית מעט יותר אשר עדיין תהיה פשוטה יחסית היא פעולה מרחבית הפועלת על סביבת שכנים מצומצמת. כלומר, מוצא המערכת הפעם יהיה תלוי בפיקסל הכניסה ומספר פיקסלים שכנים לו. פעולה כזו יכולה להיות ליניארית או לא ליניארית, קבועה או משתנה במקום, ופעולת LUT הופכת להיות מקרה פרטי של מערכת מעין זו. במסגרת סעיף זה נתמקד בפעולות קבועות במקום (ליניאריות ושאינן ליניאריות).

באופן כללי, אם אנו דנים במערכת ליניארית על סביבה סמוכה לפיקסל, אנו בעצם דנים במסנן FIR (Finite Impulse Response – מסנן בו התמך של תגובת ההלם סופי) דו-מימדי. ואמנם, ברוב מוחלט של המקרים, סינון ליניארי של תמונה נעשה ע"י מסנני FIR. מסנן כזה יאופיין ע"י תגובתו להלם אשר ניתנת לתיאור כסדרת המשקלות המכפילות כל פיקסל בסביבה, המתוארות במטריצה דו-מימדית הקרויה גרעין (Kernel). נמחיש זאת דרך דוגמה.

דוגמה: אם פעולת הסינון על הפיקסל $[m,n]$ תהיה

$$\text{Out}[m,n] = \text{In}[m,n]a_1 + \text{In}[m-1,n]a_2 + \\ + \text{In}[m,n-1]a_3 + \text{In}[m+1,n]a_4 + \text{In}[m,n+1]a_5$$

אזי הגרעין של מסנן ה-FIR נתון ע"י

$$K = \begin{bmatrix} 0 & a_2 & 0 \\ a_3 & a_1 & a_5 \\ 0 & a_4 & 0 \end{bmatrix}$$

כאשר הראשית במישור האינדקסים מצויה למעלה בצד שמאל, והגרעין מופעל על התמונה ע"י הנחתו כמסכה במקום $[m,n]$ וחישוב סכום המכפלות בין המשקלות וערכי הפיקסלים של התמונה המטופלת. איבר האמצע בגרעין מתייחס למיקום $[m,n]$.

השימוש במסנן לשיפור תמונות רווח לשלושה יישומים עיקריים - ניקוי תמונה מרעש, חידוד תמונה, ומציאת שפות בתמונה. באשר לדיכוי רעש בתמונה: רעש יכול להצטרף לתמונה כתוצאה מלכלוך על תמונה שאמורה להיסרק, הפרעות שידור לתמונה מועברת בערוץ תקשורת, ועוד. בדרך כלל מקובל להניח מודל אדיטיבי (חיבורי) לפיו התמונה הרועשת $f[m,n]$ הינה סכום של התמונה המקורית $s[m,n]$ ותמונת רעש $d[m,n]$, כלומר

$$d[m,n] + f[m,n] = s[m,n]$$

ישנן גם מערכות צילום בהן הרעש הינו דווקא מולטיפליקטיבי (מכפלתי), בו הרעש כופל את התמונה האידיאלית, אך אנו לא נעסוק בהן במסגרת זו.

בדרך כלל אות הרעש יונח להיות אקראי בעל ממוצע אפס, הומוגני ולבן. משמעות הנחה זו היא שערכי הפיקסלים בתמונת הרעש d מפולגים לפי אותו חוק פילוג בכל מקום בתמונה, לחוק פילוג זה ממוצע אפס ועוצמה (ווריאנס) קבועה, ואין קורלציה בין פיקסלים סמוכים. במלים אחרות - מתוך ידיעת ערכו של $d[m,n]$ לא נוכל להקיש דבר על ערכי שכניו. בדרך כלל אף נהוג להניח כי הרעש הוא בעל חוק פילוג גאוס.

הערה: המושג רעש מתאים יותר לאות שמע, וזהו באמת מקורו. הפרעה באות דיבור למשל תהווה "רעש". בתמונות, לפי הגיון זה, הפרעה הייתה צריכה להיקרא "לכלוך". בפועל, המושג רעש תפס גם בתמונות, והכוונה לאות אקראי לא רצוי.

הבעיה אם כך היא לנסות ולבנות את התמונה f מתוך ידיעת התמונה s , ואת זאת לעשות ע"י סינון ליניארי. בניגוד לרעש, ידיעת ערכו של הפיקסל $f[m,n]$ מלמדת הרבה על ערכי שכניו, כיוון שכפי שכבר ראינו בפרק ראשון, ערכי פיקסלים סמוכים הם לרוב קרובים (ואולי אפילו זהים). זוהי בדיוק האבחנה שבזכותה אנו יכולים לנקות רעש מתמונה.

נתחיל את הניתוח מתוך הנחה שערכי השכנים צריכים להיות זהים, ונניח שהמסנן פועל בגרעין בגודל של 3 על 3 פיקסלים. כשאנו מסתכלים בחלון של 3 על 3 סביב הנקודה $[m,n]$, אנו מחפשים מהו הערך האחד $f[m,n]$ שיחליף את $s[m,n]$ ואשר יהיה קרוב לכל תשעת הערכים בחלון. למעשה, אנו מחפשים את המינימום לביטוי הבא:

$$\varepsilon\{f[m,n]\} = \sum_{k=-1}^1 \sum_{j=-1}^1 (s[m-k, n-j] - f[m,n])^2 \rightarrow \text{Minimum}$$

והתוצאה נתונה ע"י

$$\frac{\partial \varepsilon\{f[m,n]\}}{\partial f[m,n]} = -2 \sum_{k=-1}^1 \sum_{j=-1}^1 (s[m-k, n-j] - f[m,n]) = 0$$

$$\Rightarrow f[m,n] = \frac{1}{9} \sum_{k=-1}^1 \sum_{j=-1}^1 s[m-k, n-j]$$

כלומר, יש להפעיל על התמונה מסנן ליניארי בעל הגרעין

$$K = \frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

כשמסנן זה פועל על $s[m,n]$ (אשר כזכור הוא $s[m,n] = f[m,n] + d[m,n]$), הוא נותן סכום של שני ממוצעים - מיצוע מקומי של התמונה f ומיצוע מקומי של התמונה d . אם אמנם ב- f הערכים בסביבת 3 על 3 קרובים מאוד זה לזה, הרי שממוצע מקומי של f כמוהו כ- f עצמו. לעומת זאת, מיצוע של 9 משתנים אקראיים שווי פילוג הוא משתנה אקראי חדש בו השונות קטנה פי $3 = \sqrt{9}$, וכך יצא שעוצמת הרעש הוקטנה.

מדוע אם כך להסתפק בחלון של 3 על 3 - אם נעבוד על חלון של 5 על 5 נקבל דיכוי עוצמת הרעש בפקטור 5. אבל - עם הגדלת החלון, הנחתנו על ערכי רמות אפור קבועים ב- f נשברת והולכת. וכך נקבל כי אמנם הרעש ייקטן יפה, אך עמו נקבל גם אפקט של טשטוש התמונה. הציור הבא מראה תוצאת הסינון עם חלונות בגודל שונה, וכפי שרואים, הרעש אמנם פוחת, אך עמו גם פירטי התמונה. האזורים בהם הכי גדול הנזק אם אזורי המעברים (שפות בתמונה), בהם הנחת רמת אפור קבועה היא הנחה שגויה.



ניקוי רעש ע"י מסנן ליניארי - שמאל למעלה - תמונת מקור. ימין למעלה - תמונה עם רעש גאוס ($\sigma = 15$) (רמות אפור עד 255). שמאל למטה - ניקוי ע"י חלון 3 על 3. ימין למטה - כנ"ל עם חלון 5 על 5.

כדי לראות אם שיפרנו, נוכל לחשב את ההפרש בין תמונת המקור $f[m,n]$ לכל אחת מהתמונות הנ"ל (עם רעש, סינון ע"י 3 על 3 וסינון ע"י 5 על 5), ולסכם את ריבועי הפרשים אלו. רצוי גם לחלק במספר הפיקסלים בתמונה ולהוציא שורש מהתוצאה, על מנת לקבל מספר ביחידות של רמות אפור. הנוסחה אם כך היא:

$$\text{Error} = \sqrt{\frac{1}{MN} \sum_{m=1}^M \sum_{n=1}^N (f[m,n] - \hat{f}[m,n])^2}$$

כאשר $\hat{f}[m,n]$ היא התוצאה שלנו המנסה לקרב את $f[m,n]$. קריטריון זה קרוי שגיאה ריבועית ממוצעת (Mean Squared Error - MSE), והוא רווח בעיבוד תמונות כמדד לאיכות תהליכי עיבוד. ברור כי שימוש בו אפשרי רק כאשר זמינה לנו תמונת המקור (בטרם הצטרפות הרעש). ברור כי ככל שערכו של סכום זה קטן יותר השיפור טוב יותר. התוצאה למקרה הנ"ל נותנת כי השגיאה המיוחסת

לתמונה הרועשת היא 15.01 - זוהי כמעט בדיוק השונות בה השתמשנו. לתמונה לאחר סינון 3 על 3 התקבלה שגיאה של 9.8, ולסינון 5 על 5 שגיאה של 13.38. בשני המקרים חל שיפור, אך סינון 3 על 3 טוב יותר. נשים לב כי אם התמונה הייתה קבועה לגמרי היינו צפויים לקבל שגיאות של 5 ו- 3 בקירוב טוב (לפי יחסי דיכוי רעש בלבד 3 ו- 5). בשל היות התמונה לא קבועה, השגיאה גדולה יותר, וזו הולכת ומחריפה עם הגדלת החלון.

שאלה מעניינת היא השאלה הבאה: עברנו מחלון של 1×1 (זהו המקרה בו בעצם לא הופעל סינון), לחלון בגודל 3×3 , וממנו לחלון בגודל 5×5 , כשבכל המקרים קיבלנו מסנן אחד (ערכים זהים בגרעין). האם אנו יכולים להציע דרך אותה גישה מגוון רחב יותר של פילטרים? התשובה לכך נעוצה בביטוי שאותו הבאנו למינימום. אם נכניס לשגיאות הריבועיות משקלות לפי החשיבות לכל פיקסל

$$\varepsilon\{f[m,n]\} = \sum_{k=-L}^L \sum_{j=-L}^L c[k,j] (s[m-k,n-j] - f[m,n])^2 \rightarrow \text{Minimum}$$

נקבל:

$$\frac{\partial \varepsilon\{f[m,n]\}}{\partial f[m,n]} = -2 \sum_{k=-L}^L \sum_{j=-L}^L c[k,j] \cdot (s[m-k,n-j] - f[m,n]) = 0$$

$$\Rightarrow f[m,n] = \frac{\sum_{k=-L}^L \sum_{j=-L}^L c[k,j] \cdot s[m-k,n-j]}{\sum_{k=-L}^L \sum_{j=-L}^L c[k,j]}$$

והפעם הגרעין יהיה:

$$K = \frac{1}{\sum_{k=-L}^L \sum_{j=-L}^L c[k,j]} \begin{bmatrix} \ddots & & \vdots & & \ddots \\ & c[-1,-1] & c[-1,0] & c[-1,1] & \\ \dots & c[0,-1] & c[0,0] & c[0,1] & \dots \\ & c[1,-1] & c[1,0] & c[1,1] & \\ \ddots & & \vdots & & \ddots \end{bmatrix}$$

כיצד נקבע את המשקלות? גישה אפשרית אחת היא כערכים הדועכים לפי המרחק האויקלידי מהמרכז. קצב הדעיכה יקבע בפועל את התמך האפקטיבי של המסנן שיווצר (רוחבו). למשל, לחלון 3 על 3 ניתן לקבוע את המשקלות כ:

$$c[-1,-1]=c[-1,1]=c[1,-1]=c[0,1]=c[1,0]=c[-1,0]=c[0,-1]=0.5, c[0,0]=1$$

$$1]=c[1,1]=0.25$$

והגרעין יהיה:

$$K = \frac{1}{4} \begin{bmatrix} 0.25 & 0.5 & 0.25 \\ 0.5 & 1 & 0.5 \\ 0.25 & 0.5 & 0.25 \end{bmatrix}$$

את התהליך הנ"ל יכולנו לתאר מתוך שיקולים ספקטריים, כלומר תכנון המסנן כמסנן LPF דו-מימדי במישור התדר. נראה את עיקרי הדברים כאן: לומר שהתמונה $f[m,n]$ קבועה בקירוב במקום שקול לאמירה שתמונה זו בעלת התמרת תדר הדועכת מהר לאפס כשמתרחקים מהראשית (כלומר יש ספקטרום עם תדרים נמוכים בלבד). לעומת זאת, הרעש הוא בעל ספקטרום אחיד בכל התדרים (מכאן למעשה בא המושג "רעש לבן" - לבן הוא צבע העושה שימוש בכל אורכי הגל באופן אחיד). לכן, סינון הרעש תוך כדי אי-פגיעה בתמונת המקור מחייב מסנן LPF שאינו מזיק כמעט בכלל לתמונה $f[m,n]$. לכן, המסנן לא יכול להיות LPF צר כרצוננו, אלא תלוי בהתמרת התדר של התמונה $f[m,n]$.

מסנן חציון

בקריטריון שהוצע לעיל לבניית המסנן העלינו כל שגיאה בריבוע. ניתן כתחליף להגדיר את השגיאה ע"י ערך מוחלט כ:

$$\varepsilon\{f[m,n]\} = \sum_{k=-L}^L \sum_{j=-L}^L |s[m-k, n-j] - f[m,n]| \rightarrow \text{Minimum}$$

השינוי הוא במהותו שינוי המטריקה המודדת את המרחק בין שתי תמונות, והמטריקה החדשה אינה פחות נכונה מקודמתה. הבאת ביטוי זה למינימום מניבה את המסנן הידוע בשם סינון חציון (Median).

$$\frac{\partial \varepsilon\{f[m,n]\}}{\partial f[m,n]} = -2 \sum_{k=-L}^L \sum_{j=-L}^L \text{sign}(s[m-k, n-j] - f[m,n]) = 0$$

נניח כי אוסף ה- $(2L+1)^2$ הפיקסלים בחלון סודרו בסדר עולה לפי ערכיהם (ללא כל חשיבות למקומם). אזי, הערך שיביא את הביטוי הנ"ל למינימום הוא

הערך האמצעי - $(2L^2 + 2L + 1)$. למשל, עבור חלון 3 על 3 (בו $L=1$), האיבר האמצעי הוא החמישי מתוך 9. בחלון 5 על 5 יהיה זה האיבר ה-13. מדוע האיבר האמצעי דווקא? כיוון שאז מחצית מהאיברים בסכימה הנ"ל בעלי סימן שלילי, מחצית בעלי סימן חיובי, ואיבר האמצע נותן ערך אפס. סכימת הסימנים תיתן התאפסות הנגזרת כנדרש.

המסנן שהתקבל - סינון חציון - אינו ליניארי. מסנן זה שייך למשפחה של מסנני סדר (Order Filters), בהם מוצא המסנן תלוי בסידור (Sorting) של ערכי הכניסה. ניתן גם להציע בחירה של המקסימלי/מינימלי בחלון, וישנם יישומים העושים שימוש במסננים מעין אלו.

אז איזה סינון יותר טוב - ליניארי או חציון? התשובה תלויה בטיבו של הרעש האדיטיבי (מהו חוק הפילוג שלו), מהי הגישה לבחון את התוצאות (האם MSE או מדד אחר) ועוד. ככלל, מקובל כי ברעשי שיאים סינון חציון יותר טוב. רעש שיאים, הקרוי גם רעש מלח ופלפל Salt & pepper (נשמע יותר אמין משום מה), הוא רעש בו פיקסלים במיקום אקראי משנים את ערכם לשחור/לבן מוחלט באופן אקראי. בדוגמה הבאה ננסה להמחיש מדוע סינון מדיין טוב כל כך בטיפול ברעשי שיאים.

דוגמה: נניח כי אנו משווים סינון ליניארי אחד וסינון חציון בחלון של 3 על 3. בחלון מסוים התקבלו הערכים הבאים - {50 55 60 60 65 65 70 75 75} (תשעה ערכים ומיקומם לא חשוב). ערך הממוצע יהיה $64 \approx$ והחציון יהיה 65 - אין הבדל גדול. נניח כעת כי הערך 75 השתנה כתוצאה מרעש וערכו גדל ל-255. הממוצע החדש יהיה 84, בעוד שהחציון לא ישתנה כלל! כל שינוי באחד המספרים יגרום לשינוי מידי (אך מוקטן) בתוצאת הממוצע. לעומת זאת, בחישוב החציון שינוי של אחד מארבעת הערכים הנמוכים מטה או שינוי של אחד מארבעת הגבוהים מעלה לא תשנה דבר. כמקרה שני, נניח כי חוזרים לערכים המקוריים והערך 60 עובר להיות 255. במקרה כזה, החציון יהיה עדיין 65, ואילו הממוצע האלגברי ייתן 85 - שוב שינוי משמעותי.

בציור הבא מתוארת תמונה אשר אליה צורף רעש Salt & pepper בה 20% מהפיקסלים "נגועים" ברעש שיאים לשני הכיוונים (שחור ולבן בהסתברות

שווה). כפי שניתן לראות, מסנן חציון מצליח להתגבר על מרבית הרעש תוך שמירת מרבית הפרטים בתמונה, בעוד שהמסנן הליניארי מניב תוצאה ירודה למדי באיכותה.



סינון רעש ע"י מסנן ליניארי ומסנן חציון. עמודה שמאלית - מקור (למעלה) ותמונה רועשת. עמודה אמצעית - חציון 3 על 3 (למעלה) וחציון 5 על 5 (למטה). עמודה ימנית - סינון ליניארי 3 על 3 (למעלה) וסינון ליניארי 5 על 5 (למטה).

חשוב לציין כי כאשר מניחים ידע כלשהו על סטטיסטיקת הרעש האדיטיבי (כגון האמירה - רעש לבן גאوسی וכו'), דיכוי הרעש יכול להיעשות מתוך שיקולים תיאורטיים מוצקים יותר, בהשוואה לגישה שתוארה כאן. אנו נראה זאת במסגרת פרק הבא, בו נדבר על שחזור תמונות.

גזירת תמונה וחידוד

משימה שנייה שמסננים (ליניאריים ושאינם ליניאריים) מגשימים היא פעולת חידוד תמונה. המידע העיקרי בתמונה הם שפות העצמים המופיעים בה, ולכן הדגשת שפות היא משימה חשובה. בהסתכלות בתחום התדר, שפות הם מידע בעל תדרים גבוהים (בגלל המעברים החדים), בעוד שפנים העצמים

מיוחס לתדרים נמוכים. לכן, חידוד שפות פירושו מסנן אשר יעבוד על התדרים הגבוהים וידגיש אותם. זוהי בדיוק פעולה הפוכה לפעולת ניקוי הרעש ע"י LPF כפי שתוארה קודם. לכן, חידוד תמונה בא בחשבון רק עבור תמונות נקיות יחסית מרעש. העיקרון באלגוריתמי חידוד תמונה הוא להעביר את התמונה דרך HPF (מסנן מעביר גבוהים) ולחבר את מוצא מסנן זה לאחר מישקול נאות לתמונה המקורית. השאלה המתבקשת היא - כיצד מממשים HPF ? ובכן, HPL הוא עצם LPF-1, וליישם LPF אנו כבר יודעים. לדוגמה, עבור ה-LPF הפשוט בו מוצע גרעין של 3 על 3 אחד, ה-HPL יהיה

$$K_{HPF} = 1 - K_{LPF} = 1 - \frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} = \frac{1}{9} \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$

מסנן HPF יהיה תמיד בעל סכום איברים השווה 0 (למה?). לכן, המסנן הכולל שיבצע את פעולת החידוד במקרה זה יהיה:

$$K_{Sharp} = 1 + K_{HPF} = 1 + \frac{1}{9} \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix} = \frac{1}{9} \begin{bmatrix} -\alpha & -\alpha & -\alpha \\ -\alpha & 9+8\alpha & -\alpha \\ -\alpha & -\alpha & -\alpha \end{bmatrix}$$

עבור $\alpha = 0$ נקבל את תמונת המקור בדיוק, וכש- α חיובי וגדל, נקבל אפקט חידוד חזק יותר ויותר, ומנקודה מסוימת ואילך, תתקבל תוצאה כה חדה שתיראה לא טבעית ולא יפה. הציור הבא מראה תוצאת חידוד תמונה למגוון ערכי α עם המסנן הנ"ל. כתמונת מקור בחרנו לקחת את Lena לאחר טשטוש קל.

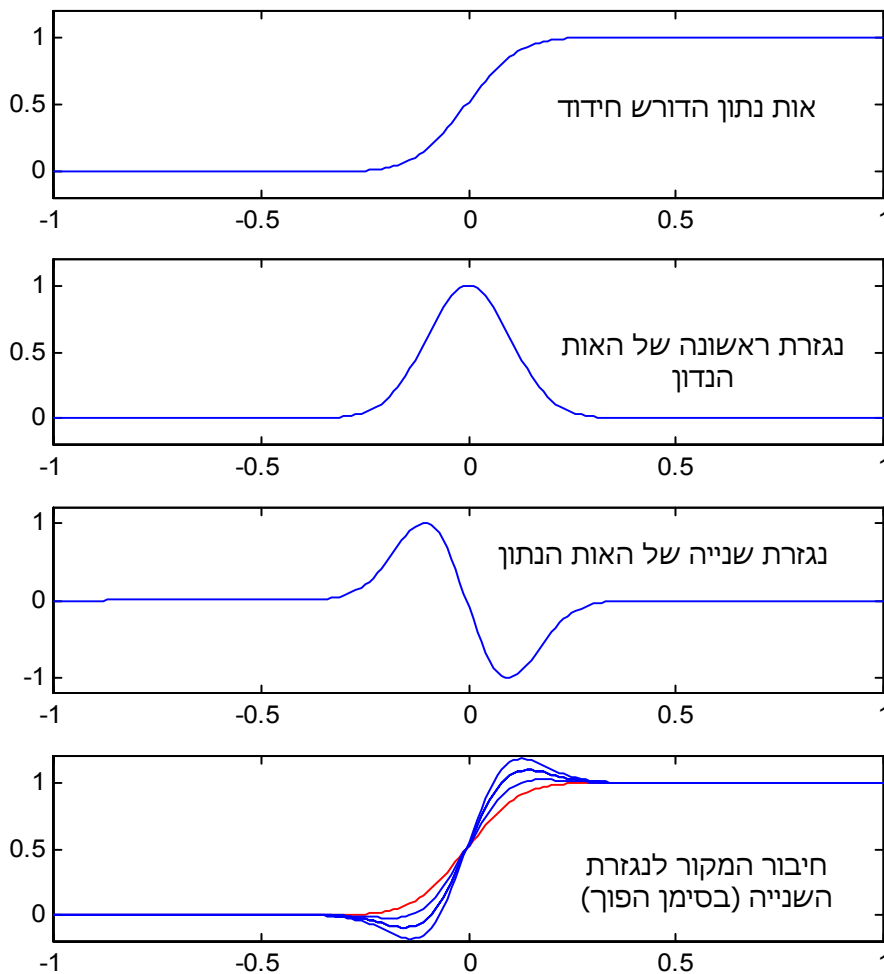


תמונת מקור (שמאל למעלה) ותוצאות חידוד שפות למגוון ערכי α - 0.1 (ימין למעלה), 0.3 (שמאל למטה), ו-1 (ימין למטה).

מסנן שרווח בו השימוש כמסנן HPF לטובת חידוד שפות הוא אופרטור הלפלסיאן הנתון ע"י

$$K_{\text{Laplacian}} = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix}$$

כפי שנראה בהמשך, לאופרטור זה קשר הדוק לאופרטור הנגזרת השנייה הרציף. מה קורה בתהליך החידוד? הציור הבא ממחיש את התהליך עבור אותות חד-ממדיים.



המחשת תהליך החידוד לאותות חד-מימדיים

אנו רואים שפה מטושטשת, ורואים כי נגזרת שנייה המוחסרת מאות המקור נותנת אפקט של מעבר חד יותר, כנדרש. מדוע סימן שלילי לנגזרת? מדוע נגזרת שנייה? לעת עתה שאלות אלו פתוחות כיוון שלא ענינו על שאלה בסיסית יותר - מה הקשר בין נגזרות (ראשונה ושנייה) ומסנן ה-HPF שרצינו מלכתחילה? בכל אופן, אינטואיטיבית נוכל לומר את הדבר הבא - הפעלת

מסנן כדוגמת הלפלסיאן נותנת ערך קרוב לאפס באזורים בהם רמת האפור מהווה מישור מוטה. במקומות אלו לא יורגש כלל החידוד. במקומות בהם תוצאת הלפלסיאן גבוהה משמעותית מאפס, פירוש הדבר הוא שהפיקסל הנוכחי גבוה יותר מממוצע ארבעת שכניו. כלומר - ישנו שם פרט. על מנת לחדדו, יש להגדיל עוד את ההפרש בין הפיקסל לבין שכניו, וזוהי בדיוק הפעולה המתבצעת. תהליך דומה קורה לערכים שליליים.

נעבור כעת לדון במציאת שפות בתמונה, וקירוב נגזרות מעל שריג דיסקרטי דו-מימדי. כזכור, אנו יכולים להתייחס לתמונה כמשטח (פונקציה) מעל המישור, כשרמת האפור $f(x,y)$ היא גובהו של המשטח בנקודה (x,y) . לפונקציה זו אנו יכולים לחשב בנקודה (x,y) את הגרדיאנט בכל נקודה. הגרדיאנט הוא וקטור הבנוי מהנגזרת האופקית והנגזרת האנכית, אשר ברצף מוגדר לפי:

$$\nabla f(x,y) = \frac{\partial f(x,y)}{\partial x} \hat{x} + \frac{\partial f(x,y)}{\partial y} \hat{y}$$

לוקטור הגרדיאנט חשיבות רבה. חישוב ערכו המוחלט יכול לומר לנו האם הנקודה בה אנו עומדים היא שפה או לא, שהרי נורמה גדולה לגרדיאנט פירושה שיפוע חזק. נורמה זו נתונה ע"י:

$$\|\nabla f(x,y)\| = \sqrt{\left(\frac{\partial f(x,y)}{\partial x}\right)^2 + \left(\frac{\partial f(x,y)}{\partial y}\right)^2}$$

אופרטור הלפלסיאן מתקבל ע"י:

$$\Delta f(x,y) = \nabla^2 f(x,y) = \frac{\partial^2 f(x,y)}{\partial x^2} + \frac{\partial^2 f(x,y)}{\partial y^2}$$

כך שאם נחזור למה שהצגנו קודם, חידוד הפונקציה, לו היה נעשה ברצף, היה מתבצע ע"י

$$g(x,y) = f(x,y) - \alpha \Delta f(x,y)$$

כל האמור לעיל טוב ויפה אך אינו מועיל בבואנו לטפל בתמונות דיסקרטיות. לכן, עלינו למצוא דרך לקרב נגזרות באות דגום. הרעיון הבסיסי פשוט עד טריוויאלי. כיוון שנגזרת ברצף מייצגת הפרשים בין ערכים סמוכים במרחק אינפיניטסימלי וחלוקה במרחק זה, הצעה מתבקשת היא לקבוע את הנגזרת האופקית, למשל, כהפרש בין דגימות סמוכות באותה שורה, תוך הנחה

שהמרחק ביניהם הוא 1. הבעיה היא שכשאנו עומדים על הדגימה $[m,n]$, ישנן שתי אפשרויות:

$$\frac{\partial f}{\partial x} \rightarrow f[m,n] - f[m-1,n] \text{ or } f[m+1,n] - f[m,n]$$

האפשרות הראשונה קרויה גזירה אחורית, והשנייה גזירה קדמית. באופן דומה לגזירה אנכית:

$$\frac{\partial f}{\partial y} \rightarrow f[m,n] - f[m,n-1] \text{ or } f[m,n+1] - f[m,n]$$

כתחליף להגדרות הנ"ל ניתן להציע גזירה מרכזית לפי:

$$\frac{\partial f}{\partial x} \rightarrow \frac{1}{2} [f[m+1,n] - f[m-1,n]]$$

$$\frac{\partial f}{\partial y} \rightarrow \frac{1}{2} [f[m,n+1] - f[m,n-1]]$$

בהמשך ישיר לקו הגיון זה, גזירה שנייה היא הפרש על הנגזרת עצמה. לכן, גזירה שנייה אופקית תהיה:

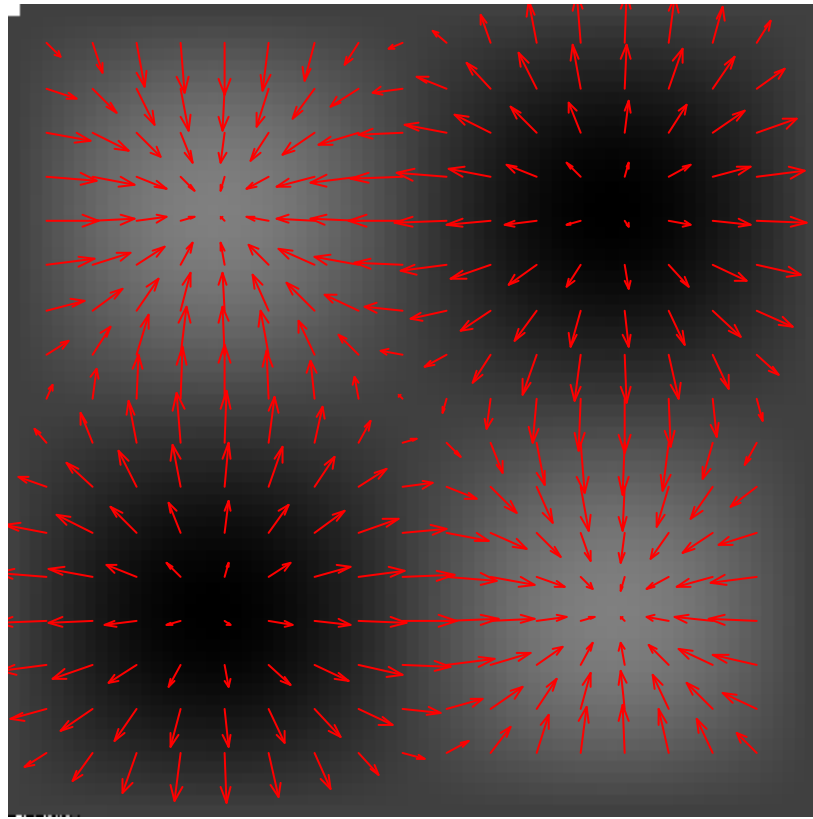
$$\begin{aligned} \frac{\partial^2 f}{\partial x^2} &\rightarrow (f[m+1,n] - f[m,n]) - (f[m,n] - f[m-1,n]) = \\ &= f[m+1,n] + f[m-1,n] - 2f[m,n] \end{aligned}$$

כשהפעם עשינו שימוש בגזירה הקדמית והאחורית במשולב על מנת לקבל נגזרת שנייה מאוזנת (כלומר כזו שממוקצת במקום הנכון). לפי האמור לעיל, אופרטור הלפלסיאן הדיסקרטי מתקבל ע"י:

$$\begin{aligned} \Delta f[m,n] &= (f[m+1,n] - f[m,n]) - (f[m,n] - f[m-1,n]) + \\ &\quad + (f[m,n+1] - f[m,n]) - (f[m,n] - f[m,n-1]) = \\ &= f[m+1,n] + f[m-1,n] + f[m,n+1] + f[m,n-1] - 4f[m,n] \end{aligned}$$

וכבר ראינו מסנן זה קודם לכן, כשדנו בחידוד.

בציור הבא מוצגת תמונה ועליה חיצו הגרדיאנט, אשר חושבו ע"י נגזרת מרכזית. כצפוי, חיצו הגרדיאנט מצביעים לכיוון העלייה המרבית במשטח התמונה.



תמונה ועליה מפת וקטורי הגרדיאנט שחושבו נומרית לפי נגזרת מרכזית

הציור הבא מראה תמונה ומציאת שפותיה ע"י חישוב הגרדיאנט והשוואת הנורמה שלו לסף. הסף נבחר כך שאחוז שנקבע מראש של ערכים יעבור אותו. הציור מראה העברה של 10%, 15% ו- 20% מכלל הפיקסלים, וניכר כי השפות בתמונה מתגלות יפה, ובאופן רובסטי לערכו של הסף.

כנקודה אחרונה בסעיף זה נזכיר כי חישוב נגזרות רגיש מאוד לרעש אדיטיבי. אין זה מפתיע כיוון שרעש מתבטא חזק יותר בתדרים הגבוהים, בהם עיקר האינפורמציה היא הרעש ולא תמונת המקור. פעולת גזירה היא פעולת חיזוק התדרים הגבוהים, ולכן אם רעש קיים, הוא יוגבר ויזיק. על מנת להתגבר על בעיה זו מקובל לבצע סינון מקדים על התמונה לשם דיכוי הרעש בה, בטרם חישוב הנגזרות. אם



גילוי שפות ע"י נורמת הגרדיאנט. תמונת מקור (שמאל למעלה), תמונת גרדיאנט לאחר סף המעביר 20% מהפיקסלים (ימין למעלה), העברת 15% מהפיקסלים (שמאל למטה), ו-10% מהפיקסלים (ימין למטה).

הסינון המקדים הוא ליניארי (כפי שמקובל) והגזירה ליניארית (וכל הנגזרות שהראינו הינן ליניאריות!), הרי שהמסנן הכולל ליניארי וניתן לביצוע כמסנן אחד. למשל, אם הסינון המקדים הוא 3 על 3 אחד, ורצוננו לבצע גזירה אופקית ממורכזת, המסנן הכולל יהיה:

$$\frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \otimes [1 \ 0 \ -1] = \frac{1}{9} \begin{bmatrix} 1 & 1 & 0 & -1 & -1 \\ 1 & 1 & 0 & -1 & -1 \\ 1 & 1 & 0 & -1 & -1 \end{bmatrix}$$

ואנו רואים כאן הפרש אופקי בין ממוצעי אזורים במקום בין פיקסלים בודדים - תוצאה די צפויה.

פעולות מרחביות מסתגלות

במסגרת הסעיף הקודם הכרנו פעולות סינון ליניאריות ושאינו ליניאריות, אך משותפת לכולן הייתה התכונה שהם קבועים במקום. באופן עקרוני ברור כי עם פעולות שמשתנות במקום ניתן להשיג תוצאות טובות יותר, אך הקביעה כיצד המסנן יהיה תלוי במקום היא שאלה קשה, וללא מענה מוצלח עליה, לא השגנו דבר. במסגרת סעיף זה נראה מספר מצומצם של מסננים תלויי מקום ונציג תוצאות אופייניות להן.

האלגוריתם הראשון אותו נציג קרוי מסנן Wallis. הבעיה בה מנסה מסנן זה לטפל היא תיקון תחום דינאמי וחיידוד התמונה. כבר ראינו שכאשר בתמונה מנוצלים מספר מצומצם של ערכי רמות אפור, פעולת מתיחת קונטרסט או יישור היסטוגרמה מצליחים להשיג שיפור. מה קורה אם בתמונה מנוצלות הרמות כולן, אך באופן שאינו מוצלח, כך שלמשל באזור מסוים ישנם רק גווני כהים ולא רואים פרטים, ובאזור אחר רק גווני בהירים ושוב לא רואים פרטים? במצב כזה כל הגווני מצויים, אך בפרטים קשה להבחין. הפתרון לבעיה זו הינו מתיחה לוקלית של רמות האפור - פעולה לוקלית במקום פעולה גלובלית.

מתיחת היסטוגרמה לוקלית יכולה להיעשות ע"י חישוב בכל פיקסל של היסטוגרמת הסביבה, ותיקון ערך הפיקסל על-פי היסטוגרמה זו, כך שאם רמת האפור של הפיקסל המרכזי מצויה בצד התחתון של ההיסטוגרמה הלוקלית, יש להקטין את רמת האפור עוד, ובכך להשיג חידוד מקומי, ובאופן דומה, אם היא בצד העליון יש להעלותה. הבעיה עם גישה פשטנית זו היא מורכבותה החישובית, שכן בניית היסטוגרמה סביב כל פיקסל היא פעולה כבדה.

כתחליף, מסנן Wallis מציע להניח כי היסטוגרמות לוקליות הן בקירוב גאוסיות, ולכן, במקום לבנות במלואן, נדרש רק חישוב של שני פרמטרים - ממוצע ושונות. נתחיל את תיאור האלגוריתם בתיאור דרך החישוב של הממוצע והשונות באופן מקומי בצורה נוחה. אם נניח כי התמונה הנתונה היא $x[m,n]$, חישוב הממוצע בכל פיקסל נעשה ע"י קונבולוציה עם מסנן בצורת גאוסין רחב אשר רוחבו σ קובע עד כמה הפעולה לוקלית (ערך מומלץ הוא $\sigma = 3-10$ פיקסלים). מסנן זה נבנה כמטריצה של 15 על 15 (בחירה שרירותית) ע"י המשוואה:

$$G_{\sigma}[m,n] = \frac{\exp\left\{\frac{(m-8)^2 + (n-8)^2}{2\sigma^2}\right\}}{\sum_{k=1}^{15} \sum_{j=1}^{15} \exp\left\{\frac{(k-8)^2 + (j-8)^2}{2\sigma^2}\right\}} \quad 1 \leq n, m \leq 15$$

לכן, תמונת הממוצע תהיה:

$$M_{In}[m,n] = x[m,n] \otimes G_{\sigma}[n,m]$$

קירוב השונות בכל פיקסל נעשית ע"י החסרת תמונת הממוצע מהתמונה המקורית, העלאת התוצאה בריבוע, ושוב קונבולוציה עם אותו מסנן גאوسی, כלומר:

$$S_{In}[m,n] = (x[m,n] - M_{In}[m,n])^2 \otimes G_{\sigma}[n,m]$$

ידוע כי בהינתן משתנה אקראי x בעל ממוצע M ושונות S , הערך $(x-M)/S$ הוא משתנה אקראי חדש בעל ממוצע אפס ושונות 1. באופן דומה, אם נתון משתנה אקראי x בעל ממוצע אפס ושונות 1, ורצוננו להמירו להיות בעל ממוצע M ושונות S , עלינו לבצע את הפעולה $Sx+M$. לכן, אם רצוננו להגיע בכל פיקסל לממוצע יעד M_0 (למשל - 128), ולשונות יעד S_0 (למשל 40), אזי עלינו לבצע את הפעולה הבאה:

$$x_{new}[m,n] = \frac{x[m,n] - M_{In}[m,n]}{S_{In}[m,n]} S_0 + M_0$$

כפי שרואים, תחילה נורמל המשתנה האקראי להיות בעל ממוצע 0 ושונות 1, ומשם הומר לערכי היעד. הפעולה הנ"ל היא פעולה מקומית, אך תלויה (בשל השימוש ב-תמונות M ו- S) בסביבת הפיקסל. מסנן זה אינו ליניארי ואינו חסר זכרון!

למעשה שיטה זו מניבה תוצאות קיצוניות מידי, כיוון שכפיית ממוצע ושונות יעד (S_0, M_0) זהים בכל מקום היא דרישה חזקה. ריסון לאלגוריתם זה מוצע ע"י קביעת תמונות לממוצע ולשונות היעד, כממוצע משוקלל של הערכים המקוריים וערכי היעד הקבועים. לכן:

$$\begin{aligned} M_{Dest}[m,n] &= (1 - \alpha)M_{In}[m,n] + \alpha M_0 \\ S_{Dest}[m,n] &= (1 - \alpha)S_{In}[m,n] + \alpha S_0 \\ \Rightarrow x_{new}[m,n] &= \frac{x[m,n] - M_{In}[m,n]}{S_{In}[m,n]} S_{Dest}[m,n] + M_{Dest}[m,n] \end{aligned}$$

בגישה זו, שימוש ב- $\alpha = 0$ לא משנה את התמונה, ושימוש ב- $\alpha = 1$ נותן את האלגוריתם הקיצוני. עבור ערכי ביניים נקבל אפקט נאה של חידוד ומתיחת קונטרסט לוקליים. הציור הבא מראה תמונת מקור ותוצאות העיבוד עם מסנן Wallis למגוון ערכי α .

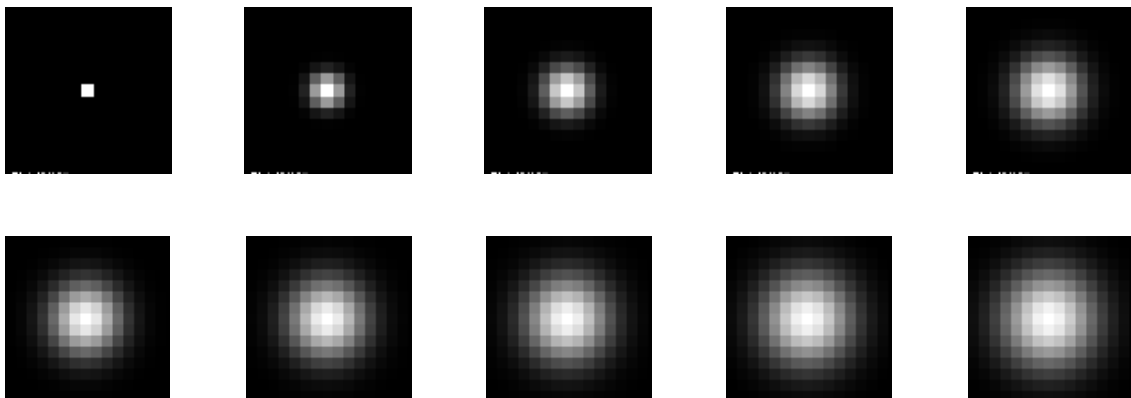


תוצאות סינון Wallis - שורה עליונה (שמאל) תמונת מקור, (ימין) תמונת הממוצע המקומי, שורה אמצעית (שמאל) תמונת השונות המקומית, (ימין) תוצאת הסינון עם $\alpha = 0.1$, שורה תחתונה (שמאל) תוצאת הסינון עם $\alpha = 0.4$, (ימין) תוצאת הסינון עם $\alpha = 1$.

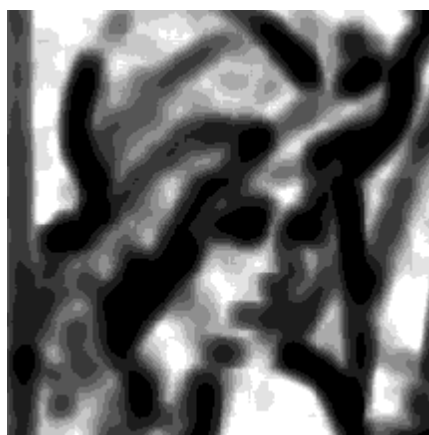
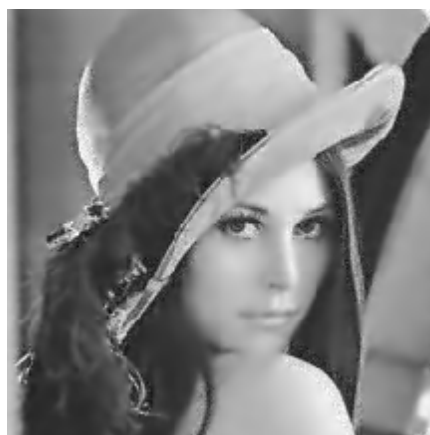
האלגוריתם הבא בו נדון נועד לנקות תמונה מרעש אדיטיבי. כבר ראינו כי החלקה (LPF) מצליחה יפה בניקוי רעש כאשר מדובר באזורים חלקים. לעומת זאת, באזורי שפות, החלקה גורמת לתופעה הצורמת לצופה האנושי. הרעיון אם כך הוא להפעיל מסנן גאוסי בעל שונות תלויה מקום - באזור חלק המסנן יהיה רחב, ובאזור שפה או מרקם המסנן יהיה צר. כך נקבל כי רעש יידחה יפה במקומות החלקים, וכמעט לא יידחה כלל באזורי שפה.

למעשה, במערכת הראיה האנושית קיימת תופעה הקרויה "תופעת המיסוך", אותה אנו מנצלים כאן. מסתבר כי קיומו של הרעש צורם לנו הרבה יותר באזורים חלקים, ושם אנו עושים עבודה טובה של דיכוי. לעומת זאת, באזורי שפות איננו כה רגישים לקיומו של רעש, ולכן ניתן להשאירו.

כיצד נקבע אם אזור הוא כולל שפה? אפשרות אחת היא לבצע גזירה כפי שתואר בסעיף קודם, ולהתייחס לערכים המוחלטים של הגרדיאנט. ברור כי גישה זו חייבת לבוא עם החלקה מקדימה של התמונה לשם מניעת השפעה רעה של הרעש המצוי בתמונה. גישה אחרת בא אנו נעשה שימוש היא להשתמש בתמונת שונות מקומיות, בדיוק כמו זו שנבנתה עבור מסנן ה-Wallis. נדרש לבנות סט של מסננים גאוסיים בעלי רוחב משתנה, ולבחור מסנן לפי ערך השונות (ביחס הפוך לה - שונות גבוהה - רוחב נמוך). הציור הבא מראה את המסננים הגאוסיים בהם בחרנו לעשות שימוש, והציור שאחריו מראה תוצאות של ניקוי רעש ע"י מסנן אדפטיבי כזה. בציור זה מתוארת גם תמונת בחירת המסננים (לפי אינדקסים בין 1 ל-10, כמספר המסננים הקיימים).



עשרת המסננים הגאוסיים בעלי הרוחב השונה

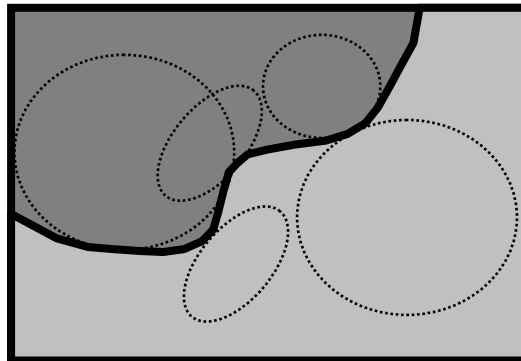


תוצאות סינון אדפטיבי עם גאוסיאן ברוחב משתנה. למעלה (שמאל) תמונת מקור, (ימין) תוצאת אלגוריתם הסינון, למטה (שמאל) תוצאת סינון קבוע במקום ברוחב ממוצע זהה, (ימין) תמונת האינדקסים של המסננים לפי המקום

מפעילים מסנן תלוי פעילות מקומית. כשאנו באזור שקט אנו פועלים נכון כשאנו מטשטשים ע"י מסנן רחב תמך, וקשה יהיה להתעלות על כך. באזורי שפה, לעומת זאת, אנו בוחרים לא לטשטש, וכתוצאה מכך נותר הרעש כבמקורו. כתחליף לכך, היה ראוי להפעיל באזור כזה מסנן בעל צורת שאינה עיגול אשר היה מכסה רק פיקסלים מהצד הנכון של השפה. הציור הבא ממחיש רעיון זה באופן גרפי. תמך המסנן נבחר להיות עיגול אם אין שפה באזור התמך. באשר מתקרבים לשפה, מתעוות המסנן ומנה להשיק לקו השפה, ולא לחצותו. לכן, ממוצע משוקלל המבוצע עם מסנן כזה מערב רק אוכלוסייה מצד אחד של השפה.

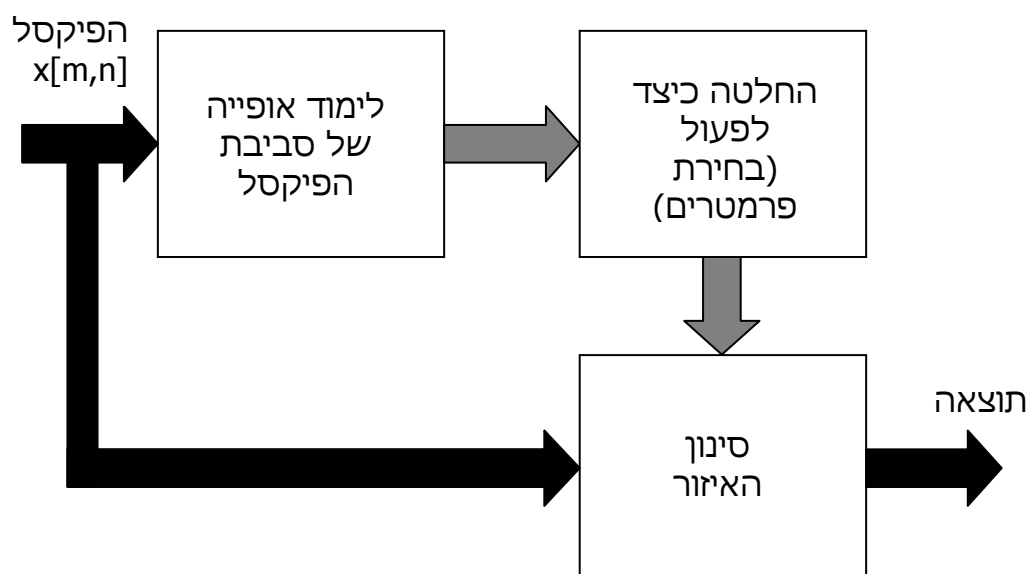
לרעיונות אלו של סינון כיווני עומק תיאורתי שלא יוצג כאן. מסתבר כי ניתן לבסס תכנון סינון שכזה ע"י כלים מחשבון וריאציות וגיאומטריה דיפרנציאלית, ולהראות כי התוצאה הינה נקודת מינימום של פונקציה על מידת

חלקות מסוימת של התמונה. אנו נכיר גישה זו באופן מפורט יותר במסגרת הפרק הבא, הדין בשחזור תמונה.



שימוש במסנן בעל חתך לא איזורופי לשם השגת דחיית רעש גם ליד שפות

ניתן להמשיך ולתאר עוד מגוון של מסננים אשר מתאימים עצמם לסביבת הפיקסל לשם קביעת אסטרטגית הטיפול בו, אך אנו נעצור את הדיון כאן. למרבית המסננים מהסוג שתואר כאן מבנה עקרוני כמיואר בציור הבא, בו נעשית בדיקה של אופי המידע בסביבת הפיקסל, עפ"י מידע זה נלקחת החלטה איזו פעולה לבצע, ולבסוף מבוצעת פעולה תלויה תוכן. חשוב להבהיר שאפילו אם המסנן שמושם בסוף הוא ליניארי (כלומר, צירוף משוקלל של רמות אפור), אין המסנן הכולל ליניארי, בשל שני הבלוקים האחרים בסכימה.



סכימה עקרונית של מסנן תלוי מקום המתאים עצמו לסביבה המטופלת

עיבוד תמונות בעזרת מחשב (236327)

הרצאה מס' 9

שחזור תמונות

בעיית השחזור הקלאסית

במסגרת פרק זה נדון בשחזור תמונות מקלקול מסוים - טשטוש ורעש. הבעיה חשובה ומתעוררת במערכות רבות. אנו נניח כי הן האופרטור המטשטש והן אופי הרעש האדיטיבי (חיבורי) ידועים לנו, ונבסס שיטות לבנייתן של התמונה המקורית אשר קדמה לקלקולים. זוהי בעיה קלאסית בעיבוד אותות בכלל, ובעיבוד תמונות בפרט, אשר זכתה לטיפול עשיר. למרות זאת, כשמסתכלים על התוצאות המתקבלות, נדמה כי עדיין לא נאמרה המילה בתחום.

תיאור נאות של תחום זה מחייב היכרות עם כלים מתורת השערוך (פרק בסטטיסטיקה והסתברות). אנו לא נעמיק ולא נציג את כלים אלו, ונשאיר את הדיון ברמה בסיסית יותר.

נתחיל בתיאור הבעיה, ונמחיש אותה דרך דוגמה. בידינו מצלמה דיגיטלית ורצוננו לצלם יעד כלשהו. לו היתה המצלמה שלנו אידיאלית (נטולת הפרעות), היינו מצלמים את התמונה $X[m,n]$ בגודל $L \times L$. לצערנו, בשל בחירתנו לעשות שימוש במצלמה מאיכות ירודה, אנו מקבלים תמונה אחרת $Y[m,n]$ בגודל $L \times L$ (ניתן להכניס לתוך הסיפור גם מצב של יצירת תמונה קטנה יותר). מהם הקלקולים שהכניסה המצלמה? ובכן, אנו נניח כי המצלמה עושה שימוש באופטיקה מאיכות ירודה (כלומר עדשות ירודות באיכותן) אשר גורמת לטשטוש H . אנו נניח כי טשטוש זה הוא ליניארי וקבוע במקום. לכן, גרעין הטשטוש $H[m,n]$ מספר את סיפורה המלא של פעולה זו.

ישנם מקרים בהם מקור הטשטוש אינה האופטיקה אלא תנועה של המצלמה. כאשר המצלמה זהה בזמן החשיפה, תתקבל מריחה כיוונית הקרויה

Motion Blur. מתוך התמונה המקולקלת ניתן לרוב לחשב את גרעין הטשטוש אשר הינו ליניארי וקבוע במקום.

נחזור לסיפור על המצלמה: בנוסף, מטריצת הגלאים בה המצלמה עושה שימוש סובלת מרעשים, ולכן התמונה לאחר הטשטוש מקבלת מרכיב אדיטיבי $N[m,n]$ של רעש אקראי בעל ממוצע אפס. אנו נניח לכל אורכו של פרק זה כי רעש זה הינו רעש לבן והומוגני בעל פילוג גאوسی. פירוש הדבר שבכל פיקסל מצטרף מספר אקראי גאوسی בעל שונות σ_N קבוע במקום, וממוצע אפס.

לכן, הקשר בין התמונה הרצויה $X[m,n]$ לתמונה המתקבלת $Y[m,n]$ נתון ע"י הביטוי:

$$Y[m,n] = X[m,n] \otimes H[m,n] + N[m,n]$$

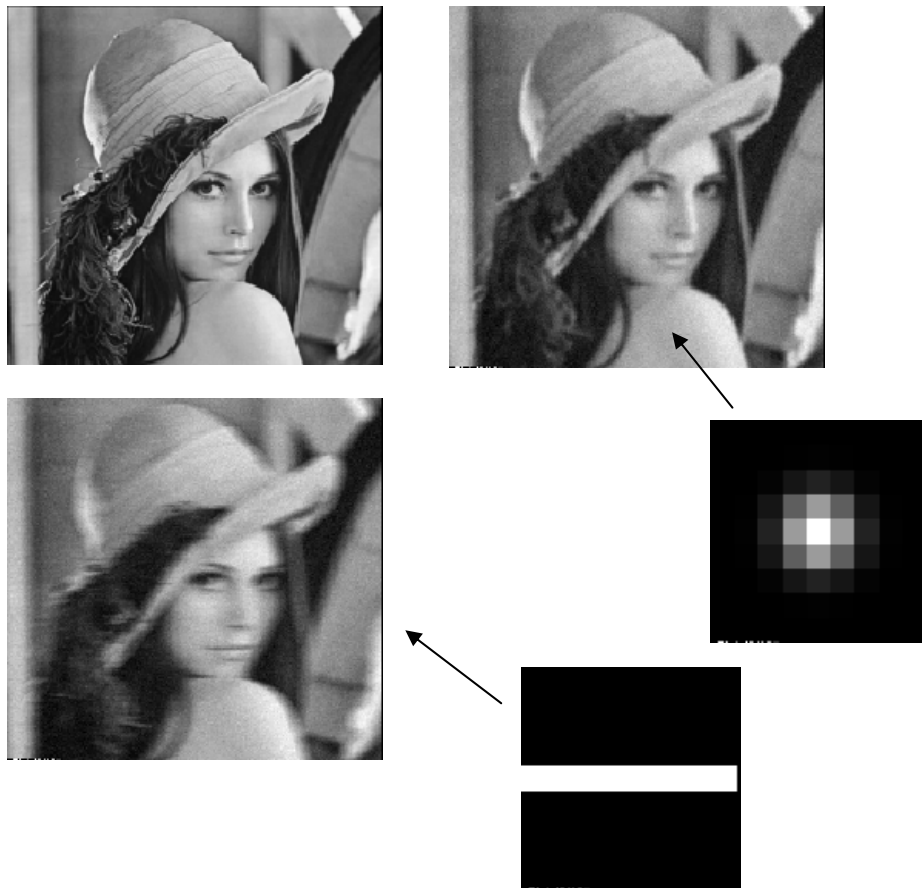
ניתן להמיר ייצוג זה לייצוג תדר ע"י שימוש בתכונת הקונבולוציה (הרי הטשטוש קבוע במקום), ולקבל:

$$Y_T[k,j] = H_T[k,j] \cdot X_T[k,j] + N_T[k,j]$$

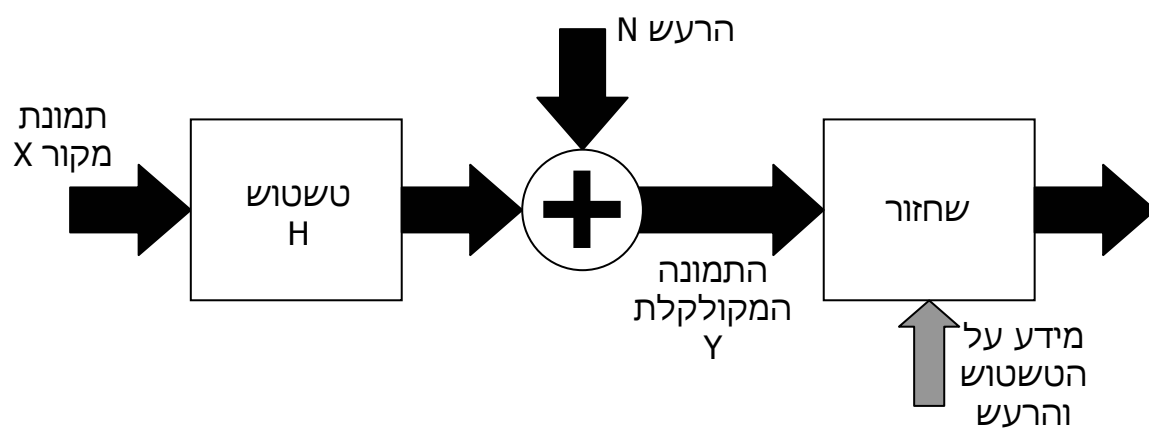
$$\begin{aligned} \text{where : } Y_T[k,j] &= F\{Y[m,n]\} \\ H_T[k,j] &= F\{H[m,n]\} \\ X_T[k,j] &= F\{X[m,n]\} \\ N_T[k,j] &= F\{N[m,n]\} \end{aligned}$$

היתרון בייצוג זה של הבעיה הוא היות הקשרים מכפלות של סקלרים אלו באלו, ולא פעולות מרחביות מורכבות.

הציור הבא מראה תמונת מקור X וגרסה מקולקלת שלה ע"י טשטוש אופייני לאופטיקה (טשטוש גאوسی בעל חתך רדיאלי), וגרסה של מריחה הנובעת מתנועה. בשני המקרים הוסף רעש אדיטיבי. כאמור, המטרה היא לקחת את התמונה המקולקלת, את המידע על גורמי הקלקול, ולשחזר את תמונת המקור. הציור שאחריו מראה סכימת בלוקים של התהליך במלואו (קלקול ותיקון). המידע המוזן לתהליך השחזור הוא התמונה המקולקלת עצמה, גרעין הטשטוש, ואופיו הסטטיסטי של הרעש (במקרה שלנו רק הפרמטר σ_N).



תמונת מקור וגרסאות מקולקלות שלה ע"י טשטוש גאוסי (למעלה בצד ימין) וטשטוש כתוצאה מתנועה (שמאל למטה). בשני המקרים הוסף רעש גאוסי לבן בעל שונות $\sigma_N = 5$.



סכימה המתארת את תהליך הקלקול והשחזור

שיטות שחזור בסיסיות

אם נתעלם לרגע מהרעש, כיוון שהקלקול הוא טשטוש H , מדוע לא ליישם אופרטור הפוך לו, H^{-1} , לשם השחזור? שימוש בייצוג בתדר מקל מאוד יישום זה ונותן

$$\frac{1}{H_T[k, j]} \cdot Y_T[k, j] = \frac{H_T[k, j]}{H_T[k, j]} X_T[k, j] + \frac{1}{H_T[k, j]} N_T[k, j]$$

ומיד רואים כי מתעוררות שתי בעיות: ראשית, כלל לא ברור כי לטשטוש קיים היפוך. למעשה, במרבית המקרים ניתן לקבוע בודאות כי לא קיים היפוך! אם אחד ממקדמי הפורייה של הטשטוש הינו אפס ($H_T(k, j) = 0$), פירוש הדבר הוא שתדר זה אשר יכול היה להיות חלק מתוכן X אופס ע"י הטשטוש ולכן אבד ולא ניתן יהיה לשחזרו.

תיקון לבעיה זו יהיה (ושוב, אנו "חושבים" בתדר) לקחת כל ערך כזה שהוא אפס, ולהמירו בערך אחר קרוב לאפס - נניח $1e-10$. גישה דומה אך מעט שונה היא להשתמש במסנן המשחזר הבא (הנתון בתדר):

$$G_T[k, j] = \frac{H_T^*[k, j]}{|H_T[k, j]|^2 + \text{Const}}$$

כאשר * מייצג צמוד קומפלקסי. כאשר הקבוע קטן משמעותית מ-1 (סדר גודל מומלץ - 0.001). לכן, בתדרים נמוכים בהם $H_T^*[k, j]$ בערכו המוחלט קרוב ל-1, הקבוע זניח ואנו מקבלים התנהגות של מסנן הפוך. עבור תדרים גבוהים בהם $H_T^*[k, j]$ מתקרב לאפס, הקבוע משמש כהגנה מחלוקה באפס.

עד כה בחרנו להתעלם מקיומו של הרעש. מסתבר כי אפילו אם הרעש חלש, השפעתו בתהליך של סינון הפוך היא דרמטית. נשתמש שוב בתיאור הבעיה ע"י ייצוג בתדר: נניח כי באיבר $[k, j]$ כלשהו ערך נמוך יחסית (נניח $1e-3$) ל- $H_T[k, j]$. נניח כמו כן כי בנקודת תדר זו, עוצמת האות המקורי ועוצמת הרעש נמוכות מאוד אך הרעש חזק יותר מהאות. תופעה זו מתרחשת כמעט תמיד בתדרים גבוהים בהם כמעט אין נוכחות לאות המקור ואז יוצא שהרעש דומיננטי. נניח לכן - $X_T[k, j] = 1e-3$ ואילו הרעש חזק יותר - $N_T[k, j] = 1e-2$. הפעלת המסנן ההפוך כבמקורו פירושה הכפלת שני ערכים אלו בערך $1e3$. גרמנו באופן זה להכנסה חזקה מאוד של רעש לתמונת התוצאה.

המסנן המשחזר שהוצע כשיפור יתנהג יותר טוב. במקרה כזה, נקבל שהן הרעש והן האות מועברים כשהם מוכפלים בגודל:

$$\frac{1 \times 10^{-3}}{1 \times 10^{-6} + \text{Const}} \underset{\substack{\uparrow \\ \text{Const}=0.001}}{\approx} 1$$

תוצאה זו לא נוראית כמו הכפלה ב-1000, אך אינה טובה לכשעצמה. בנקודת תדר זו היה ראוי שנכפול את הכניסה בערך נמוך מ-1 לשם דיכוי הרעש. ממילא האות טובל שם ברעש גדול ממנו, ולכן אין מה להפסיד. נציע אם כך את הגישה הבאה: בכל נקודת תדר נברר מהו היחס בין עוצמת האות לעוצמת הרעש, ולבצע מסנן הפוך רק בתדרים בהם יחס זה גבוה. במקומות בעלי יחס נמוך, בהם הרעש דומיננטי, כתחליף למסנן הפוך, יש להפעיל דיכוי (כלומר הכפלה בערך נמוך מאוד מ-1. דיכוי זה יקטין את האות שם, אך גם את הרעש שממילא דומיננטי יותר.

הבעיה עם גישה זו היא הצורך בשני גדלים שאת שניהם אין לנו - עוצמת האות המקורי ועוצמת הרעש בכל נקודת תדר. כתחליף, ניתן להניח כי עוצמת הרעש קבועה וזהה בכל נקודת תדר, ושווה σ_N . הנחה זו נכונה כאשר הרעש לבן. באשר לאות המקורי, הנחה סבירה היא שעוצמתו יחסית הפוך למרחק מהראשית בתדר. כלומר, עבור תדרים נמוכים עוצמתו צפויה להיות גבוהה, ועם העלייה בתדר עוצמתו יורדת. ביצוע התמרות פורייה למגוון גדול של תמונות מגלה כי כלל זה נכון ברוב מוחלט של המקרים. ההבדל בין התמונות מתבטא בדרך כלל כמעט רק בפאזה, ומזו אנו מתעלמים פה.

לסיכום, התהליך המוצע הוא ביצוע התמרת פורייה לאופרטור הטשטוש, ולקבל את $H_T[k, j]$. מסנן השחזור בתדר, $G_T[k, j]$ יתקבל ע"י:

$$G_T[k, j] = \frac{H_T^*[k, j]}{|H_T[k, j]|^2 + \alpha \sigma_N^2 (r[k, j])^2}$$

כאשר $r[k, j]$ יהיה המרחק של התדר $[k, j]$ אל הראשית במישור התדר (לדוגמה, $r[k, j] = \sqrt{k^2 + j^2}$. כך למשל, עבור $k=j=0$ זהו איבר ה-DC ולכן $r[0,0]=0$. גישה זו דומה להצעתנו הקודמת, למעט העובדה שכעת ה-"קבוע" שלנו גדל עם התדר, ולכן הופך דומיננטי מהר יותר.

מדוע ביטוי זה? בשל היות המכנה גדול מאפס לכל $[k,j]$, ברור כי לא נקבל היפוך של 0, וכך נפתרה הבעיה הראשונה. שנית, עבור $[k,j]$ נמוכים, כאשר האות חזק מאוד ביחס לרעש, הגודל $z[k,j]$ קטן מאוד, ונקבל כי המסנן הוא בקירוב טוב היפוך מדויק כיוון שהאיבר הימני במכנה יהיה זניח. כאשר הרעש דומיננטי מאוד, איבר זה (הימני במכנה) יהיה הדומיננטי במכנה ולכן המסנן המשחזר יהיה קרוב לאפס, כנדרש.

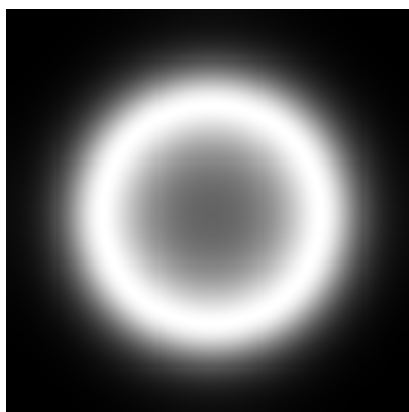
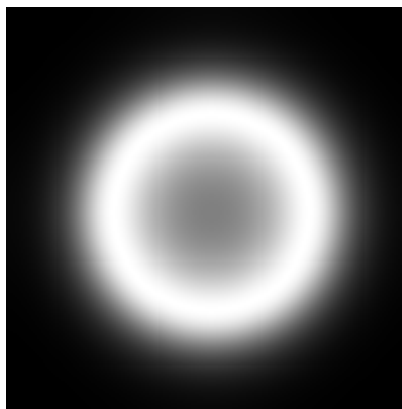
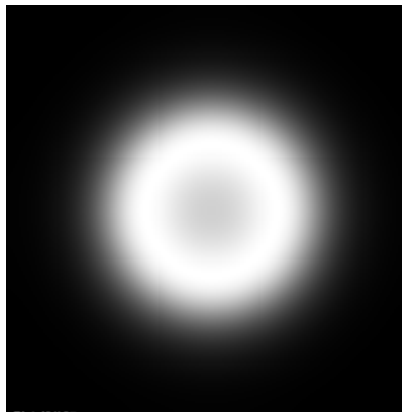
לאחר שהתקבל G_T ניתן לבצע את השחזור בתדר ע"י הכפלת G_T ב- Y_T וביצוע התמרת תדר הפוכה על תוצאה זו. המסנן המוצע מוכר מאוד וקרוי מסנן Wiener. בנייתו באופן מדויק וקבלת הביטוי שהוצע נעשית מתוך שיקולים הסתברותיים, שלא הצגנו כאן. הציור הבא מראה תמונת מקור מטושטשת ורועשת, ושחזור שלה ע"י מסנן היפוך. בעוד שהמרחק (במובן MSE) בין תמונת המקור לתמונה המקולקלת הוא 103, הרי שהמרחק בין תמונת המקור לשחזור כאן הוא $7e6$! התוצאה ללא כל ספק גרועה!



שחזור ע"י מסנן הפוך. תמונת מקור (שמאל למעלה), תמונה מקולקלת (ימין למעלה), תמונת המסנן בתדר (שמאל למטה), ותוצאת השחזור (ימין למטה).

בציור הבא אנו מראים לאותה תמונה בדיוק את תוצאת מסנן Wiener, עבור שלושה ערכי a שונים. עבור $a = 5e - 6$ מתקבלת שגיאה של 87.1, עבור $a = 1e - 6$ השגיאה היא 78.8, ועבור $a = 5e - 7$ השגיאה עולה ל- 94.8. אנו

רואים כי חל שיפור בתהליך השחזור, הן באופן ויזואלי, והן באופן כמותי. עם זאת, קשה להתנתק מתחושה של החמצה כשמסתכלים בתוצאות - השיחזור אינו מתקרב כלל לאיכות של תמונת המקור, ונשאלת השאלה האם זה הכי טוב שניתן להשיג? למזלנו, התשובה על כך היא לא! ניתן להציע פתרונות איכותיים יותר, אך הדבר מחייב שיטות מורכבות יותר, אשר יידונו בסעיף הבא בקצרה.



שחזור ע"י מסנן Wiener. שמאל - המסננים (בערך מוחלט), ימין - תוצאות השחזור. שלושת התוצאות מתייחסות לשלושה ערכים שונים של הפרמטר α .

דוגמה: נניח כי אנו עוסקים באותות חד ממדיים, הטשטוש הוא הפונקציה הבאה (בתדר):

$$H_T[k] = \frac{10}{k^2 + 10} \quad k = 0, 1, \dots, 100.$$

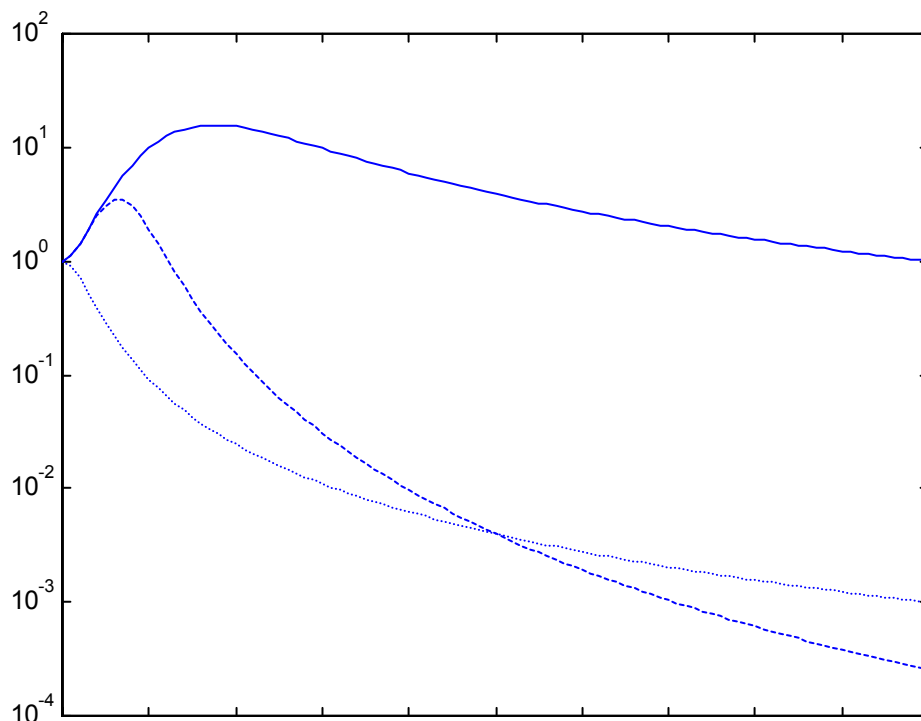
המסנן המשחזר שהוצע ראשון בסעיף זה יהיה:

$$G_T[k] = \frac{\frac{10}{k^2 + 10}}{\left[\frac{10}{k^2 + 10}\right]^2 + 0.001} = \frac{10(k^2 + 10)}{100 + 0.001(k^2 + 10)^2}$$

ואילו המסנן המשחזר השני יהיה (נבחר כאן קבוע באופן שרירותי):

$$G_T[k] = \frac{\frac{10}{k^2 + 10}}{\left[\frac{10}{k^2 + 10}\right]^2 + 0.001[0.4k^2]} = \frac{10(k^2 + 10)}{100 + 0.0004(k^2 + 10)^2 k^2}$$

בציור הבא אנו מראים את שלושת מסננים אלו (המסנן המטשטש, ושני המשחזרים).



המסנן המטשטש בתדר (קו מנוקד) ושחזורים המבוססים עליו. בקו מלא - מסנן היוריסטי ע"י הוספת קבוע בהיפוך, בקו מרוסק - מסנן דמוי Wiener המוסיף גורם הגדל בתדר.

אנו רואים כי המסנן הבנוי על הוספת קבוע נותן אמנם דעיכה בהגבר אחרי תדר מסוים, אך בכל מקרה הוא גבוה. מסנן ה-Wiener מצליח לרדת חזק יותר ואף לתת ניחות חזק יותר משל המסנן המטשטש - תוצאה רצויה כיוון שאנו מניחים שבאזור זה חזק הרעש מהאות. בשני המקרים, ההתנהגות באזור התדרים הנמוכים דומה - התנהגות של מסנן הפוך אשר אמורה לשחזר יפה מרכיבי תדר אלה.

גישות מתקדמות לשחזור תמונות – בסיס

במסגרת זו לא נראה באופן מלא את גישות מתקדמות לשחזור תמונה מטשטש ורעש, אך נציג את עקרונותיהן המנחים. אנו ננהל את כל הדיון הזה תחת המודל לבעיה כפי שתואר בתדר.

תחילה נשים לב לכך שאת המסנן ההפוך (שפועל רע – יש לזכור) יכולנו לקבל כמענה לשאלה הבאה – בהינתן תמונה מטשטשת Y עם טשטוש H , מהי התמונה X אשר אם נטשטש אותה נקבל תוצאה קרובה ככל האפשר למדידה Y ? ניסוח בעיה זו במונחים אלגבריים מוביל ל:

$$\text{Err}\{X_T[k, j]\} = \sum_{k=1}^L \sum_{j=1}^L (Y_T[k, j] - H_T[k, j] \cdot X_T[k, j])^2$$

בביטוי זה אנו לוקחים פתרון מועמד X , מטשטשים אותו ע"י H , ומבקשים שהתוצאה תהיה קרובה במידת האפשר ל- Y , כשהמידה לקירבה היא MSE. הפתרון כאמור טריוויאלי ומוביל למסנן ההפוך, כך שברור כי גישה זו לא מחדשת דבר ואינה מועילה. עם זאת, קורה פה דבר מעניין שיכול להתפתח למשהו חשוב – ניסחנו את בעיית השחזור כבעיית אופטימיזציה בה הנעלם היא התמונה X , ואנו רוצים להביא למינימום את השגיאה Err . בשל הייצוג בתדר, למרות ש- X כולל L^2 נעלמים, בעיית האופטימיזציה ניתנת להפרדה ל- L^2 בעיות סקלריות קלות לטיפול.

ניקח את השגיאה המוצעת Err ונוסיף לה מרכיב נוסף: מבין כל הפתרונות שיוצעו, נרצה להעדיף את אלו שנראים כתמונות. למשל, ראינו כי תוצאת המסנן ההפוך היא תמונה בה ערכי רמות האפור ב- X הם עצומים. לכן,

נוכל להציע בעיית אופטימיזציה חדשה בה נפעיל כוח למנוע ערכים גבוהים כאלה,

$$\text{Err}\{X_T[k, j]\} = \sum_{k=1}^L \sum_{j=1}^L (Y_T[k, j] - H_T[k, j] \cdot X_T[k, j])^2 + \lambda \sum_{k=1}^L \sum_{j=1}^L (X_T[k, j])^2$$

אם מוצע פתרון עם X_T גבוה, הוא ישלם מחיר בפונקצית השגיאה ובחירתו תימנע. גם במבנה המוצע הבעיה ניתנת לפירוק להרבה בעיות סקלריות מהמבנה

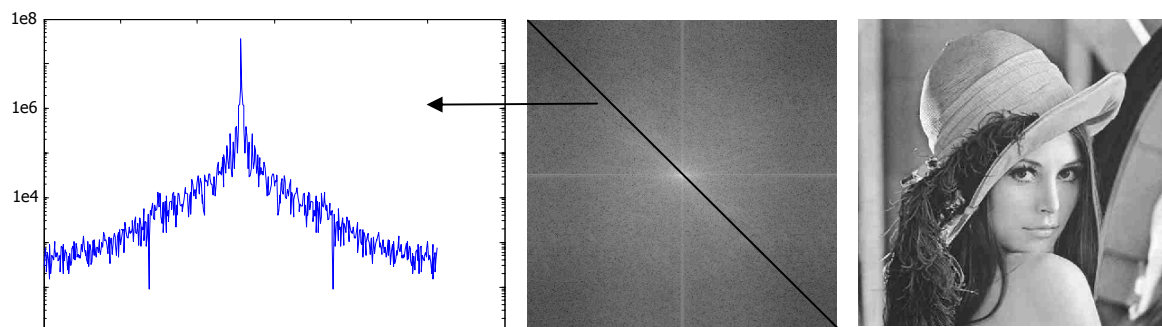
$$f(x) = (y - hx)^2 + \lambda x^2$$

פירוק זה נובע ישירות מהשימוש בתיאור בתדר. עבור תיאור דומה במקום לא היה מתקבל ייצוג כל-כך נוח של הבעיה. הפתרון המתקבל הוא

$$\frac{df(x)}{dx} = 2h^*(hx - y) + 2\lambda x = 0 \Rightarrow \hat{x} = \frac{h^*}{h^*h + \lambda} h$$

וזו הרי תוצאה שכבר הצענו על בסיס אינטואיטיבי (וגם ראינו שהיא לא מוצלחת).

גישה אחרת לעניין הכוחות שיש להפעיל על הפתרון הרצוי קשורה לתכונה גלובלית של תמונות להיות בעלי התמרת פורייה הדועכת עם העלייה בתדר (בדו-מימד הכוונה לעליה במרחק מהראשית בתדר). קל לאושש טענה זו – קחו תמונות רבות ועיינו בהתמרת הפורייה שלהן ותגלו כי תכונה זו נכונה. הציור הבא ממחיש זאת – שימו לב כי ההתמרה מוצגת בסקלה לוגריתמית כי הדעיכה מהירה מידי.



תמונה אופיינית, התמרתה בתדר (עם פעולת log על הערכים) וחתך אופייני

לכן, נוכל להציע פונקצית מחיר אלטרנטיבית בה מקדמים בתדרים נמוכים יורשו לגדול בעוד שעל אלה שבתדרים גבוהים נכפה משקל חזק. נקבל

$$\text{Err}\{X_T[k, j]\} = \sum_{k=1}^L \sum_{j=1}^L (Y_T[k, j] - H_T[k, j] \cdot X_T[k, j])^2 + \lambda \sum_{k=1}^L \sum_{j=1}^L (k^2 + j^2)(X_T[k, j])^2$$

גם במבנה זה הבעיה ניתנת לפירוק להרבה בעיות סקלריות מהמבנה

$$f(x) = (y - hx)^2 + \lambda Cx^2$$

כאשר $C = k^2 + j^2$ משתנה מנקודת תדר אחת לאחרת. הפתרון המתקבל הוא

$$\hat{x} = \frac{h^*}{h^*h + \lambda C} h$$

וכשנציג זאת לכל הפתרונות יחדיו נקבל:

$$\hat{X}_T[k, j] = G_T[k, j]Y_T[k, j] = \frac{H_T^*[k, j]}{|H_T[k, j]|^2 + \lambda(k^2 + j^2)} Y_T[k, j]$$

וגם תוצאה זו מוכרת לנו כתוצאת מסנן Wiener, ואנו ראינו כי אמנם תוצאותיו טובות מסינון היפוך, אך גרועות לכשעצמן.

מסנן Wiener אדפטיבי לשחזור תמונות

ברוח הגישה שתוארה לעיל, נוכל להציע שיטות משופרות לשחזור תמונה. כבר ראינו כי מסנן Wiener אינו מתפקד יפה בדרך כלל. ראינו שעבור בחירת λ גדול ישנה נטייה לטשטוש יתר של התמונה מתוך רצון להילחם ברעש, בעוד שעבור בחירת λ קטן ישנה נטייה לחידוד יתר ולצלצולים – תופעת התנדנדות פרועה – מתוך רצון להילחם בטשטוש. ישנו ערך λ מיטבי הנותן ביצועים הכי טובים, אך אלו לכשעצמם אינם מלבבים.

גישה אחת להתגבר על "תקרת הזכוכית" של ביצועי מסנן Wiener היא להפוך אותו לאדפטיבי למקום. ככלל, קשה ליישם מסנן תלוי במקום בתדר, כיוון שתכונת הקונבולוציה רלוונטית רק למסננים קבועים במקום. עם זאת, נראה פה גישה פשוטה ויעילה שתיתן אדפטיביות ותפעל בתדר. הרעיון פשוט – ניישם על התמונה המטושטשת שתי (או יותר) פעולות סינון – האחת עם λ קטן אשר תיתן אפקט חידוד חזק אך תביא להתנדנדויות באזורים חלקים, והאחרת עם λ גדול שתמרח המוצא ולא תחדד, אך תביא למחיקת רעש יעילה באזורים חלקים בתמונה. כלומר, עבור $\lambda_1 < \lambda_{\text{opt}} < \lambda_2$, שתי תמונות התוצאה תהיינה

$$\hat{X}_T^1[k, j] = \frac{H_T^*[k, j]}{|H_T[k, j]|^2 + \lambda_1(k^2 + j^2)} Y_T[k, j]$$

$$\hat{X}_T^2[k, j] = \frac{H_T^*[k, j]}{|H_T[k, j]|^2 + \lambda_2(k^2 + j^2)} Y_T[k, j]$$

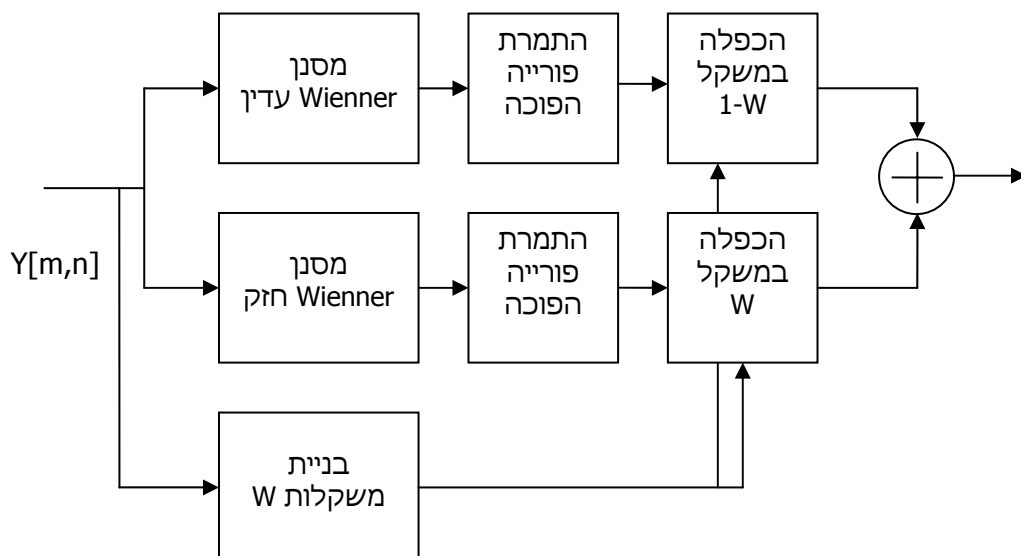
נניח כעת כי נוכל לבחון את התמונה Y ולקבוע את האזורים בהם היא חלקה – נעשה זאת פשוט ע"י פעולת גזירה והשוואה לסף של הערך המוחלט, כפי שעשינו לשם גילוי שפות. נקבע מפת משקלות $W[m, n]$ אשר תהיה 1 לאזורים חלקים מאוד, תרד בערכה כשהאזור אינו חלק דיו, ותגיע ל- 0 עבור אזורים עתירי פעילות מפה זו תוכל להיות מחושבת, למשל, ע"י

$$W[m, n] = \frac{1}{1 + \beta |\nabla X[m, n]|^2}$$

באזורים עם גרדיאנט נמוך המשקל יהיה קרוב ל-1 כרצוי, ועם עליית הגרדיאנט המשקל דועך ל-0. נוכל להציע את הפתרון המשולב הבא כתמונה המנצלת את הטוב שבין שני העולמות:

$$\hat{X}_{Final}[m, n] = F^{-1} \{ \hat{X}_T^1[k, j] (1 - W[m, n]) + F^{-1} \{ \hat{X}_T^2[k, j] W[m, n] \}$$

בצורה זו, באזורים בהם W קרוב ל-1 (בהם התמונה חלקה) נבחר את התמונה השנייה בה המסנן פעל חזק. עם התגברות הפרטים בתמונה נטה את המוצא לבחור את התמונה הראשונה בה המסנן עדין יותר.



תיאור האלגוריתם שייתן הפעלת מסנן Wiener אדפטיבי

גישה רובסטית לשחזור תמונות

בבניית מסנן Wiener למעשה הצענו פונקצית מחיר אותה יש להביא למינימום, ופתרונה היה לתמונה הרצויה. עם זאת, הגבלנו את בחירת המחיר השני – התנהגות התמונה הנעלמת – לביטוי ריבועי, בו האופטימיזציה קלה. נראה עתה גישה שתיראה בתחילה כקשה לפתרון, אך תהיה למעשה די פשוטה. בגישה זו נשתמש במדדים שאינם ריבועיים. נציע את פונקציית המחיר הבאה:

$$Err\{X_T[k, j]\} = \sum_{k=1}^L \sum_{j=1}^L (Y_T[k, j] - H_T[k, j] \cdot X_T[k, j])^2 + \lambda \sum_{k=1}^L \sum_{j=1}^L |X_T[k, j]|^p$$

אנו נניח כי ערכי p הינם בתחום $[0, 2]$. עבור $p=0$ ביטוי זה סופר את כמות המקדמים השונים מאפס, ולכן בניסוח הבעיה תהיה נטייה להציע פתרון בעל מספר מקדמי פורייה מעטים ככל האפשר (כמותם תיקבע בפועל על-ידי λ). גישה זו טובה עבור מקרים בהם ידוע לנו כי התמונה הרצויה בנויה ממספר מצומצם של הרמוניות.

עבור $p=1$ אנו נבקש מהפתרון להיות בעל סכום ערכים מוחלטים של המקדמים קטן. זהו מן ריכוך של גישת הדלילות שעלתה קודם, אך גם כאן הרוח השורה היא על מיעוט מקדמי פורייה. לעומת זאת, עבור $p=2$ אנו חוזרים למקרה בו כבר טיפלנו, וכאן דווקא ישנה העדפה להרבה מקדמי פורייה קטנים בעוצמתם.

גם כאן ניתן לנסח את הבעיה הנ"ל כאוסף בעיות פרידות, ובשל כך הפתרון לא יהיה קשה. הבעיה הסקלרית היא

$$f(x) = (y - hx)^2 + \lambda |x|^p$$

הפתרון המתקבל הוא

$$\frac{df(x)}{dx} = 2h^*(hx - y) + p\lambda \text{sign}(x) \cdot |x|^{p-1} = 0$$

נשים לב לכך שהפעם לא יהיה קל כמו קודם לקבל ביטוי סגור לפתרון.

נתחיל בניתוח של המקרה $p=0$. אנו נמצא הפתרון ללא שימוש בנגזרת המוצעת. במקרה זה עלינו להבחין בין שתי אפשרויות – $x=0$ או $x \neq 0$. נציב את שני מקרים אלו ונקבל את המחירים

$$\begin{aligned}
 x = 0 &\Rightarrow f(x) = y^2 \\
 f(x) = (y - hx)^2 + \lambda |x|^0 &\Rightarrow \\
 x \neq 0 &\Rightarrow f(x) = \lambda
 \end{aligned}$$

בביטויים הנ"ל הנחנו כי אם $x \neq 0$ אזי מן הראוי שיהיה $x = y/h$ (ונניח כי h אינו אפס, כיוון שאם הוא אפס, רצוי שגם x יהיה אפס למינימיזציה של פונקציה זו). לכן, הפתרון במקרה זה פשוט:

$$\hat{x} = \begin{cases} y/h & y > \sqrt{\lambda} \\ 0 & \text{אחרת} \end{cases}$$

עבור $p=1$ הנגזרת ניתנת לרישום כ - $2h^*(hx - y) + \lambda \text{sign}(x) = 0$. אם נניח כי h ו- y ממשיים וחיוביים, אזי ברור כי גם x אי-שלילי. אם הוא אפס, הנגזרת שלילית ושווה $2h^*y > 0$. אם הוא חיובי ממש, הנגזרת תהיה

$$2h^*(hx - y)^2 + \lambda = 0$$

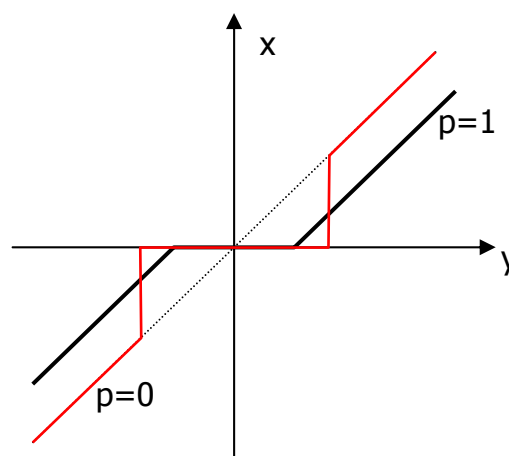
וזו תתאפס עבור

$$\hat{x} = \frac{1}{h} \left(y - \frac{\lambda}{2h^*} \right) = \frac{y}{h} - \frac{\lambda}{2|h|^2}$$

אלא שפתרון זה חייב להיות חיובי - דבר המציב דרישה על הקשר בין y ל- λ . לכן, לסיכום נקבל כי הפתרון במקרה של $p=1$ הינו

$$\hat{x} = \begin{cases} \frac{y}{h} - \frac{\lambda}{2|h|^2} & |2h^*y| > \lambda \\ 0 & \text{אחרת} \end{cases}$$

בציור פתרונות אלו נראים כך:



שחזור רובסטי

עיבוד תמונות בעזרת מחשב (236327)

הרצאה מס' 10

מבוא לדחיסת תמונות

יסודות בתורת האינפורמציה

תורת האינפורמציה הינה תורה חדשה יחסית אשר פותחה לראשונה ע"י Shannon בשנות הארבעים של מאה זו. בבסיסה, תורה זו באה לכמת מידע, ולאפיין מערכות בהקשר של העברת מידע. תורה זו מהווה את התשתית התיאורטית לנושא דחיסת מידע בכלל ודחיסת תמונות בפרט. ברור כי במסגרת זו לא ניתן יותר מסקירה ראשונית לא ממצה בשל נפחו העצום של תחום זה וריבוי התוצאות בו.

נתחיל עם מושג האנטרופיה - מידת אי-ודאותו של מקור מידע. נתון לנו מקור המייצר מספרים אקראיים דיסקרטיים x חסרי תלות ביניהם (i.i.d) עם חוק פילוג $P(x)$. האנטרופיה של מקור זה מוגדרת ע"י הביטוי:

$$H\{x\} = \sum_x P(x) \log_2 \frac{1}{P(x)} \left[\frac{\text{bits}}{\text{symbol}} \right]$$

משמעות גודל זה היא כמות הסיביות התיאורטית הנחוצה לייצוג מדויק של תו ממקור זה בממוצע. נראה מספר דוגמאות להמחשת מידה זו.

דוגמה: מקור שמייצר מספרים שלמים בתחום $[1,16]$ בהסתברות זהה - $1/16$ יהיה בעל אנטרופיה

$$H(x) = 16 \cdot \frac{1}{16} \log_2 16 = 4 \text{ bits / symbol}$$

תוצאה זו אינטואיטיבית לנו כיוון שעבור 16 אפשרויות, ניתן לייצג כל אחת ב- 4 סיביות, כפי שהתקבל בפועל.

דוגמה: עבור אותו מקור כשהפעם ההסתברות נתונה כ: $P(x) = x/136$ (המספר 136 מנרמל את ההסתברות לקבלת סכום הסתברויות שווה 1), האנטרופיה תהיה:

$$H(x) = \frac{1}{136} \left[1 \cdot \log_2 \frac{136}{1} + 2 \cdot \log_2 \frac{136}{2} + \dots + 16 \cdot \log_2 \frac{136}{16} \right] \cong 3.76 \text{ bits / symbol}$$

תוצאה זו מעניינת כיוון שעדיין יש לנו 16 אפשרויות, אך לטענת התיאוריה, די ב- 3.76 סיביות לתיאורן באופן מלא. ניתן למצוא בדבר הגיון אם נחשוב על הדברים באופן הבא - כאשר כל ההסתברויות זהות, אי הודאות היא מירבית, כלומר, אין לנו כל דרך לדעת משהו בטורם קיבלנו את התו. עבור הסתברות לא אחידה, עוד בטורם נקבל את התו נוכל לומר כי ישנה נטייה חזקה לכיוון הערכים הגדולים, בהיותם בעלי הסתברות גבוהה יותר. לכן, מידת אי-הודאות המתקבלת נמוכה יותר.

דוגמה: נניח כי ל- 16 התווים האפשריים, הסתברותו של 1 היא 1, ושל כל היתר 0. במקרה כזה האנטרופיה תהיה:

$$H(x) = 1 \cdot \log_2 1 + 15 \cdot 0 \log_2 0 = 0 \text{ bits / symbol}$$

אין כאן כל הפתעה: במקרה זה, לפני שנקבל את התו נדע בדיוק מה יקרה, ולכן אין כל אי-ודאות.

- למעשה, דרך דוגמה זו המחשנו שתי תכונות חשובות של פונקציית האנטרופיה:
1. למקור בעל N תווים, האנטרופיה H חסומה מלמעלה לפי - $H \leq \log_2 N$.
 2. למקור בעל N תווים, האנטרופיה H חסומה מלמטה לפי - $H \geq 0$. שיויון מתקיים כאשר תו אחד בעל הסתברות 1, וכל יתר התווים בעלי הסתברות 0.

משפט הצפינה של Shannon קובע כי עבור מקור בעל אנטרופיה $H(x)$, לא ניתן לדחוס מידע ממנו בפחות מהאנטרופיה ללא הכנסת שגיאות. יתרה מזו: לכל ε חיובי וקטן כרצוננו, קיים **בעקרון** אלגוריתם דחיסה אשר ייתן דחיסת המקור לקצב של $H(x) + \varepsilon$. תוצאה זו קובעת כי האנטרופיה מהווה אומדן לביצועיו של אלגוריתם דחיסה, אם רצוננו לדחוס ללא שגיאות. משפט זה של Shannon לא מציע את אלגוריתם הדחיסה עצמו, אלא מוכיח את קיומו באופן תיאורטי. בהמשך נכיר דרכים מעשיות למימוש טענה זו.

חשוב להבהיר כי מושג האנטרופיה אינו מוגבל לטיפול בתווים בודדים. לדוגמה, עבור מקור המייצר מספרים בתחום $[1,16]$, ניתן להתייחס לצמדי תווים כ-"סופר-תווים" חדשים, ואז יש לנו 256 סוגי תווים אפשריים, וניתן לכל צירוף לקבוע את הסתברותו, לחשב אנטרופיה כמקודם וכו'. תהליך זה חשוב ונחץ כאשר יש קשר סטטיסטי בין התווים. נמחיש זאת דרך דוגמה.

דוגמה: עבור המקור שתואר בדוגמה הראשונה, לקיחת צמדי תווים תביא לכך שיהיו 256 תווים אפשריים שווי הסתברות. האנטרופיה תהיה:

$$H([x_1, x_2]) = 256 \cdot \frac{1}{256} \log_2 256 = 8 \text{ bits / symbol}$$

מכור כי הפעם תו אחד הוא בעצם שני תווים, ולכן אין כל שינוי ביחס לטיפול בתווים בודדים.

דוגמה: נתון מקור המייצר את המספרים $1+$ ו- $1-$ לפי חוק הפילוג הבא:

$$P(x_k / x_{k-1}) = \begin{cases} 0.8 & x_k = x_{k-1} \\ 0.2 & x_k \neq x_{k-1} \end{cases}$$

כלומר, לדוגמה, אם התו הקודם היה $1-$, סבירות גבוהה יש להיותו של התו הבא $1-$ גם כן. בטיפול בתווים בודדים נקבל כי $P(-1) = P(+1) = 0.5$ כיוון ש:

$$\begin{aligned} P(+1) &= P(+1 / +1)P(+1) + P(+1 / -1)P(-1) \\ P(-1) &= P(-1 / +1)P(+1) + P(-1 / -1)P(-1) \\ \Rightarrow \begin{bmatrix} P(+1 / +1) - 1 & P(+1 / -1) \\ P(-1 / +1) & P(-1 / -1) - 1 \end{bmatrix} \begin{bmatrix} P(+1) \\ P(-1) \end{bmatrix} &= \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ \Rightarrow P(+1) = P(-1) &= 0.5 \end{aligned}$$

לכן האנטרופיה תהיה:

$$H(x) = 0.5 \log_2 2 + 0.5 \log_2 2 = 1 \text{ bits / symbol}$$

עבור טיפול בצמדי תווים נקבל כי חוק ההסתברות הוא $P([+1, +1]) = P([-1, -1]) = 0$ ו- $P([-1, +1]) = P(+1, -1) = 0.1$, כיוון ש:

$$P([x_k, x_{k-1}]) = P(x_k / x_{k-1})P(x_{k-1}) = \begin{cases} 0.8 \cdot 0.5 & [+1, +1] \\ 0.2 \cdot 0.5 & [+1, -1] \\ 0.2 \cdot 0.5 & [-1, +1] \\ 0.8 \cdot 0.5 & [-1, -1] \end{cases}$$

במקרה זה האנטרופיה תהיה:

$$H([x_1, x_2]) = 2 \cdot 0.4 \log_2 2.5 + 2 \cdot 0.1 \log_2 10 = 1.7219 \text{ bits / symbol}$$

כך שלתו יחיד נדרשים לפחות מסיבית אחת, וברור כי הושגה יעילות גבוהה יותר כי נוצל הקשר בין התווים. למעשה, ניתן להמשיך במגמה זו ולטפל בשלוש, רביעיות וכו', כשצפוי שיפור ככל שעובדים עם בלוקים גדולים יותר.

למעשה, יכולנו לנצל את הקשר הסטטיסטי בין התווים הסמוכים בצורה יעילה יותר לו פעלנו על ההסתברויות המותנות בצורה ישירה. נניח כי אנו נמצאים בתו ה-L וכבר קיבלנו אותו במפענח, ונניח כי הוא היה +1. לכן, בבואנו לדחוס (ובמקלט - לפרוס) את התו הבא, אנו יודעים כי בסבירות גבוהה הערך הבא הוא +1 גם כן. למעשה, האנטרופיה של התו הבא תהיה:

$$H(X_{L+1}) = 0.8 \log_2 \frac{1}{0.8} + 0.2 \log_2 \frac{1}{0.2} = 0.7219$$

כך שקיבלנו תוצאה טובה יותר מזו של העבודה בצמדים. גישה זו אינה גישה עבודה בבלוקים אלא גישה חלופית של חיזוי - על פי התו הנוכחי שכבר שודר (ונקלט), נדחס התו הבא, כשעושים שימוש בהסתברות המותנית.

למעשה, לו קידדנו את המקור הזה בבלוקים שאורכם שואף לאינסוף היינו מקבלים תוצאה זו בדיוק. כאשר אנו עובדים בצמד תווים אנו מנצלים את הקשר ביניהם, אך לא את הקשר לצמד הקודם או לצמד הבא, ולכן התוצאה פחות טובה. כאשר הבלוקים מתארכים אנו מצליחים טוב יותר לנצל את הקשרים בין התווים, כשאנו מאבדים רק את הקשרים בתפרים בין הבלוקים. אם הבלוקים באורך אינסופי - אין אובדן. ברור כי הגישה המותנית יעילה יותר לאין-ערור.

נציג עוד מספר תכונות המאפיינות את פונקציית האנטרופיה, הנוגעות לטיפול בבלוקים:

$$1. \text{ קיים אי-השוויון הבא: } H([x_1, x_2, \dots, x_N]) \leq \sum_{k=1}^N H(x_k)$$

כשוויון מתקיים עבור חוק פילוג פריד:

$$P([x_1, x_2, \dots, x_N]) = P(x_1)P(x_2) \dots P(x_N)$$

משמעות תכונה זו היא שתמיד יש רווח או לפחות אי-הפסד בעבודה על בלוקים במקום על תווים בודדים.

$$2. \text{ קיים אי-השוויון הבא: } H([x_1, x_2, \dots, x_N]) \geq \max_k \{H(x_k)\}$$

כששוויון מתקיים עבור מקרה בו כל המשתנים נקבעים באופן דטרמיניסטי מתוך אחד מהם. פירוש הדבר הוא שכאשר הינו בלוק של תווים אך למעשה רק אחד מהם אקראי וכל היתר נגזרים ממנו בצורה מדויקת, די לקודד תו אקראי זה. במקרה כללי, אם נחשב אנטרופיה לכל אחד מהתווים בנפרד, לא יתכן שהאנטרופיה המשותפת תהיה קטנה מהגרוע שבהם (גרוע במובן של מכיל הרבה אינפורמציה).

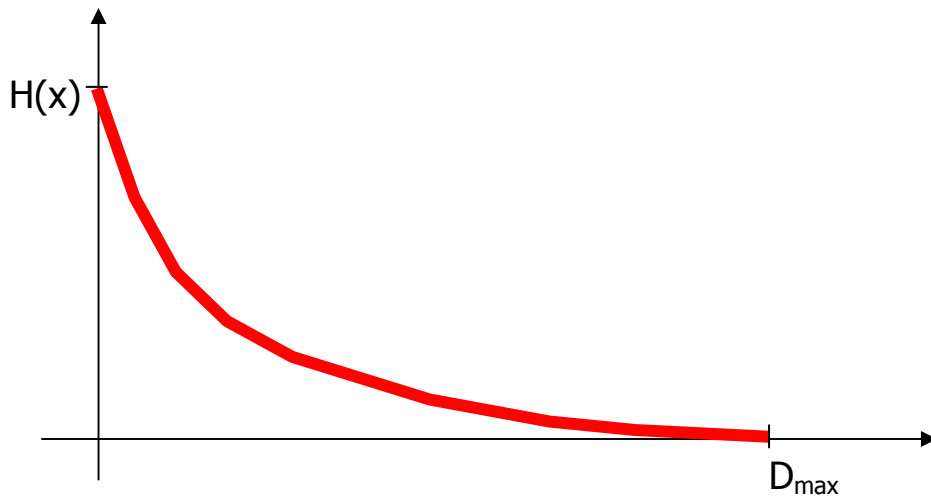
$$3. \text{ קיים אי-השוויון הבא: } H([x_1, x_2]) \geq 2 \cdot H(x_1/x_2)$$

כששוויון מתקיים עבור מקרה בו אין קשר סטטיסטי בין התווים.

עד כה עסקנו בהגדרת מידת האינפורמציה של מקור כאשר רצוננו לדווחסו ללא שגיאות. מה קורה כאשר אנו מוכנים להתיר שגיאות? מסתבר שגם ליישום זה הציע Shannon אומדן אינפורמציה, הידוע בכינויו פונקצית קצב-העיוות $R(D)$ - Rate-Distortion Function. הרעיון מכליל בצורה ישירה את מושג האנטרופיה ממקודם באופן הבא. נניח כי אנו מוכנים לקבל כי ייווצר מרחק של עד D בין סידרת המקור x והסדרה הנפרסת לאחר דחיסה y . פער זה מסומן ב- $d\{x, y\} \leq D$ וברור כי עלינו להגדיר מהי המטריקה בה אנו עושים שימוש. אזי, מידת האנטרופיה של המקור x תחת שגיאה של עד D תהיה $R(D)$. נציג מספר תכונות של פונקציה זו:

1. עבור $D=0$ אנו בפועל לא מרשים שגיאה ולכן ברור כי $R(0) = H(x)$.
2. עבור D מעל סף מסוים התלוי בעוצמתו של x יתקבל כי אם פשוט נשלח את ממוצע המספרים לכל תו כניסה תתקבל שגיאה קטנה מ- D . במקרה זה נקבל כי $R(D) = 0$.
3. פונקציה ה- $R(D)$ מונוטונית לא עולה. עליה לא תיתכן כיוון שאז לשגיאה גדולה יותר נדמה כי נדרש קצב גבוה יותר - זה לא יתכן.
4. פונקציה ה- $R(D)$ היא פונקציה קמורה.

הציור הבא מראה כיצד נראית פונקצית ה- $R(D)$ באופן עקרוני.



מבנה עקרוני של פונקצית ה- $R(D)$

חשוב להבהיר כי בעוד שאנטרופיה של מקור קלה לחישוב, חישובה של $R(D)$ (באופן אנליטי או נומרי) קשה מאוד, ולרוב בלתי אפשרי. חישובה של האנטרופיה דורש את חוק פילוגם של תווי המקור, חישוב ה- $R(D)$ דורש בנוסף את הגדרת המטריקה בין אותות הכניסה והיציאה ממערכת הדחיסה\פריסה.

דוגמה: נניח כי המקור שלנו מייצר מספרים שלמים בתחום $[0,127]$ בהסתברות אחידה. לכן האנטרופיה של מקור זה היא 7 סיביות לאות. כעת נניח כי אנו מתירים לקידוד לבצע סטייה של 1, כלומר,

$$D(x, y) = (x - y)^2 \leq 1$$

במילים אחרות, אנו מתירים לסטות ברמה אחת במקלט בהשוואה למשדר. נציע הצעה פשוטה שתגשים שגיאה זו תוך שידור פחות סיביות - ניקח את המספר ונחליף אותו במספר הזוגי הקרוב אליו מלמטה. כך יקרה כי מספרים זוגיים ייוותרו כמות שהם, ומספרים אי-זוגיים יומרו למספרים זוגיים. כיוון שכעת קיימים רק מספרים זוגיים, נדרשים 6 סיביות במקום 7. בתהליך דומה, עבור שגיאה של 2 נוכל לבחור את השכן הקרוב ביותר מבין המספרים המתחלקים ב-3 (0,3,6,9, ..., 126). כעת יש לנו רק 43 ערכים אפשריים שווי הסתברות ולכן האנטרופיה היא 5.246.

נוכל להמשיך תהליך זה ולהציע מקודד עם שגיאה למספרים אלו, ובעצם לקבל פונקציה המקרבת את ה- $R(D)$. הפונקציה המוצעת היא פונקציה החוסמת מלמעלה את פונקציית קצב-העיוות כיוון שאין כל הבטחה כי לא קיימת אסטרטגיה טובה יותר לשידור עם שגיאות במקרה המוצע עם פחות אנטרופיה.

יתירויות בתמונות וניצולן

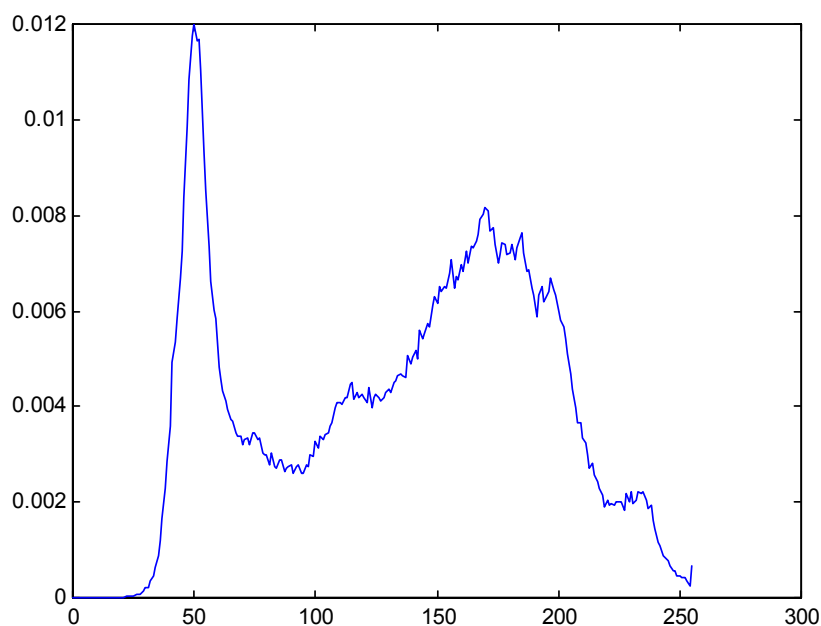
מהדין בנושא האנטרופיה ופונקציית קצב-העיוות ברור כי דחיסה מצליחה בשל יתירות באות הנדחס. יתירות פירושה שהמידע מכיל כפילויות מיותרות הניתנות לניצול. עד כה ראינו מספר סוגי יתירויות עקרוניים. נזכיר אותם ונראה כיצד הם ישימים בתמונות:

1. יתירות הסתברותית - אפילו אם המקור שלנו מניב תווים מתוך סט נתון של N תווים אפשריים ובלי תלות ביניהם, אם לתווים שונים הסתברות שונה נקבל כי האנטרופיה נמוכה מ- $\log_2 N$.

אם נחזור לרגע לתמונות, ונזכור כי היסטוגרמה של תמונה היא פונקציית ההסתברות של רמות האפור בה, הרי שאם ההיסטוגרמה אינה אחידה, ישנה כאן יתירות הסתברותית הניתנת לניצול לשם קבלת דחיסה. למעשה, משיקולים אלו ברור כי ככל שההיסטוגרמה רחוקה מלהיות אחידה, כך התמונה דחיסה יותר.

דוגמה: ניקח את התמונה Lena בה יש 512 על 512 פיקסלים המיוצגים ע"י 8 סיביות כל אחד. עפ"י ההיסטוגרמה המנורמלת שלה נחשב את האנטרופיה. ההיסטוגרמה מתוארת בציר 8.2, והאנטרופיה המתקבלת היא:

$$H_{\text{Lena}}(\text{pixel}) = \sum_{k=0}^{255} h(k) \log_2 \frac{1}{h(k)} = 7.58 \text{ bits / pixel}$$



היסטוגרמה מנורמלת של התמונה Lena. האנטרופיה עפ"י היסטוגרמה זו (ללא ניצול של קשר בין פיקסלים) היא 7.58 סיביות לפיקסל.

כלומר, רק מתוך ההיסטוגרמה ברור כי ניתן בפחות מ- 8 סיביות לייצג כל פיקסל. כמו כן אנו רואים כי הרווח אינו גדול - נוכל (תיאורטית בשלב זה) לחסוך כחצי סיבית לפיקסל.

2. יתירות בין תווים - ראינו כי כאשר עובדים על בלוקים של תווים וקיימת תלות סטטיסטית ביניהם, מתקבלת יעילות טובה יותר של הדחיסה. בהשאלה לתמונות נאמר כי כאשר אנו עומדים לקודד את הפיקסל ה- $[i,j]$ בתמונה, אם ננצל את המידע שמספקים שכניו של פיקסל זה, נצליח טוב יותר בדחיסה. זאת כיוון שקיימת תלות בין פיקסלים סמוכים בתמונה - הם אמורים להיות בעלי רמות אפור קרובות, ולכן קיימת תלות סטטיסטית חזקה ביניהם.

דוגמה: נחזור לתמונה Lena. על מנת לעבוד בבלוקים נדרש לבנות היסטוגרמה של צמדים. היסטוגרמה זו תדרוש כ- 65000 תאים (256 בריבוע), וברור כי מלאכת בניית היסטוגרמה כזו לא תהיה נעימה. כמו כן ברור כי לבלוקים גדולים יותר נקבל היסטוגרמות בנפחים בלתי-אפשריים. מעבר לכך, אפילו אם נתעלם מהנפחים הנוצרים, הרי שבתמונה זו יש 512^2 פיקסלים, כלומר, בממוצע - רק פי

ארבע יותר צמדים מכמות התאים לעבודה על צמדים. לכן, אין מספיק מידע בתמונה לשם בניית פונקציה הסתברות דו-מימדית.

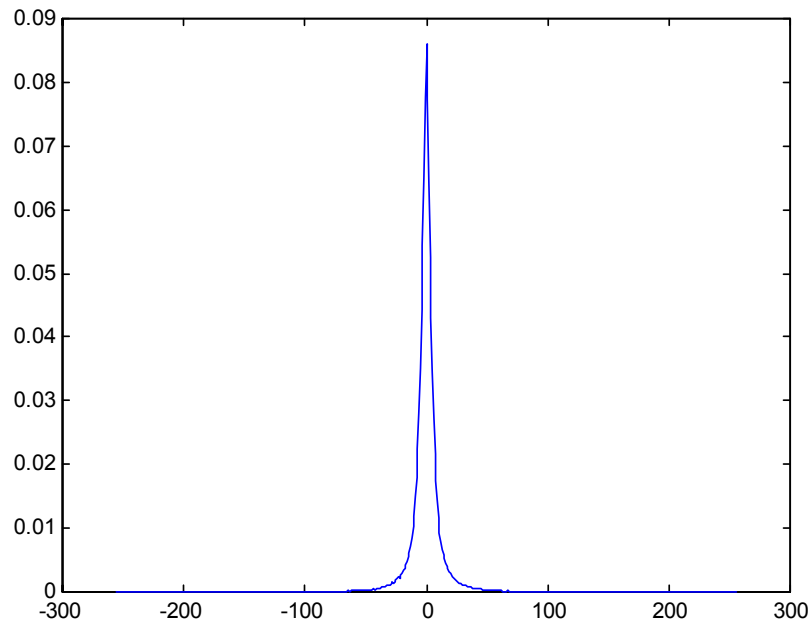
כתחליף לעבודה בבלוקים, ראינו כי ניתן להציע עבודה עם חיזוי. נמחיש רעיון זה על דחיסת התמונה הנתונה. נסדר את כל הפיקסלים בתמונה בעמודה בסידור לקסיקוגרפי, ונחליט כי במקום לשדר את הפיקסל $X[k]$, נשדר את ההפרש שלו משכנו הקודם - $X[k-1]$. במקלט כבר נקלט מידע על-פיו יודעים את $X[k-1]$, ולכן בהינתן ההפרש - $E[k] = X[k] - X[k-1]$, ניתן יהיה לקבוע את $X[k]$ עצמו. נתבונן אם כך בהיסטוגרמה של הוקטור $E[k]$ - היסטוגרמה זו נתונה בציור 8.3. ברור כי ערכי ההפרש נעים בתחום -255 ועד +255, כך שהתחום הדינמי גדל. לעומת זאת, עיון בהיסטוגרמה זו מגלה כי רוב השגיאות מרוכזות קרוב לראשית (לאפס), והפילוג נראה פילוג לפלס בעל הצורה:

$$P(e) = \frac{\lambda}{2} \cdot \exp\{-\lambda|e|\}$$

מסתבר כי שגיאות חיזוי בתמונות בדרך כלל בעלות חוק פילוג כזה.

האנטרופיה במקרה זה היא 5.14 סיביות לפיקסל. ברור כי החיזוי המוצלח ביותר הוא לעשות שימוש בפונקציה ההסתברות המותנית, אך זו קשה לקבלה (היא גם משתנה במקום) וקשה לשימוש. כפי שאנו רואים, ע"י חיזוי פשוט למדי (אנו נראה בהמשך שניתן להציע טובים ממנו שיהיו מעשיים) קיבלנו אנטרופיה נמוכה בכ- 3 סיביות מהשמונה המקוריים.

כאשר דנים בדחיסת סדרות של תמונות אנו ננסה לנצל גם את הקשר ההדוק בין פיקסלים בין תמונות. מסתבר שקשר זה אף חזק יותר בהשוואה לקשר בין פיקסלים שכנים מאותה תמונה, כיוון שתמונות עוקבות בוודאק מאוד דומות. זהו בעצם המקור ליכולת הדחיסה החזקה כל-כך בוודאק.



היסטוגרמה מנורמלת של תמונת ההפרשים של Lena. האנטרופיה עפ"י היסטוגרמה זו (ללא ניצול של קשר בין פיקסלים) היא 5.14 סיביות לפיקסל.

3. יתירות פסיכו-ויזואלית - שתי היתריות הנ"ל מתייחסות למושג האנטרופיה. כעת ננצל את שלמדנו על פונקציית קצב-העיוות. מערכת הראיה שלנו אינה מושלמת, וניתן לנצל עובדה זו לשם השגת דחיסה טובה יותר.

לכן, אם רמות אפור בתמונה הנפרסת (לאחר דחיסה אובדנית) שונות במקצת מרמות האפור בתמונת המקור, במרבית המקרים אין אנו מסוגלים לזהות שינוי זה. כיוון שכך, במושגים של תורת האינפורמציה, אנו מתירים מרחק D כלשהו בין תמונת הכניסה ותמונת היציאה של תהליך הדחיסה-פריסה. כבר ראינו כי מרחק כזה מניב אנטרופיה קטנה יותר - כלומר - המידע מכיל אפקטיבית יותר יתירות.

לסיכום דיון זה על יתירויות, יתירות הסתברותית סקלרית מציעה חסכון קטן בייצוג האות. אם מוסיפים לה התייחסות להסתברויות בין פיקסלים, אנו בעצם עושים שימוש ביתירות מרחבית (או זמנית) בין פיקסלים, ואז ניתן לדחוס בערך פי שתיים. כאשר מכניסים למשחק את היתירות הפסיכו-ויזואלית ומתירים שגיאות קלות, יחס הדחיסה יכול לעלות לפי 10 ועד מאות - תלוי באיכות הרצויה ובאופי התמונה/ות המטופל.

עד כה עסקנו באנטרופיה תיאורטית. נשאלת השאלה - כיצד בונים מקודד מעשי המצליח להגשים את האנטרופיה או לפחות להתקרב לכך. באשר לקידוד ללא שגיאות בו אנו חותרים אל האנטרופיה, הכלי אותו נציג מיד הוא קוד הופמן. באשר לקידוד עם שגיאות, נזכור כי לרוב לא ניתן כלל לקבוע מהי האנטרופיה האפקטיבית. אנו נכיר מספר כלי יסוד המשמשים באלגוריתמי דחיסה אובדנית.

קידוד אנטרופיה - אלגוריתם הופמן

נניח כי נתון לנו מקור X המייצר תווים מעל אלף-בית של N תווים אפשריים (למשל - קוביה המייצרת מספרים בתחום $[1, 6]$). עבור מקור זה ידוע לנו חוק ההסתברות $P(X)$, ועל פי פונקציה זו אנו יכולים לחשב את האנטרופיה התיאורטית $H(X)$.

רצוננו לבנות קוד אשר יגשים קצב שידור שיהיה האנטרופיה או קרוב אליה. לשם דוגמה, בהינתן תמונה בת מיליון פיקסלים כשכל פיקסל בעל רמת אפור בתחום $[0, 255]$, ברור כי ע"י הקצאת 8 סיביות לפיקסל בנינו קוד מעשי, אך אם האנטרופיה של רמות האפור (כסקלרים - אנו עדיין לא מדברים על בלוקים או על חיזוי) ידועה להיות 4.5 סיביות לפיקסל - כיצד נתכנן קוד שירד לרמה זו? לשם מתן תשובה, נציג דוגמה פשוטה וממנה נכיר שני עקרונות חדשים - קוד בעל אורך מילה משתנה, וקוד Prefix.

דוגמה: נתון מקור המייצג 6 תווים אפשריים $\{A, B, C, D, E, F\}$ עם ההסתברויות הבאות:

$$P\{A\} = 1/2, P\{B\} = 1/8, P\{C\} = 1/8, P\{D\} = 1/8, P\{E\} = 1/16, P\{F\} = 1/16$$

רואים כי סכום הסתברויות אלו הוא 1, כנדרש. בקידוד טריוויאלי נוכל להציע הקצאת 3 סיביות לכל תו ויהיה בכך במזל כיוון שיוותרו שני צירופים ללא שימוש. תוצאה זו צפויה כיוון שתיאורטית, עבור 6 תווים די ב- $\log_2 6 = 2.585$ סיביות לתו. במקרה שלנו האנטרופיה היא 2.125 סיביות לתו.

אם הגישה הטרוויאלית מקצה אורך זהה של מילות קוד לכל תו, גישה יעילה יותר תהיה להקצות מילים באורך שונה לתווים שונים - תו בעל הסתברות גדולה יקבל מילה קצרה, ותו בעל הסתברות נמוכה יקבל מילה ארוכה. בדוגמה שלנו - נשתמש בקוד הבא:

אורך	קוד	הסתברות	תו
1	0	1/2	A
3	100	1/8	B
3	101	1/8	C
3	110	1/8	D
4	1110	1/16	E
4	1111	1/16	F

האורכים אותם בחרנו הם בדיוק LOG לפי בסיס 2 של היפוך ההסתברות של כל תו. לכן, אורכה של המילה הממוצעת בקוד יהיה:

$$E[L] = \sum_{k=1}^N P\{X_k\} L_k = 1 \cdot \frac{1}{2} + 3 \cdot \frac{1}{8} + 3 \cdot \frac{1}{8} + 3 \cdot \frac{1}{8} + 4 \cdot \frac{1}{16} + 4 \cdot \frac{1}{16} = 2.125 \text{ bits / pixel}$$

זוהי בדיוק האנטרופיה, כך שיש לנו קוד מוצלח למדי. השאלה הנותרת היא האם הוא ניתן לפענוח? בקוד בו יש מספר קבוע של סיביות לכל תו, כל שעל המקלט לעשות הוא לפרק את סידרת הסיביות המגיעה לקבוצות שוות סיביות וכל קבוצה תייצג תו. כעת, כשמגיעה סידרת סיביות נדרש לפרקה לאורכים שונים, והשאלה אן זה אפשרי. נשים לב לתכונה הבאה המתקיימת בקוד שהוצע - אף מילת קוד קצרה אינה נהווה את הפתיח למילות ארוכות יותר. תכונה זו נקראת Prefix Code, והיא זו המבטיחה פענוח יחיד ונכון.

לדוגמה - נניח כי רצוננו לשדר את הסדרה - AABEACAADAFDBACA - נבנה עבור צירוף תווים אלו את סידרת הסיביות שתשודר - 0,0,100,1110,0,101,0,0,110,0,1111,110,100,0,101,0 סימנים אותם המקלט לא יכול לקבל - עליו לפרק את הסדרה לבדו, ובזכות היות הקוד בעלת תכונת ה-Prefix הוא יכול. התו הראשון הוא 0, ולכן המילה היחידה האפשרית היא A. כך גם לסיבית השניה. בסיבית השלישית יש 1 ולכן זה יכול להיות B עד F. 1 כשלאחריו 0 יכולים להיות רק B ו-C, ו-1 כשלאחריו

שני אפסים זה רק B. כך התהליך נמשך כשישנו רק פסק אפשרי אחד - וזהו הפסק הנכון.

לסיכום דוגמה זו - ראינו כי קיים קוד המגשים את האנטרופיה בדיוק וזאת ע"י קוד בעל אורך מילה משתנה - מילים קצרות לתווים הסבירים יותר ומילים ארוכות לתווים נמוכי ההסתברות. כמו כן ראינו כי היות הקוד קוד Prefix מאפשרת פענוח יחיד ונכון של הסדרה. סוגיה שלא זכתה למענה עד כה - מאיפוא נולד הקוד שהוצע, וכיצד נוכל תכנן קוד למקרים אחרים?

קוד הופמן דן במקרה הכללי ומנסה להכליל את התוצאות הנ"ל. אנו דנים במקור X המייצר תווים מעל אלף-בית של N תווים אפשריים, ויש בידנו את חוק ההסתברות $P(X)$, והאנטרופיה $H(X)$. אלגוריתם הופמן מציע תהליך פשוט לפיו ייבנה קוד Prefix מוצלח. התהליך הוא תהליך של בניית עץ:

1. כתוב את סט N ההסתברויות בשורה. קבע $k=N$.
2. מתוך k התווים הקיימים, אחד את שתי התווים בעלי ההסתברויות הנמוכות ביותר. האיחוד ייעשה ע"י שני ענפי-עץ המתאחדים לצומת אחת. צומת זאת מגדירה תו חדש שהוא צירוף שני התווים שאוחדו, והסתברותו היא סכום הסתברותם. בסיום שלב זה מתקבל סט של $k-1$ תווים עם הסתברות לכל אחד, כשסכום ההסתברויות הוא 1.
3. אם $k > 1$ חזור לסעיף 2. ואחרת - סיים.

התהליך שתואר בונה עץ. הקוד מתקבל ע"י הקצאת 0 ו-1 לכל לשני הענפים היוצאים מכל צומת. סיברת הסיביות המגדירה מסלול משורש העץ ועד לעלה היא מילת הקוד של התו המתייחס לעלה זה. נמחיש על תהליך זה דרך דוגמה.

דוגמה: נחזור לדוגמה הקודמת בה נתון מקור המייצג 6 תווים אפשריים -

עם ההסתברויות $\{A, B, C, D, E, F\}$

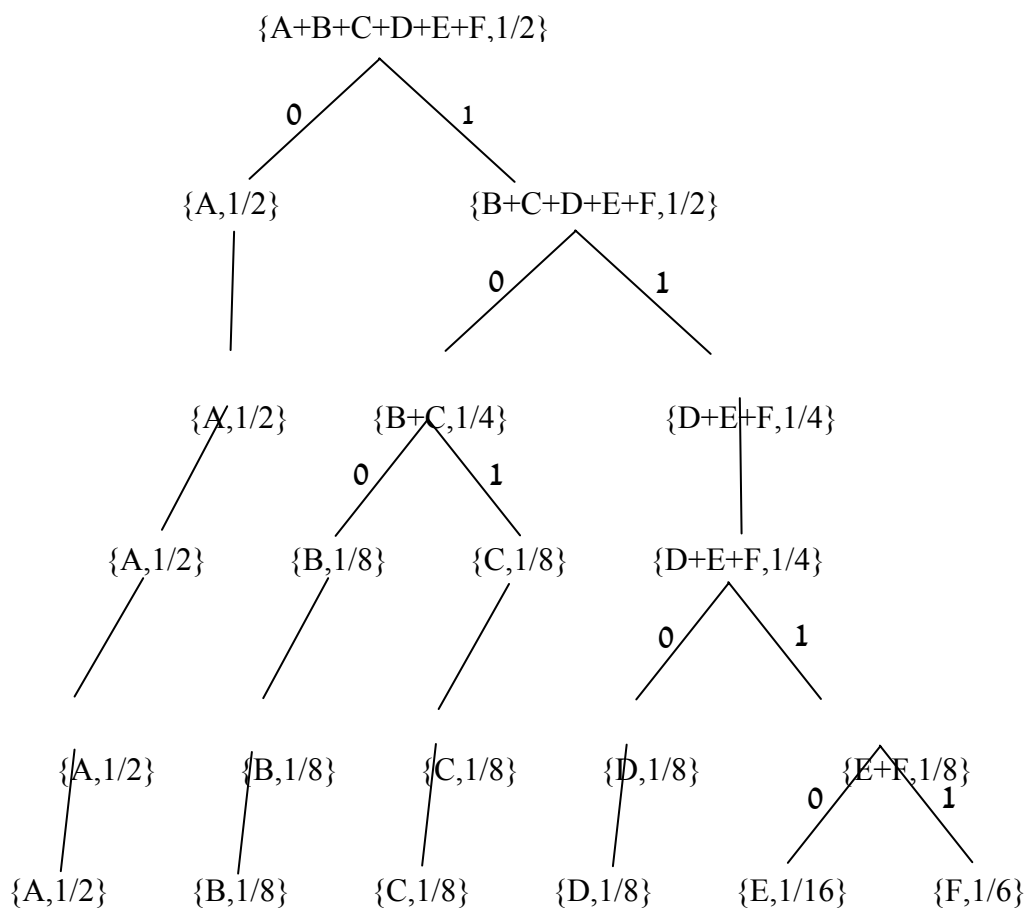
$$P\{A\} = 1/2, P\{B\} = 1/8, P\{C\} = 1/8,$$

$$P\{D\} = 1/8, P\{E\} = 1/16, P\{F\} = 1/16$$

הציור הבא מראה את העץ הנוצר.

בשורה התחתונה ישנה רשימה של כל התווים על הסתברויותיהם. בשלב ראשון מאוחדים E ו-F ביותרם בעלי ההסתברות הנמוכה ביותר. נוצר תו חדש $E+F$ והסתברותו שמינית. בשלב שני יש ארבעה תווים בעלי הסתברות שמינית - בחרנו לאחד את D על $(E+F)$. כך התהליך ממשיך עד לקבלת תו אחד בעל הסתברות 1, שהוא למעשה כל התווים כולם.

בשלב שני אנו מקצים סיביות לכל שני ענפי צומת. כך למשל, על מנת לצאת משורש העץ (למעלה) ועד התו E עלינו ללכת במסלול שמותווה ע"י ההחלטות 1, 1, 1, ולבסוף 0 - זוהי בדיוק המילה שהוצעה במסגרת דוגמה קודמת. למעשה - בדיקה קצרה מגלה כי עץ זה מניב במדויק את הקוד שהוצע בדוגמה קודמת.



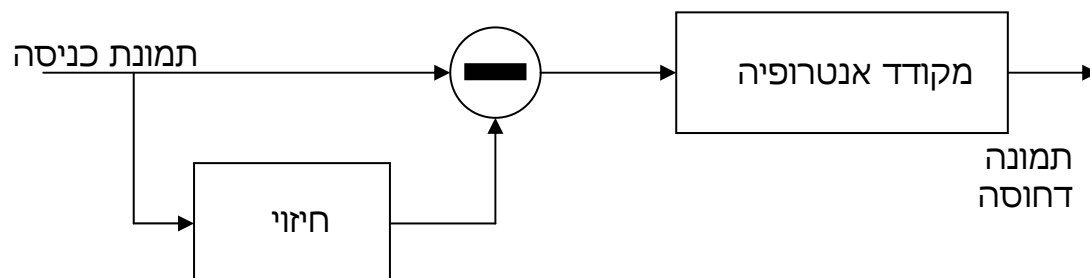
בניית העץ לשם בניית קוד הופמן

במקרה בו ההסתברויות מהוות חזקות שלמות שליליות של 2 (כמו הדוגמה הנ"ל) מתקבל כי קוד הופמן מגשים בצורה מדויקת את האנטרופיה, דהיינו, אורך המילה הממוצעת היא האנטרופיה. במקרה כללי יותר קוד הופמן ייתן אורך מילה שיהיה גבוה מהאנטרופיה בעד סיבית אחת לתו, כלומר:

$$H(X) \leq L_{\text{Average}} < H(X) + 1$$

כאשר נתון סט הסתברויות בו קוד הופמן סוטה בקרוב לסיבית אחת, ניתן להתגבר על כך ע"י עבודה בבלוקים. גם אם אין כל קשר סטטיסטי בין התווים, אם נתייחס לדוגמה לצמד תווים כ-"סופר-תו" חדש, הרי שנקבל סטייה של עד סיבית אחד ל-"סופר-תו". לכן, לתו בודד תתקבל סטייה של עד חצי סיבית. במקרה הכללי, עבודה על בלוק של N תווים תיתן סטייה של עד $1/N$ סיביות. ברור כי תרגיל זה בא עם מחיר - עבודה על בלוקים קשה יותר. לדוגמה - עבור מקור בין 100 תווים, קוד הופמן המקורי דורש בנייה של 100 מילות קוד ושימוש בהן. עבור צמדים נדרש לתכנון ואחסון של 10,000 מילות קוד, ותהליך הדחיסה\פריסה מסתבך בהתאם.

בהסתמך על כל האמור לעיל נוכל להציע אלגוריתם דחיסה פשוטני ללא שגיאות לתמונה נתונה. אלגוריתם זה יקודד את ההפרשים בין פיקסלים סמוכים, וכבר ראינו כי האנטרופיה הסקלרית של מידע זה היא 5.14 סיביות. אם נשתמש בקוד הופמן נוכל לקבוע כי במקרה החמור ביותר נצליח להגיע ל-6.14 סיביות לפיקסל - דחיסה לא חזקה אך סבירה, ובמיוחד לנוכח פשטות האלגוריתם. אם נשתמש בחיזוי מוצלח יותר, ו\או נקודד את שגיאות החיזוי בבלוקים נוכל לסחוט דחיסה טובה יותר. הציור הבא ממחיש בסכימת בלוקים את אלגוריתם הדחיסה המוצע.



מקודד פשוט ללא שגיאה המבוסס על חיזוי וקידוד אנטרופיה כדוגמת אלגוריתם הופמן

נקודה אחרונה בדיון זה על קודים בעלי אורך משתנה היא מידת רגישותם לשגיאות ערוץ. אם מופעל קוד הופמן על סידרת תווים ומתקבלת סידרת סיביות, די בשינוי של סיבית אחת בסדרה כדי לשבש את הפענוח של כל מה שיבוא אחריה. ניתן להילחם בתופעה זה במגוון דרכים, אך אנו לא ניכנס לנושא זה.

עיבוד תמונות בעזרת מחשב (236327)

הרצאה מס' 11

דחיסת תמונות – שיטת החיזוי

גישת החיזוי בדחיסה

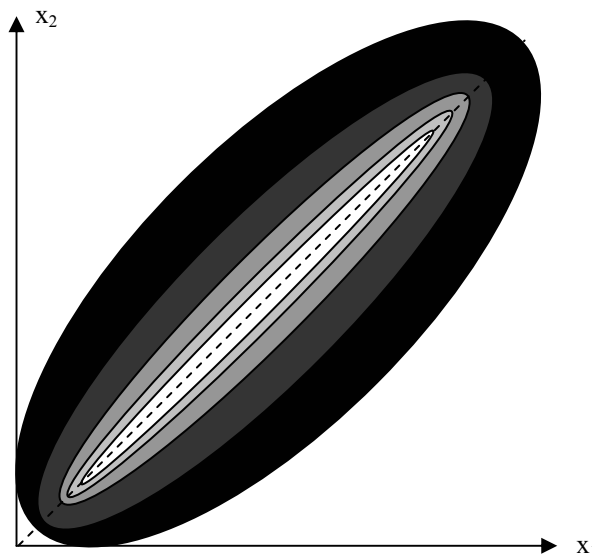
באופן כללי ישנן שתי גישות לביצוע דחיסה עם שגיאות - גישת החיזוי וגישת ההתמרה. אנו נציג את שתי גישות אלו במסגרת הדיון בדחיסת תמונות. בשתי גישות אלו, המרכיב האחראי על אובדן המידע הוא קוונטיזר, כפי שכבר הוצג בעבר. ההבדל המהותי הוא בשאלה מהי הפעולה המקדימה שיש לבצע על הפיקסלים בטרם יוכנסו לקוונטיזר - חיזוי (כפי שכבר עשינו בסעיפים קודמים), או התמרה. שתי אפשרויות אלו בעלות מכנה משותף חשוב - שתיהן ממירות את הפיקסלים עתירי היתירות במספרים בעלי אנטרופיה נמוכה יותר, וקשר סטטיסטי דל בין האלמנטים השונים. כתוצאה מכך, המידע לאחר פעולת החיזוי/התמרה נוח יותר לדחיסה, ומאפשר עבודה בפיקסלים בודדים במקום בוקטורים.

למה חיזוי? ראינו את המוטיבציה קודם, בהקשר לדחיסה ללא אובדן, אך נעמיק כעת את ההסבר. ראינו כי ניתן לדחוס תמונה ע"י טיפול בכל פיקסל במנותק משכניו. ראינו גם שהדבר נותן יחסי דחיסה גרועים כי היתרות המנוצלות אינה גדולה. אם כך יש לנו מוטיבציה לעבור לטיפול בצמדים של פיקסלים, שלשות, ואף קבוצות גדולות יותר. נניח כי נעבוד בצמדים ונראה כי אפילו אז עולים קשיים בלתי עבירים.

כדי לדחוס תמונה עי טיפול בכל צמד פיקסלים במנותק מצמדים אחרים עלינו לספק את ההסתברות המשותפת $P(x_1, x_2)$ של צמד כזה, ואז לבנות קוד שישתמש בהסתברות זו. בניית $P(x_1, x_2)$ מחייבת צבירת היסטוגרמה משותפת בה יהיו 256^2 תאים. אם אנו מחזיקים תמונה בגודל של 256 על 256 פיקסלים, יהיו לנו פחות צמדים מתאים בהיסטוגרמה זו, וכך שערור $P(x_1, x_2)$ יהיה בעל

דיוק ירוד. כך שכבר אנו נתקלים במכשלה של אימון ההיסטוגרמה על מנת לבסס קוד יעיל לדחיסה – נדרשת כמות נתונים גדולה.

נניח שעברנו שלב זה בהצלחה (ע"י בניית היסטוגרמה כזו על מקבץ תמונות גדול). כיצד זו תיראה? ניתן לבצע ניסוי ולבנות פונקציה כזו, אך אפילו ללא נתונים, קל יהיה להתנבא שהפונקציה $P(x_1, x_2)$ תיראה כך:



תיאור ההיסטוגרמה המשותפת לצמדי פיקסלים.

פונקציה זו תהיה בקירוב טוב בעלת המבנה

$$P(x_1, x_2) = C \cdot \exp\{|x_1 - x_2|\}$$

שמשמעותה היא שבידיעת אחד מהצמד, השני שווה לו עם סטיות מסוימות מעלה-מטה בהסתברות הדועכת אקספוננציאלית.

נניח כי בנינו פונקציה זו – בעיה שני שתתעורר היא הצמדת מילת קוד כנגד כל צמד אפשרי וישנם 256^2 כאלו. בקוד הופמן למשל, יהיה עלינו לבנות טבלה בלתי נסבלת בגודל זה. בעיה שלישית שתתעורר היא שבשימוש ב- $P(x_1, x_2)$ נמצא עצמנו מתעלמים לחלוטין מהקורלציה הרבה הקיימת בין צמדים סמוכים.

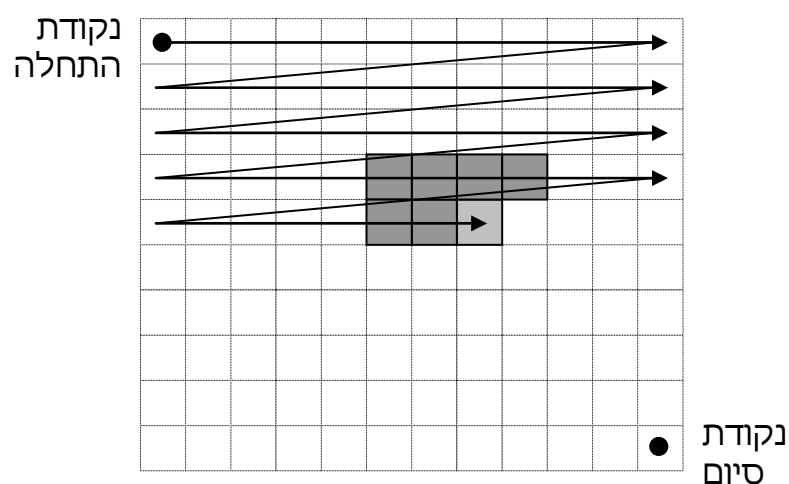
בשל כל סיבות אלו נעבור לשימוש בהסתברות מותנית. ראינו כי זו תנצל קורלציה בין כל צמד פיקסלים עוקבים, ומספר מילות הקוד הנדרש יהיה בסדר גודל של 256. אם ננצל את המבנה הנ"ל של $P(x_1, x_2)$ ונניח כי $P(x) = \text{Const}$ -

כלומר כל רמת אפור הינה בקירוב בעלת הסתברות קבועה, נקבל:

$$P(x_1 / x_2) = P(x_1, x_2) / P(x_2) = C_0 \cdot \exp\{|x_1 - x_2|\}$$

כלומר, הערך הידוע מהצמד, x_2 , יהווה חיזוי לערכו של x_1 , ובעצם הממוצע שלו, וההסתברות תהיה לפלאסית (אקפוננציאלית עם חזקה 1) כמתואר. כך "נולד" קונספט החיזוי כמימוש בפועל של ההסתברות המותנית. ואז נרוויח שימוש בפילוג סקלרי, מספר מילות קוד מצומצם, וניצול מוצלח של קורלציה לכל אורך סידרת הפיקסלים.

נתחיל את הדיון בסכימה המעשית המבוססת על חיזוי. צעד ראשון לשם יישום של אלגוריתם דחיסה עם חיזוי הוא קביעת סדר סריקה לפיקסלים. החיזוי הינו חיזוי סיבתי - כלומר ערך החיזוי נבנה ע"י פיקסלים שכבר טופלו, וזאת על-מנת שגם המקלט יוכל לבנות את החיזוי ולצרפו לשגיאת החיזוי הנקלטת. בעוד שבאות חד-מימדי סיבתיות היא עובדה טבעית, בתמונה (אות דו-מימדי) אין זה כך. נניח לעת עתה כי התמונה מסודרת שורה לאחר שורה כמתואר בציור הבא - סידור לקסיקוגרפי לפי שורות. בציור זה גם מתוארת סביבת פיקסלים סיבתית עפ"י סדר זה, אשר יכולה לשמש לשם בניית החיזוי.



סדר סיבתי בתמונה ע"י סידור לפי שורות. לפיקסל נתון (המתואר באפור בהיר), סביבה סיבתית הקרויה Non-Symmetric Half Plane (NSHP) מתוארת באפור מעט יותר כהה.

ישנן גישות שונות לסדר הפיקסלים – ראינו בעבר את שיטת Hilbert-Peano, ואמנם נעשה בה שימוש בהמשך.

הראינו קודם כיצד פועל אלגוריתם עקרוני של דחיסה מבוססת חיזוי ללא שגיאות. על מנת להכניס שגיאות (שכזכור, צפויות לשפר את עומק הדחיסה), עלינו להכניס לפני קידוד האנטרופיה בלוק חדש - קוונטיזר. הקוונטיזר יזן בשגיאת החיזוי אשר יכולה לקבל ערכים שליליים וחיוביים. במקום לקודד את השגיאות במדויק, הקוונטיזר יבצע "העגלה" של שגיאות אלו בה חלק מהמידע ייזרק, וכתוצאה מכך תיווצר שגיאה בתהליך הדחיסה. בזכות "העגלה" זו, מקודד האנטרופיה ידקק להרבה פחות סיביות לשידור שגיאות אלו, וכך יושג הרווח בדחיסה.

דוגמה: נניח כי החיזוי שלנו לפיקסל $X[k, j]$ הוא ממוצע של שכניו הסיבתיים לפי

$$\hat{X}[k, j] = \frac{1}{4} (X[k-1, j-1] + X[k-1, j] + X[k-1, j+1] + X[k, j-1])$$

בחיזוי זה עשינו שימוש בשלושת הפיקסלים המצויים מעליו, ובזה המצוי באותה שורה לשמאלו. בהנחה שכל פיקסל מקבל רמת אפור המיוצגת ע"י 8 סיביות, ערכי חיזוי זה ינועו בתחום -255 עד +255 ויקבלו כל ערך בתחום זה בקפיצות של 0.25. לכן, יהיו 2041 ערכים אפשריים (-255:0.25:255). דחיסה ללא שגיאה של אות שגיאה זה יחייב בנייה של קוד הופמן ל- 2041 ערכים אפשריים.

כתחליף, אם נפעיל קוונטיזר המעגל את השגיאה לערך השלם הזוגי, נקבל במוצא הקוונטיזר ערכי שגיאה חדשים באותו תחום אך עם 255 ערכים בלבד. זה יתקבל ע"י המשוואה $Q\{E\} = \lfloor E + 0.5 \rfloor$. באופן דומה, שימוש ב-

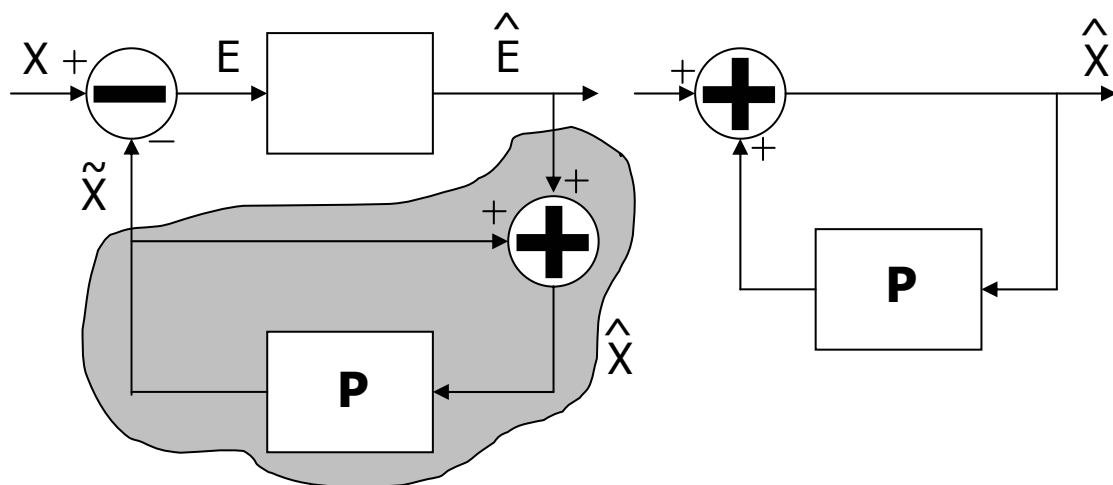
$$Q\{E\} = 2 \cdot \left\lfloor \frac{E+1}{2} \right\rfloor$$

ימפה את הערכים (-1,+1) לערך 0, (+1,+3) ל- 2 וכך הלאה. נמחיש רעיון זה על הפיקסלים בתמונה Lenna. נבנה לכל פיקסל את החיזוי כמתואר ונחשב את האנטרופיה ללא קוונטיזציה ועם קוונטיזציה כמתואר. לצורך חישוב האנטרופיה נזרוק את העמודות והשורות הקיצוניות, בהן החזאי נופל משטח התמונה. האנטרופיה ללא קוונטיזציה היא 4.86 סיביות לפיקסל, ואילו עם הקוונטיזציה מתקבלת אנטרופיה של 2.78 סיביות לפיקסל.

גישת החיזוי עם שגיאות – סכימת ה-DPCM

במקודד ללא שגיאה, הפיקסלים הנקלטים הם הפיקסלים המקוריים (ללא שגיאה מעצם הגדרתם ככאלה). לכן, המקלט מצליח לבנות חיזוי הזהה בערכו לחיזוי בו נעשה שימוש במשדר. בדחיסה עם שגיאה נוצרת בעיה - אם נבנה את החיזוי עפ"י הפיקסלים הסיבתיים המקוריים במשדר, ובמקלט ננסה לחקות תהליך זה ע"י בניית חיזוי על הערכים הנקלטים, נקבל ערכי חיזוי שונים. יתכן שהסטייה תהיה קטנה, אך זו תשפיע על הערכים המפוענחים הבאים ותיווצר תופעה של שגיאה מצטברת. כעבר מספר עשרות פיקסלים המפענח יניב ערכים שגויים לחלוטין וללא קשר לערכי המקור.

על מנת להתגבר על בעיה זו, על המשדר לבנות את החיזוי על פי ערכי רמות האפור כפי שיפוענחו במקלט. לכן, על המשדר להכיל בתוכו תת-מערכת שתהיה מקלט לכל דבר אשר מפענח את הערכים, ותת-מערכת זו היא אשר תניב את ערכי הפיקסלים על-פיהם ייבנה החיזוי. תהליך זה מתואר במלואו בצירוף הבא.



סכימת ה-DPCM המבצעת דחיסה עם שגיאות המבוססת על חיזוי. מצד שמאל המשדר ומצד ימין המקלט. החלק המסומן באפור הוא "סימולציה" של המקלט בתוך המשדר.

צירוף זה מראה סכימה עקרונית בה האות המקורי מוזן לחזאי המנבא את הדגימה $X[k,j]$ על סמך שכניה הסיבתיים. חיזוי זה מסומן ב- $\hat{X}[k,j]$. כאמור, על החיזוי במשדר להיות מבוסס על הערכים כפי ששודרו ולא עפ"י הערכים

המקוריים. אנו נסמן את ערכי הדגימות הנקלטות במקלט ב- $\hat{X}[k, j]$. החיזוי הוא למשל פונקציה מהצורה:

$$P\{X[k, j]\} = \tilde{X}[k, j] = f\{\hat{X}[k-1, j-1], \hat{X}[k-1, j], \hat{X}[k-1, j+1], \hat{X}[k, j-1]\}$$

בשלב השני מחוסר החיזוי מהערך האמיתי ומתקבלת השגיאה

$$E[k, j] = X[k, j] - \tilde{X}[k, j]$$

שגיאה זו מוזנת לקוונטיזר Q אשר ממיר את ערכי השגיאה בערכים מקורבים שלהם. זאת כיוון שאנו מוותרים על שידור השגיאות במדויק ורצוננו לשדר ערכים דומים להם. ברור כי הקוונטיזר היא פונקצית מדרגות, בה ככל שהמדרגות גדולות יותר, שגיאת הקוונטיזציה גדולה יותר. במוצא הקוונטיזר מתקבל

$$\hat{E}[k] = Q\{E[k]\}$$

גודל זה הוא המשודר החוצה, ולצורך כך נעשה שימוש במקודד אנטרופיה כדוגמת קוד הופמן. אם הקוונטיזר אינו מאבד מידע (כלומר מהווה מערכת יחידה), אנו מקבלים מערכת דחיסה ללא שגיאה. המקלט מקבל את $\hat{E}[k]$ ומחבר אליו את החיזוי $\tilde{X}[k, j]$. לכן:

$$\begin{aligned} \hat{X}[k, j] &= \hat{E}[k, j] + \tilde{X}[k, j] = \\ &= (E[k, j] + \tilde{X}[k, j]) - (E[k, j] - \hat{E}[k, j]) = \\ &= X[k, j] - (E[k, j] - \hat{E}[k, j]) \end{aligned}$$

כלומר, המוצא של המקלט הוא הערך האמיתי של הפיקסל עם תוספת שגיאה שהיא בדיוק שגיאת הקוונטיזציה. נשים לב כי המשדר מכיל בתוכו סימולציה של המקלט לשם בניית הערכים $\hat{X}[k, j]$.

דוגמה: אם נחזור לדוגמה הקודמת, הרי שקיבלנו ירידה בשתי סיביות בקירוב לפיקסל ע"י קוונטיזציה ששגיאתה היא לכל היותר רמת אפור אחת - שגיאה שאנו לא נצליח להבחין בה. בהמשך לדוגמה זו, עבור הקוונטיזר הבא:

$$Q\{E\} = 4 \left\lfloor \frac{E+2}{4} \right\rfloor$$

יתקבל 0 עבור כניסות בתחום $[-2, 2]$, 4 עבור כניסות בתחום $[2, 6]$, וכו'. במקרה זה האנטרופיה המתקבלת היא 2.1 סיביות לפיקסל, והשגיאה המירבית הפעם היא 2 רמות אפור.

עד כה הנחנו קוונטיזר פשוט ויוניפורמי. למעשה, ניתן ליישם קוונטיזר כללי יותר בעל מדרגות ברוחב משתנה, אך אנו נעצור את הדיון בנקודה זו ולא נכנס לנושא תכנון הקוונטיזר, למרות שלדבר יכולת השפעה משמעותית על ביצועי הדוחס.

אם נסכם, הרי שלמרכזו של האלגוריתם היה תהליך החיזוי והחסרתו אשר הניב תמונה חדשה בה הפיקסלים בעלי שתי תכונות חשובות:

1. אין קשר בין הפיקסלים - תכונה זו מאפשרת לנו לעבוד על כל פיקסל בנפרד מבלי לחשוש שאנו מוותרים על יתירות מרחבית.
2. ערכי הפיקסלים יוצרים פילוג נמוך אנטרופיה (לפני הקוונטיזר ובוודאי שגם אחריו). לכן, האות דחיס בצורה יעילה.

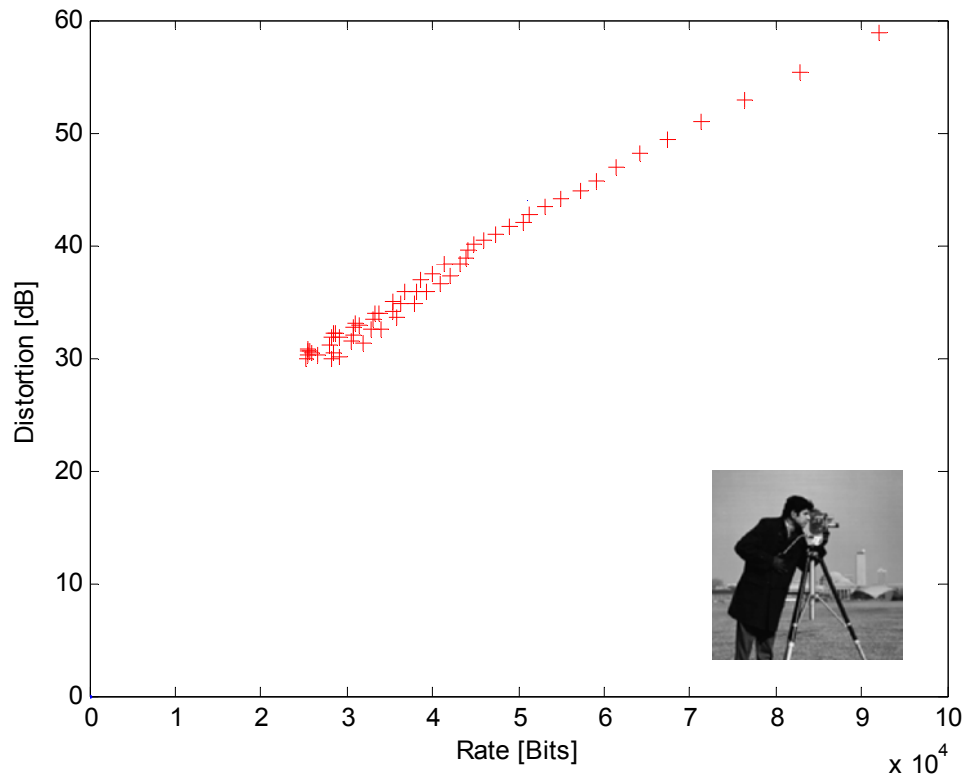
כיצד בוחנים ביצועיו של אלגוריתם דחיסה? אחת הדרכים הנוחות היא תיאור עקומת קצב-העיוות שמושגת ע"י האלגוריתם. לכל בחירה של הפרמטרים. כנגד כל בחירה כזו נוכל לקבוע את כמות הסיביות הנחוצה – זהו הציר האופקי (Rate), וכנגדו את איכות התוצאה. האיכות בדרך כלל נמדדת ע"י PSNR המחושב לפי

$$PSNR = 10 \log_{10} \left\{ \frac{255^2}{\frac{1}{N^2} \sum_{k,j} (X_{\text{true}}[k,j] - X_{\text{out}}[k,j])^2} \right\}$$

למדד זה שורשים בגישות המקובלות בהנדסת חשמל, והיחידות בהן משתמשים למדידת ה-PSNR הם דציבלים (dB). הגודל SNR הינו יחס האות לרעש, והמושג Peak SNR פירושו יחס האות המרבי לעוצמת הרעש. הרעש במקרה שלנו היא שגיאת הדחיסה-פריסה, והאות הוא רמת האפור המרבית 255. אנו רואים כי ה-PSNR פרופורציוני הפוך לשגיאה הריבועית הממוצעת. לכן ערך גבוה נחשב לערך רצוי. בפועל, ערך של 35dB נחשב לתמונה באיכות טובה. נפילה של 3dB פירושה הקטנת השגיאה הריבועית הממוצעת בפקטור 2.

הציור הבא מראה את העקומה $R(D)$ עבור דחיסה מבוססת חיזוי בסכימת DPCM. הקצב R חושב ע"י אנטרופיה ישירה על מוצא הקוונטיזר, מתוך הנחה שקיים מקודד אשר יגשים מידה זו (בניגוד לקוד הופמן למשל שיסטה בעד כדי

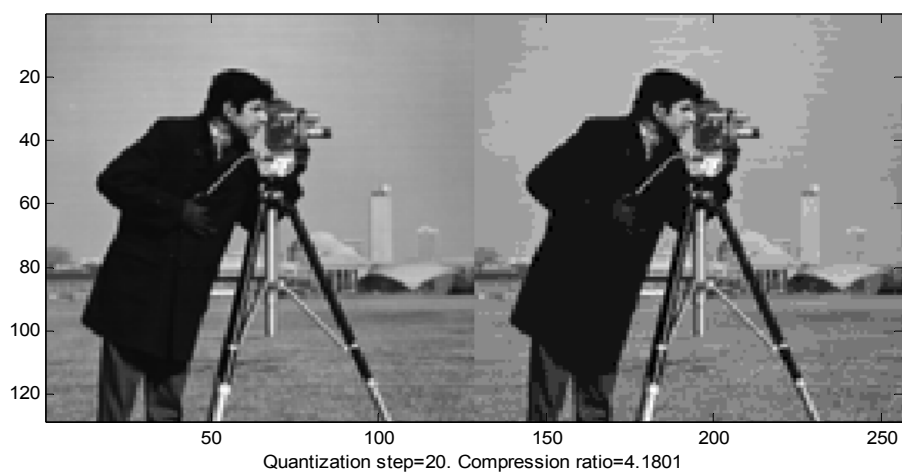
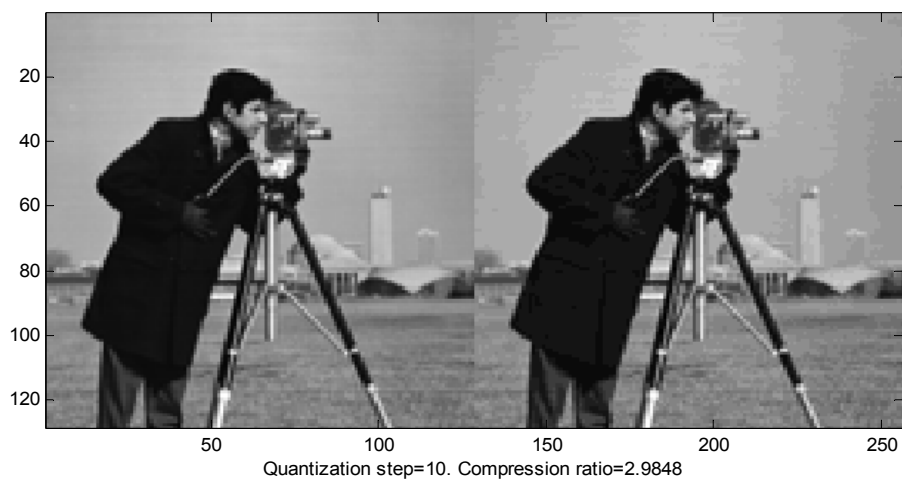
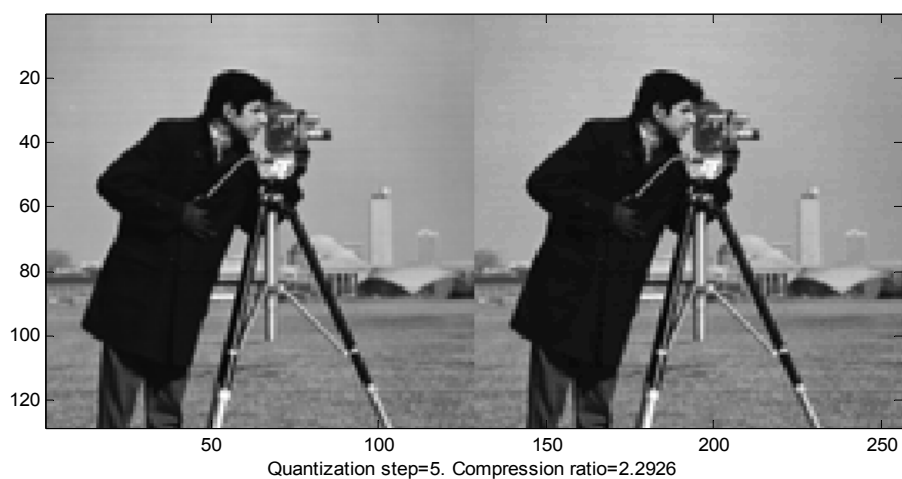
סיבית אחת). כפי שניתן לראות, עם הגדלת מדרגת הקוונטיזציה השגיאה גדלה ולכן ב- PSNR ישנה ירידה, אך איתה באה גם ירידה בכמות הסיביות הנחוצה לייצוגה של התמונה.



עקומת R(D) לדחיסת תמונה בשיטת ה- DPCM עם חיזוי על סמך פיקסל יחיד.

חשוב להבהיר – עקומה זו מתייחסת לתמונה מסוימת (התמונה המוצגת בפינת הגרף – גודלה 128-על-128 פיקסלים ולכן הכמות המקורית לייצוגה המלא היא 131000 סיביות). לכל תמונה תהיה עקומה אחרת.

התמונות הבאות מראות דוגמאות לתוצאת הדחיסה המתקבלת בגישת ה- DPCM. ניכר כי ישנן תופעות צורמות בתמונות בדחיסה ואפילו לא עמוקה. למרות זאת, סכימה זו פופולארית בשל פשטותה, ובמיוחד במערכות בהן מדובר בתמונות ענק. אלגוריתם ה- DPCM קל מבחינת מורכבותו החישובית, וניתן למימוש בחומרה. לכן מערכות לוויין משתמשות לרוב באלגוריתם זה ובדחיסה לא עמוקה (יחס 2:1).



תוצאות דחיסה באלגוריתם ה- DPCM

גישת החיזוי עם שגיאות – גישת סורק-זאבי

ישנן כמובן גם גישות אחרות הבנויות על חיזוי וקוונטיזציה. נציג כעת גישה כזו אשר הוצעה במסגרת עבודת מאסטר בטכניון בשנת 1997. שיטה זו פשוטה מרוב גישות הקידוד הקיימות ועם זאת ביצועיה יפים מאוד.

כמקודם, הכול מתחיל מסריקת התמונה והודתה לווקטור חד-מימדי. בניגוד לשיטת ה-DPCM, השיטה שנציג כאן אינה בעלת שגיאה חסומה, ולכן רצוי להשתמש בסריקה מסוג Peano לשם ערבול השגיאות במקום – אנו נבחר נקודה זו היטב בהמשך. כעת בידינו וקטור בעל N^2 פיקסלים אותו עלינו לייצג בצורה קומפקטית. וקטור זה מתאר רצף פיקסלים קשירים הלקוחים מהתמונה – ולכן ישנו קשר הדוק בין ערכיהם. בפרט, נוכל להניח כי כל פיקסל קרוב מאוד לשכניו ברמת האפור בו – ממש כשם שהנחנו קודם כשבצענו חיזוי.

הצעתם של סורק-זאבי הייתה פיצול הווקטור לקבוצות של פיקסלים דומים, כשכל קבוצה מיוצגת ע"י אורכה ורמת האפור הממוצעת בה. באופן זה הווקטור מיוצג ע"י צמדי מספרים (אורך וגובה) המאפשרים שחזור הווקטור בקירוב לא רע. כיצד נעשית החלוקה של הווקטור לאינטרוולים? הרעיון הינו להשתמש באלגוריתם סידרתי הפועל באופן הבא:

1. בחר את הפיקסל הראשון כפתיח לקבוצה הראשונה. קבע את אורך

הקבוצה $L=1$, ואת גובהה הממוצע להיות $M=x[1]$.

2. עבור לפיקסל הבא $x[k]$ ושקול את האפשרות לצרפו לקבוצה הקיימת:

a. אם $|M - x[k]| \leq T$, פירוש הדבר שהפיקסל קרוב (עד כדי קבוע T

המהווה פרמטר לאלגוריתם) לקבוצה הקיימת. לכן נצרף הפיקסל

לקבוצה ונקבע $M = (ML + x[k]) / (L + 1)$ ואז העדכון $L = L + 1$.

b. אם $|M - x[k]| > T$ הפיקסל שונה משמעותית ולכן נסגור את

הקבוצה הקודמת ונפתח בקבוצה חדשה בה $L=1$ ו- $M=x[k]$.

זהו אלגוריתם פשוט למדי, אך ביצועיו כמות שתואר אינן טובים, כי קשה לקבוע סף מתאים. סף נמוך מידי יותר קבוצות קצרות מידי וקבצי ענק, וסף גבוה אשר מוריד את הקצב יוצר שגיאות גדולות הצורמות ויזואלית. הפתרון אשר הוצע בעבודה הנ"ל הוא שימוש בסף המתחיל בערך גדול T_0 ודועך עם

גדילתה של הקבוצה – כאילו היה זה שקבוצה גדולה מידי הופכת למועדון אקסקלוסיבי המסרב לקבל חברים חדשים. בצורה כזו ישנו עידוד לקבוצות באורך בינוני כי קבוצה לא תהיה קטנה מידי כי קל למזג אליה את הפיקסלים הבאים, וקבוצה לא תהיה ארוכה מידי כי היא תסרב לקבל את החדשים. הסף יהיה

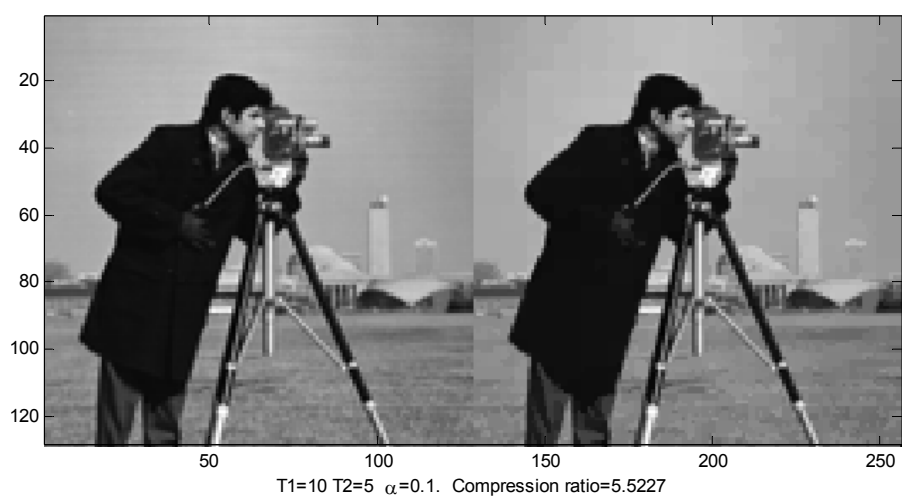
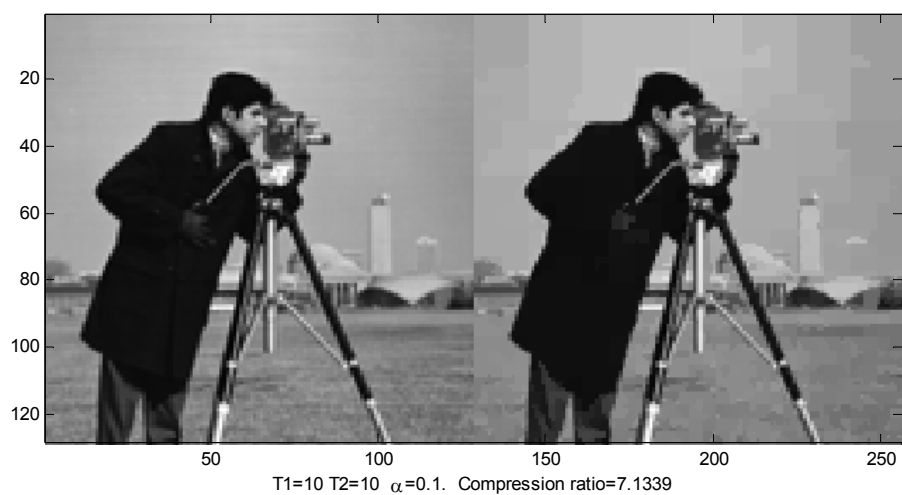
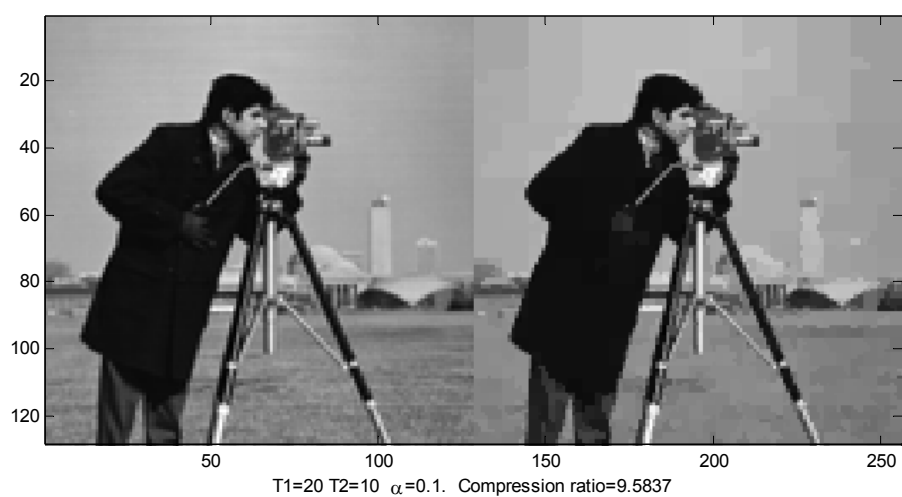
$$|M - x[k]| \leq T_1 \cdot (\exp\{-\alpha(L-1)\}) + T_2$$

כך שבאורך $L=1$ הסף הוא $T_0 = T_1 + T_2$, ועם גדילת הקבוצה הסף דועך ל- T_2 על מנת לעודד פחות הצטרפותם של פיקסלים חדשים.

מדוע סריקת Peano-Hilbert ולא Raster-Scan פשוט? ניתן כמובן לאמץ סריקה פשוטה יותר אך הפגמים בתמונה יהיו בולטים יותר בשיטה בה ישנו סדר מבני בסריקה. בגישת ההילברט ישנו ערבול של השגיאות במישור המקום באופן שאינו קופץ לעין כצורם ומכאן הבחירה הסריקה זו.

ברור כי ניתן להכליל את האלגוריתם הנ"ל לייצוג כללי יותר של הקבוצה. במקום ערך קבוע נוכל לחשוב על פולינום מסדר p אשר יותאם אל הנתונים, ופיקסל יצורף אם הצטרפותו לא גורמת לשגיאה גדולה מידי. במקרה כזה יש לשדר את אורך הקבוצות ואת $p+1$ פרמטרי הפולינום. באשר לדחיסה עצמה, ראוי לבצע קידוד אנטרופיה לאורכים ולגבהים בשיטת הקידוד הנ"ל על מנת לסחוט דחיסה עמוקה יותר לאותה איכות מוצא. קוד הופמן המתוכנן מראש על מקבץ תמונות שעברו את תהליך חלוקה לאינטרוולים כנ"ל ייתן ביצועים ראויים. נראה כי שימוש בהופמן רצוי במיוחד לאורכי האינטרוולים אשר אמורים לתת פילוג אקספוננציאלי (אורכים קצרים סבירים הרבה יותר מאשר ארוכים), בעוד הממוצעים צפויים לתת פילוג אחיד – דבר שמעודד הקצאה קבועה של 8 סיביות לכל ערך.

בציורים הבאים מתוארת תוצאת האלגוריתם למספר בחירות של הפרמטרים. יחס הדחיסה נקבע לפי כמות הסיביות שבה נעשה שימוש בפועל (חישוב אנטרופיה על האורכים והקצאת 8 סיביות לממוצעים) מול הכמות המקורית של הסיביות - 8×128^2 . ניכר כי האלגוריתם משיג איכות לא רעה ליחסי דחיסה של 5-10, ובכך מתחרה עם הטובים שבאלגוריתמי הדחיסה. יש

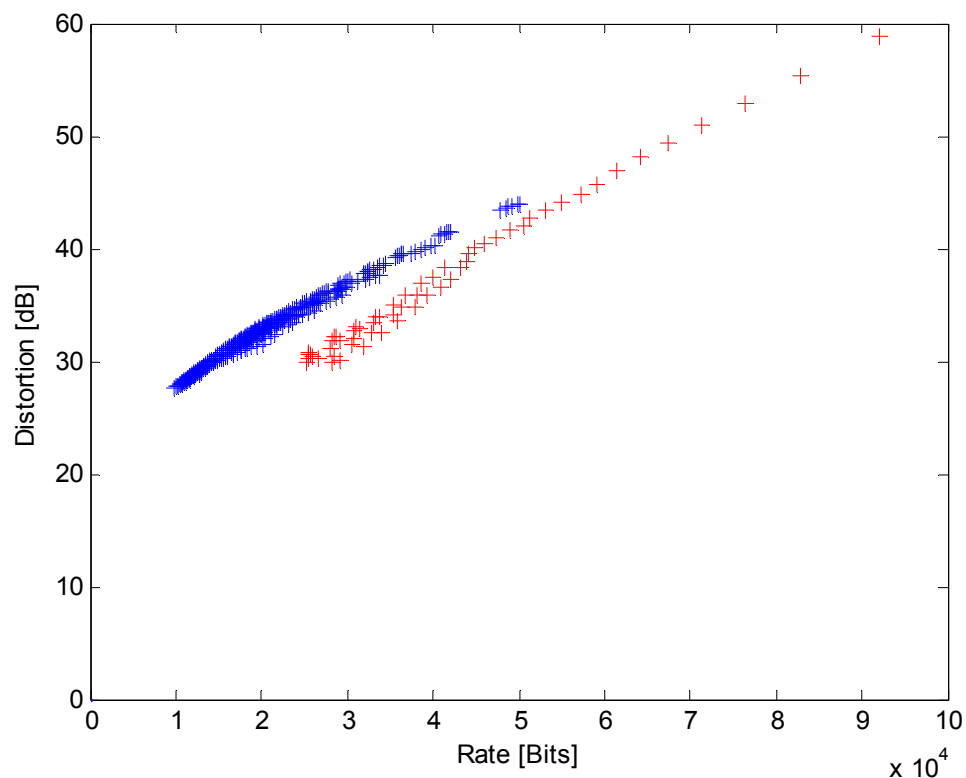


תוצאות דחיסה באלגוריתם סורק-זאבי

לזכור, שוב, כי כל זה קורה באמצעות אלגוריתם פשוט עד טריוויאלי.

נקודה מעניינת שניתן להעלות כאן היא העובדה שבנקל ניתן להציע אלגוריתם אשר יקבל את התמונה הפרוסה ע"י אלגוריתם סורק-זאבי ויבצע עליה שיפור. לשם דוגמה, נניח כי שני פיקסלים המהווים שכנים ב-2D נפרדו בשל הסריקה לחד-ממד. בחינת ערכיהם יכולה להראות כי הם קרובים אך לא מידי, וזאת בשל היותם משויכים לאינטרוולים שונים באלגוריתם הדחיסה. טשטוש ערכיהם (קירוב הערכים זה לזה) ייתן איכות ויזואלית טובה יותר. מעניין לציין כי באזורי פרטים משכיל האלגוריתם לתת דיוק טוב ולהתמקד בעיקר.

עקומת ה- $R(D)$ המתייחסת לאלגוריתם סורק-זאבי מתוארת בציור הבא למקבץ גדול של בחירת פרמטרים. אנו רואים כאן השוואה לשיטת ה-DPCM (באדום) ויתרון מובהק על פניה בשיעור ממוצע של 4-5dB. עם זאת, ניתן להניח כי פער זה יצטמצם עם שימוש בחיזוי טוב יותר (כאן נשעה שימוש בפיקסל קודם יחיד). מעבר לכך, שיטת הסריקה כאן טובה יותר ויכולה היתה להיות מופעלת גם ל-DPCM.



עקומת ה- $R(D)$ של אלגוריתם סורק-זאבי בכחול ומולה ביצועי ה-DPCM.

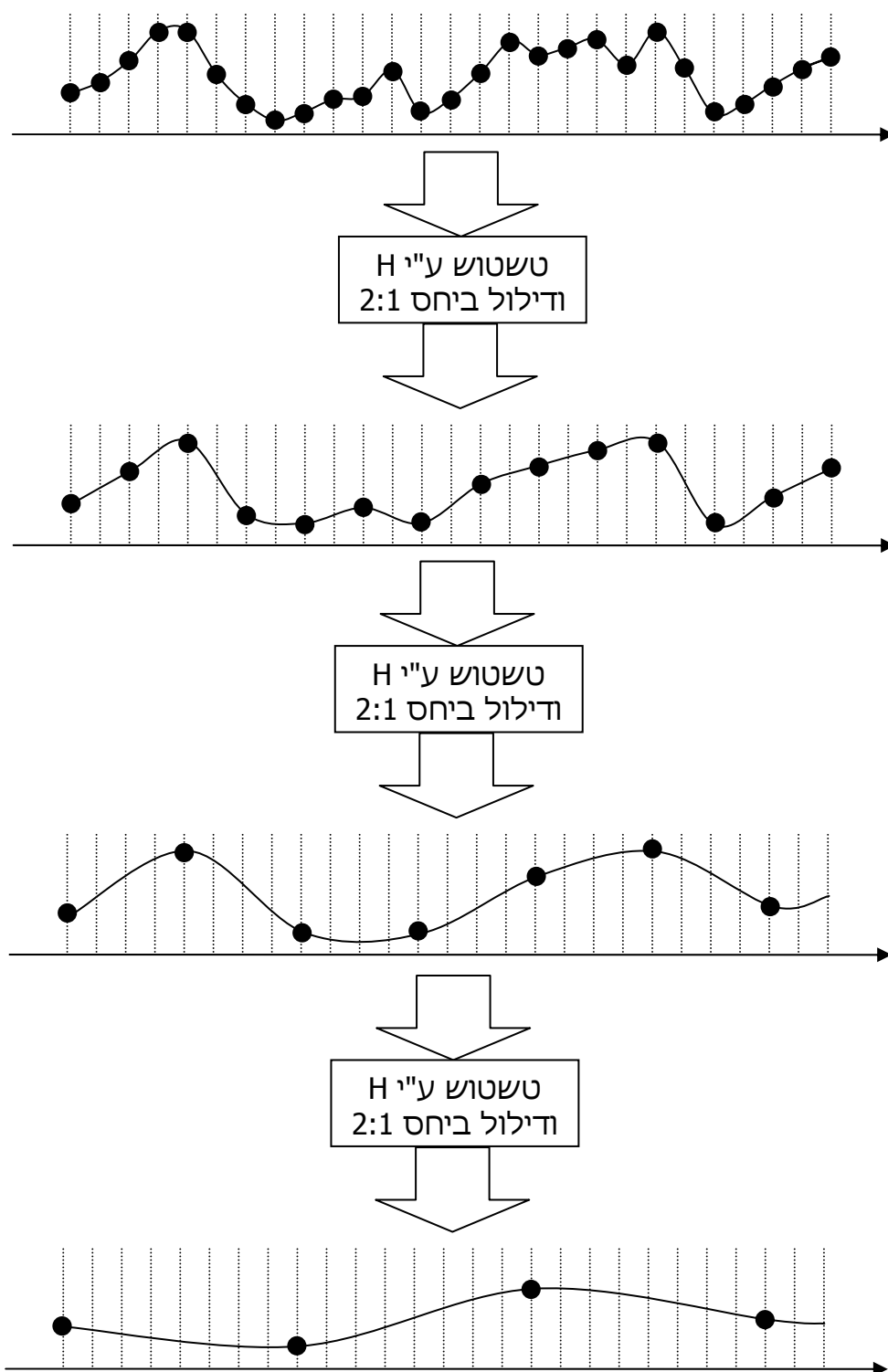
הפירמידה הגאוסית לדחיסה מבוססת חיזוי

מחלה משותפת לשתי הגישות הקודמות היא הצורך להמיר תמונה – אות דו-ממדי – לווקטור על מנת לאפשר יישום החיזוי או החלוקה לאינטרוולים. האם יש דרך להציע סכימת חיזוי דו-ממדית טבעית בה החיזוי מותר מכל כיוון? מסתבר כי גם גישה כזו קיימת, והיא הוצע בתחילת שנות ה-80. גישה זו מוכרת בשם השיטה הפירמידלית. נתחיל בתיאור מהיר של קונספט הפירמידה, ומשם נעבוד לתיאור אלגוריתם הדחיסה.

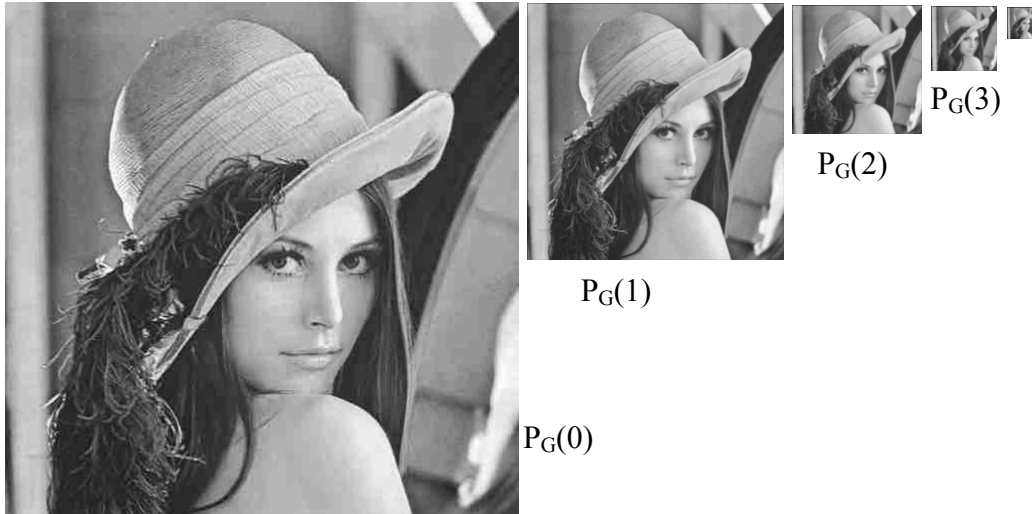
ככלל, פירמידה הינה אוסף תמונות המאורגנות בסדר כלשהו (נציע סידור אנכי המתחיל למטה ועולה למעלה) ואשר כולן נובעות מתמונה אחת שעברה עיבוד. נניח כי נתונה לנו תמונה $X(m,n)$ בגודל N על N . נגדיר תמונה זו כתחתיתה של הפירמידה P_G ונסמך אותה כ- $P_G(0)$. נפעיל על תמונה זו מסנן LPF מסוג כלשהו ונקבל תמונה בה הפרטים מטושטשים קמעה. התמונה החדשה, $P_G(1)$, תהיה השכבה השנייה בפירמידה. נמשיך ונטשטש את התמונה $P_G(1)$ ונקבל את $P_G(2)$ וכך הלאה. אוסף התמונות שתתקבלנה יהיה הפירמידה הרצויה, כאשר כל תמונה מייצגת רמת רזולוציה מסוימת.

נניח כי מסנן ה- LPF הוא כזה שמדכה בצורה סבירה (אין כל הכרח לדיוק כאן כי אין אובדן מידע – התמונה $P_G(0)$ היא תמונת המקור) תדרים גבוהים מעבר ל- $\pi/2$ (מחצית טווח התדרים). פירוש הדבר הוא שהתמונה לאחר סינון יכולה לעבור דילול ביחס 2:1 ללא אובדן. לכן, יחד עם פעולות הסינון תבואנה פעולות דילול והתמונות תקטנה. הציור הבא ממחיש את המבנה המתקבל. והדרך לבנייתו בתיאור חד-ממדי.

מהו המסנן בו יש לעשות שימוש? ישנן מגוון גישות לבחירת המסנן. לדוגמה, נוכל לחפש בחירה שתבטיח דיכוי תדרים גבוהים כיאות על-מנת שהדילול יהיה מותר ללא אובדן בקירוב טוב. בעבודתם הציעו Burt ו- Adelson גישה שונה – הם הציעו את הכללים הבאים אותם מסנן זה יקיים, בהנחה שזהו מסנן ספרבילי באורך $2N+1$ שתגובת ההלם שלו היא $\{w_k\}_k$



בניית הפירמידה הגאוסית לאות חד-מימדי.

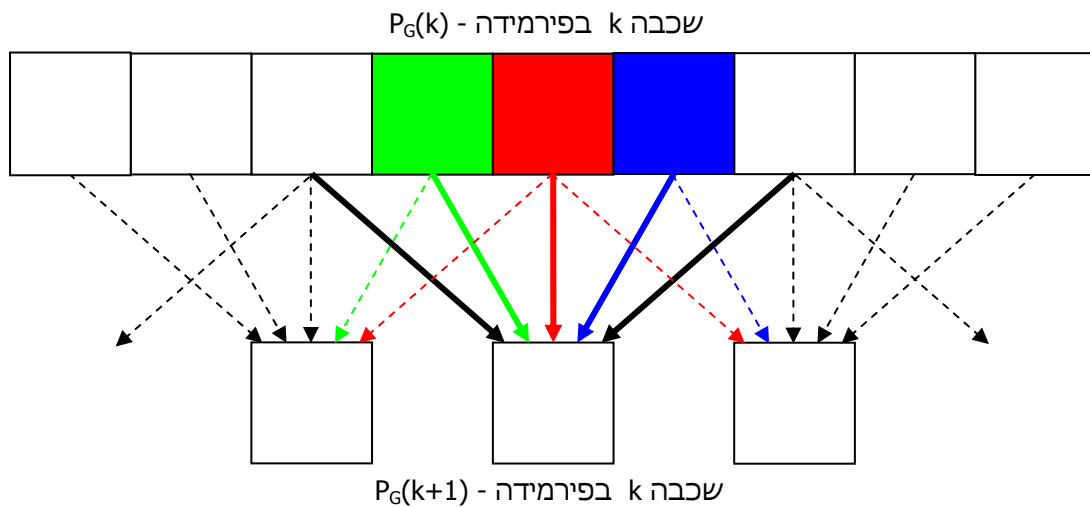


הפירמידה הגאוסית. התמונה השמאלית היא בסיס הפירמידה, כל תמונה אחרת נוצרת מטשטוש ודילול של קודמתה.

$$\begin{aligned}
 1. \sum_{k=-N}^N w_k &= 1, & 2. \forall k > 0, w_k &= w_{-k} \\
 3. \forall k > 0, w_k &\leq w_{k-1}, & 4. \forall k, \sum_{j=-\infty}^{\infty} w_{k+2j} &= \text{const}
 \end{aligned}$$

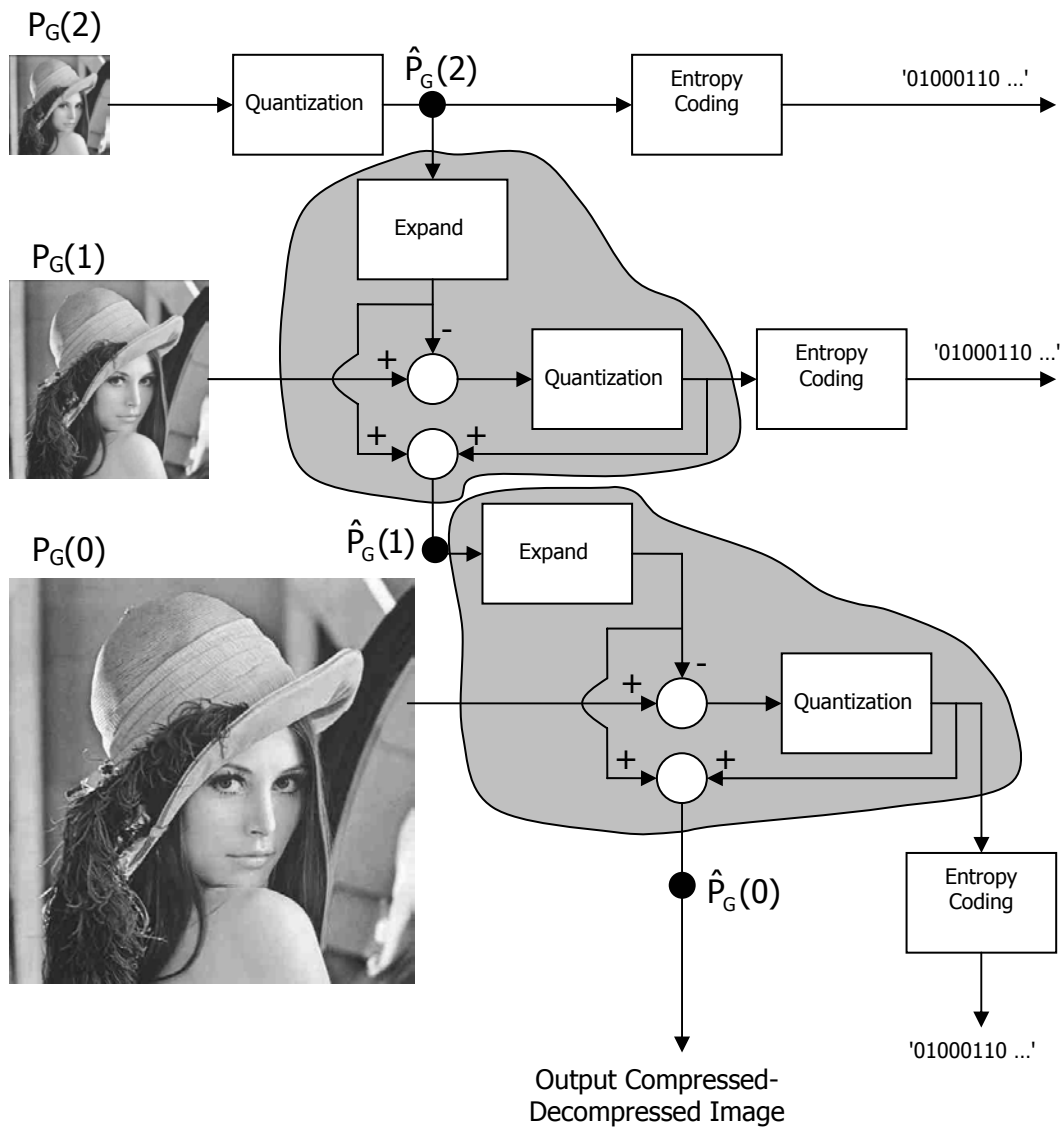
הדרישה הראשונה מבטיחה כי הפעלת מסנן זה על התמונה תשמר את הממוצע המקומי. הדרישה השנייה יוצרת מסנן סימטרי בו אין הטיה לצד אחד על פני האחר. הדרישה השלישית באה לוודא דעיכה מונוטונית של השפעת השכנים על הפיקסל המרכזי בסינון. הדרישה הרביעית והמורכבת מכולן פירושה שכל פיקסל תורם בצורה זהה ביצירת שכבת הפירמידה הבאה המדוללת. הציור הבא ממחיש עניין זה ומצדיק את ניסוח הדרישה כפי שנקבע במשוואה.

לשם דוגמה, אם נניח כי $N=2$ ונניח כי המסנן בנוי כ- $\{c \ b \ a \ b \ c\}$, הדרישה הראשונה תהיה $2c + 2b + a = 1$. הדרישה השנייה מוטמעת כבר בייצוג הפרמטרי שבחרנו, והדרישה השלישית תהיה $c \leq b \leq a$. הדרישה הרביעית תתורגם ל- $2c + a = 2b$. לכן ברור כי $b=0.25$ (צירוף דרישה ראשונה ורביעית), a צריך בהכרח להיות גדול מ- 0.25 (דרישה שלישית), ו- c יהיה $c=0.5(a-0.25)$. עבור הבחירה $a=0.4$ מתקבל מסנן דומה לפעמון גאואסי ומכאן השם "הפירמידה הגאוסית".



במיצוע משוקלל אנו רואים כיצד שכבה בפירמידה יוצרת את הבאה לה. אם נתייחס לפיקסל בודד בשיכבה ה- k , ערכו תורם לשכבה הבאה ב-3 מקומות לפי המשקל $a+2c$. באופן דומה, הפיקסלים הכחול והירוק מפעפעים לשכבה הבאה דרך משקל מצטבר $2b$.

הפירמידה הגאוסיאנית יכולה להיות הבסיס לדחיסת תמונה בגישה חזוית לא סיבתית. אם נתחיל בקצה הפירמידה העליון יש בידינו תמונה קטנה מאוד אותה נוכל לשדר כמות שהיא. זו תהווה חזוית לגרסתה המוגדלת המהווה שכבה מתחתיה בפירמידה. לכן, מהות החזוית כאן תהיה הגדלת התמונה הידועה בפקטור 2 בכל ציר תוך ניסיון לעשות זאת בצורה הכי נאמנה לתמונה הבאה. בהנחה שהחזוית מוצלח, שגיאת החזוית – תמונה של ערכי הפרשים – תהיה בעלת ערכים קטנים בעלי אנטרופיה נמוכה. נוכל לקודדם עם שגיאות ע"י ביצוע קוונטיזציה, ממש כשם שעשינו זאת בסכימת ה-DPCM. בניית החזוית לשכבה הבאה חייב להיעשות עם השגיאה לאחר קוונטיזציה, כך שמקודד ומפענח משתמשים באותו חזוית (שוב, ממש כמו ב-DPCM). התהליך יימשך הלאה עד להגיענו לבסיס הפירמידה. חשוב להבהיר כי אין כל חשיבות לסיבתיות בעת קידוד כל שכבת שגיאה. הסיבתיות הפעם נדרשת להיות סדר המעבר מהשכבה הגסה אל העדינה בפירמידה, ובזאת אנו בעצם מתקדמים מחזלוציה נמוכה לגבוהה. הציורים הבאים מראים סכימת בלוקים של אלגוריתם הקידוד המוצע, ומספר תוצאות אופייניות לאלגוריתם זה.

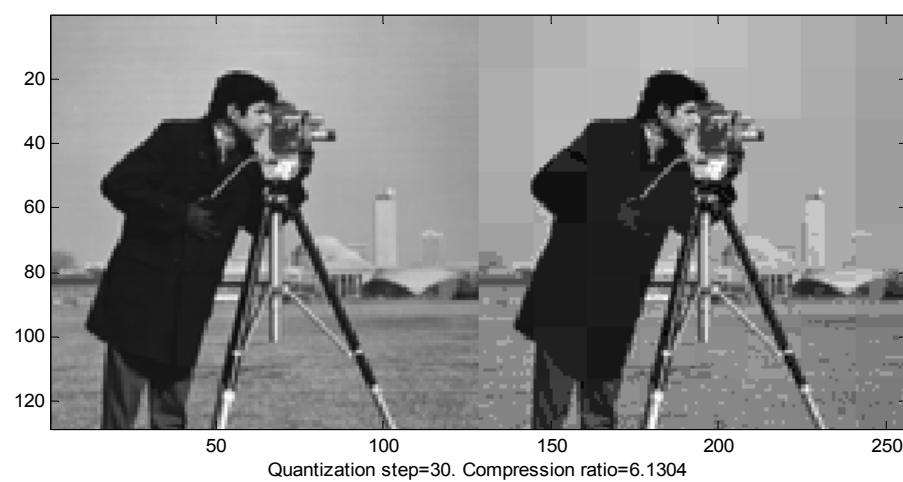
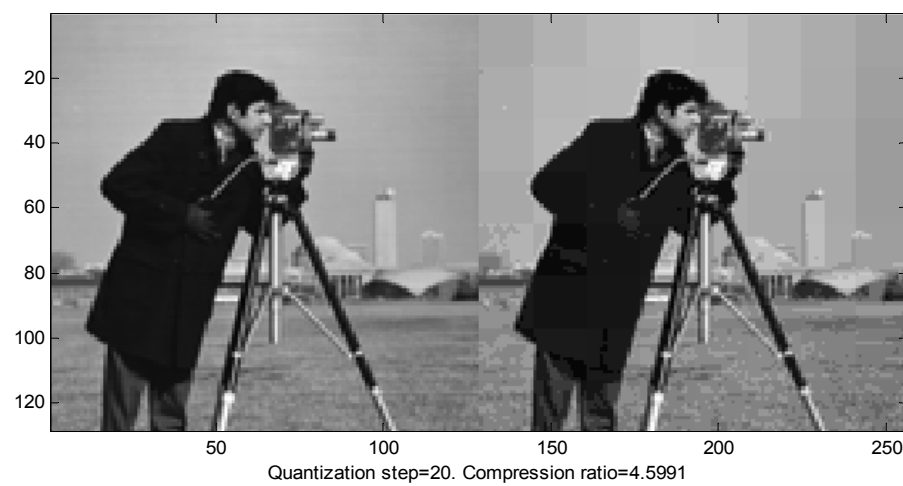
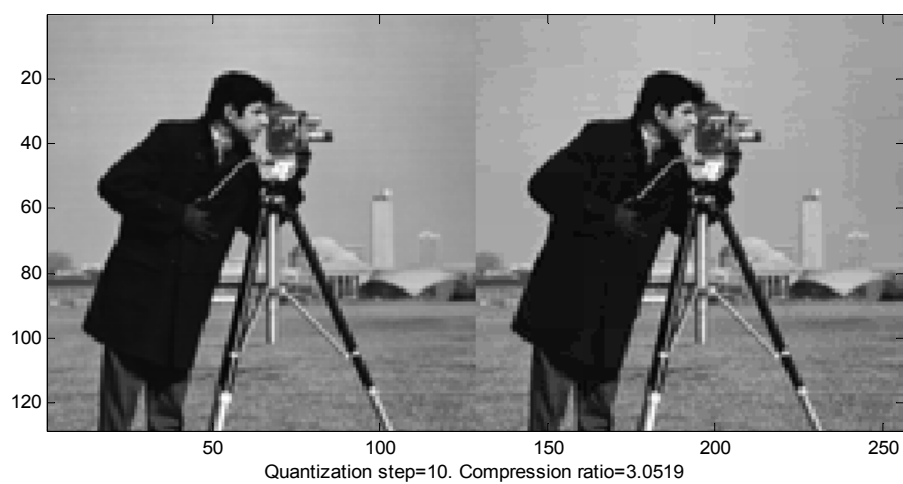


סכימת בלוקים של תהליך הדחיסה בפירמידה גאוסיאנית.

בגישת הפירמידה הגאוסית בניית כל שכבה נעשית על סמך קודמתה ע"י המשוואה

$$\hat{P}_G\{k-1\} = Q\{P_G\{k-1\} - \text{Expand}\{\hat{P}_G\{k\}\}\} + \text{Expand}\{\hat{P}_G\{k\}\}$$

במשוואה זו התוצאה עד כה כפי שמוחזקת במקלט היא $\hat{P}_G\{k\}$. יש להגדילה ע"י פעולת Expand לשם בניית חיזוי, חישוב שגיאת החיזוי וביצוע קוונטיזציה על תמונת הפרש זו, ולבסוף חיבור החיזוי הנ"ל חזרה. התוצאה, $\hat{P}_G\{k-1\}$ הינה התמונה הגדולה פי 2 בכל ציר כפי שתתקבל במפענח.



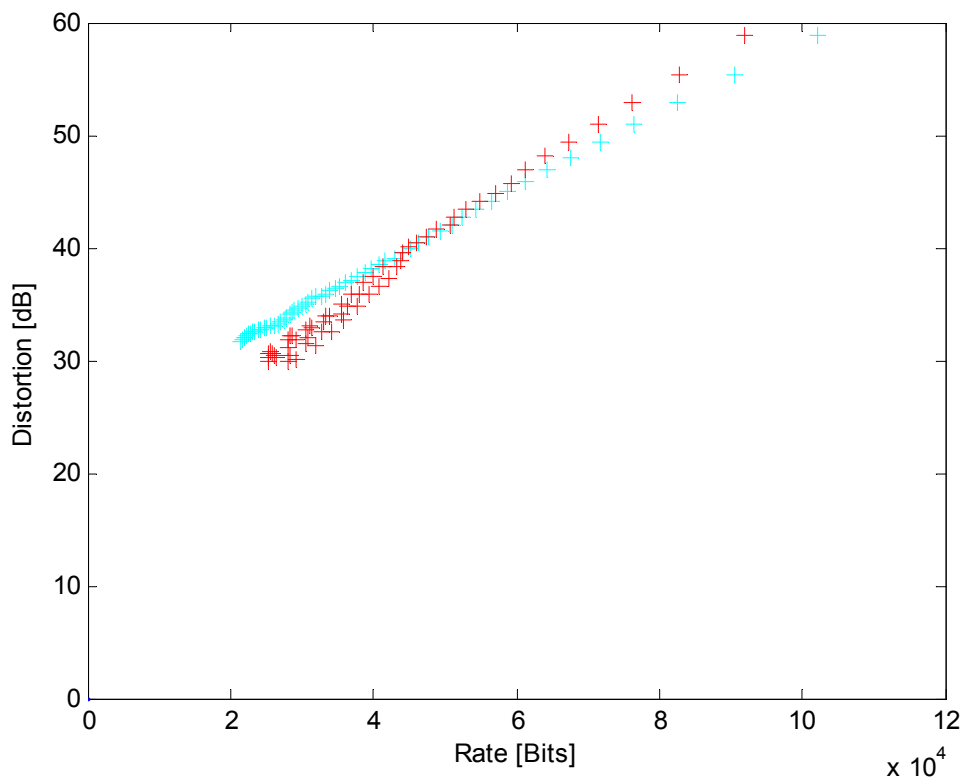
תוצאות דחיסה באלגוריתם הפירמידה הגאוסית

נשים לב כי כמות הערכים שעלינו לשדר אינה N^2 אלא יותר. רק השכבה התחתונה בפירמידה כוללת N^2 ערכי שגיאות חיזוי, וישנן כל השכבות הקודמות. הכמות בפועל הינה

$$\sum_{k=0}^L N^2 \cdot 4^{-k} = N^2 \cdot \frac{1 - 4^{-L}}{1 - 0.25} < \frac{4}{3} N^2$$

לכן ישנו ניפוח בשיעור של 33% בכמות המידע בטרם נתחיל בדחיסה, וזאת בשל המבנה הפירמידלי.

בציור הבא אנו מראים את עקומת ה- $R(D)$ עבור אלגוריתם זה כשאנו מבצעים קוונטיזציה יוניפורמית וזהה לכל שכבות הפירמידה, החיזוי נעשה בשיטת השכן הקרוב שהינה גישה גרועה למדי (בשלב ההגדלה כל פיקסל משוכפל להיות בלוק של 2-על-2 פיקסלים בעלי ערך זהה). קידוד האנטרופיה נעשה כאן ע"י חישוב ישיר של האנטרופיה כשגם כאן החישוב נעשה על כל השכבות ביחד - דבר שגוי כי אנו צריכים לצפות לסטטיסטיקות אחרות משכבה לשכבה.



עקומת $R(D)$ עבור שיטת הפירמידה הגאוסית (תכלכל) מול זו של ה- DPCM (אדום)

אנו רואים כי הביצועים משתווים פחות או יותר לאלו של ה-DPCM, עם ניצחון קלוש באזור הקצבים הנמוכים. שתי סכמות הדחיסה הללו יכולות להשתפר, ולמעשה, בספרות המקצועית מקובל לטעון כי גישת הפירמידה הגאוסית שקולה פחות או יותר לזו של ה-DPCM. לצד הרווח שבחיזוי מבוסס דו-ממד ולא סריקה חד-ממדית, ישנו ההפסד של הניפוח בשיעור של 33% בכמות המידע.

הפירמידה הלפלסיאנית לדחיסה מבוססת חיזוי

ישנה גישה פירמידלית טובה יותר המוכרת בשם פירמידה לפלסיאנית, וזו בדרך כלל מובילה לביצועי דחיסה טובים יותר. נראה כאן את עיקרי הדברים ונסתפק בזאת. הפירמידה הלפלסיאנית נבנית בעזרת שתי פעולות יסוד – Reduce ו-Expand. פעולת ה-Reduce מקבלת תמונה ומחזירה את גרסתה המוקטנת בפקטור 2 בכל ציר בדרך כלשהי (טשטוש ודילול כלשהם, בדרך-כלל). פעולת ה-Expand מקבלת תמונה ומחזירה את גרסתה המוגדלת בפקטור 2 בכל ציר (אינטרפולציה כלשהי). ניכר כי שתי פעולות אלה כבר כיכבו בבניית הפירמידה הגאוסית. הפירמידה החדשה תיבנה ע"י סידרת הפעולות הבאות:

1. התחלִי מהתמונה X . זוהי גם $P_G(0)$ - שכבת הפירמידה הגאוסית התחתונה,

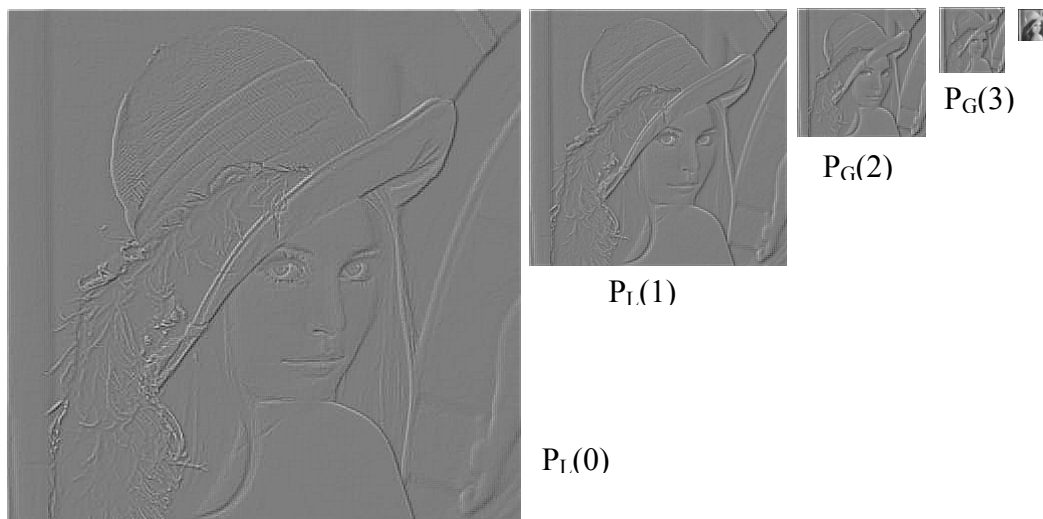
וקבע את k ל-1.

2. בצעד k (ה- $k \geq 2$), בצעִי פעולת Reduce על $P_G(k-1)$ וקבלִי את התמונה $P_G(k)$ - שכבת הפירמידה הגאוסית ה- k .

3. בצע פעולת Expand על $P_G(k)$ והגדר $\{P_L(k)\}$ $P_L(k-1) = P_G(k-1) - \text{Expand}\{P_G(k)\}$.

4. הגדל את k . אם הגענו לשכבה $k=M$ קבע $P_L(k) = P_G(k)$ ועצור. אחרת חזור לשלב 3.

בבניה זו של הפירמידה, בניית התמונה המקורית פשוטה ותיעשה בצורה רקורסיבית ע"י הפעלת Expand על התמונה בראש הפירמידה $P_L(M)$, הוספת השכבה ה- $M-1$, שוב Expand וחזור עד להגעה לתחתית הפירמידה הלפלסיאנית, כשהתמונה הנוצרת זהה במדויק לזו ממנה התחלנו. הציורים הבאים מראים את שכבות הפירמידה ודרך בניית הפירמידה ושחזור התמונה ממנה חזרה.

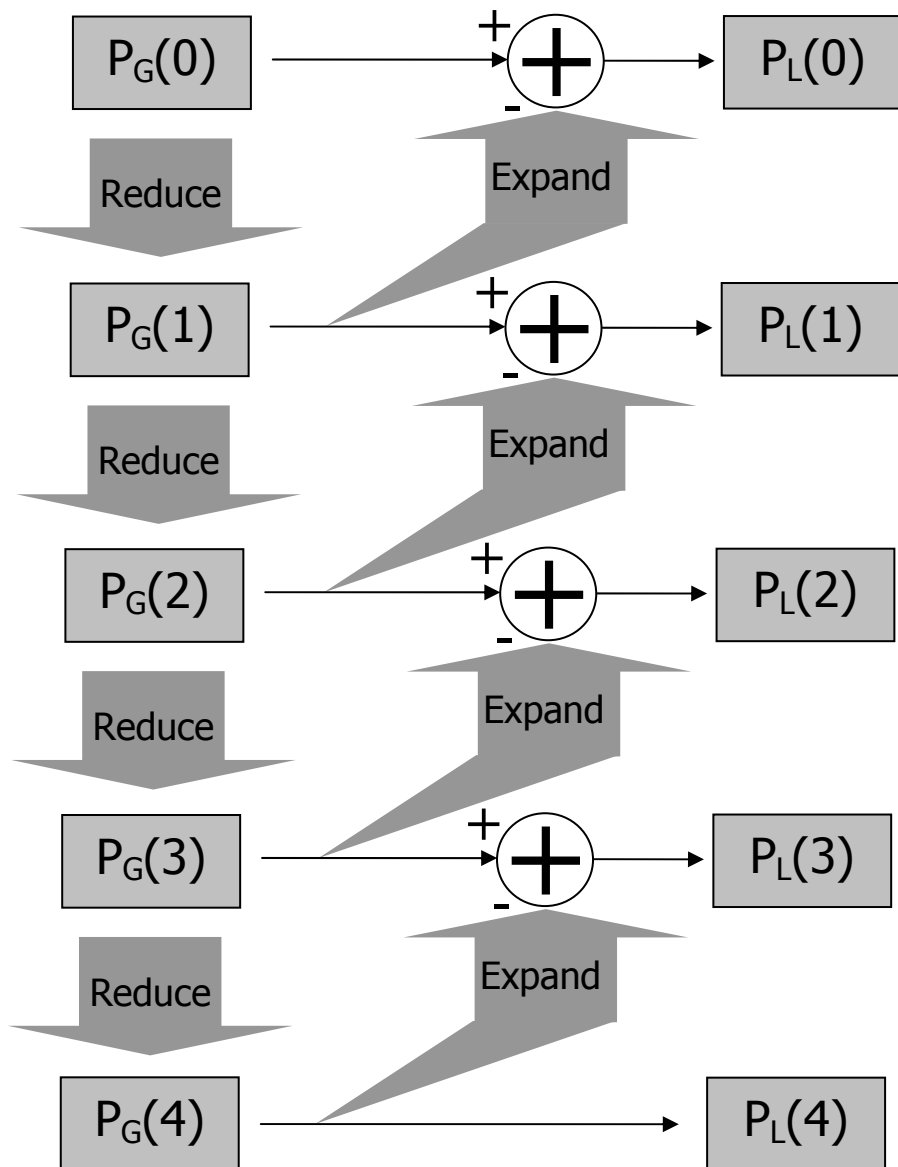


שכבות הפירמידה הלפלסיאנית (לערכי הפרשים הוסף 127).

נקודה חשובה היא הנקודה הבאה – לא חשוב כיצד נבחר את פעולות ה-Reduce וה-Expand. הסכימה הבונה את הפירמידה ניתנת להיפוך ובניית התמונה המקורית לכל בחירה חוקית (מבחינת משמעות הפעולות וגדלי הכניסה\יציאה המתחייבים) של פעולות אלה. ברור כי רצוי לבחור פעולות חכמות כך שהחיזוי שנוצר מפעולת ה-Expand יהיה מוצלח. בציור הקודם בנינו את הפירמידה על ידי פעולת Pixel Replication בו בעליה ברזולוציה כל פיקסל מייצר רביעיית ערכים זהים לו. זוהי קרוב לוודאי הגישה הפשוטה ביותר האפשרית וגם הגרועה ביותר.

נשים לב כי פירמידה זו בנתה עבורנו את שכבות שגיאת החיזוי בדרך טבעית. ובאמת, במקור הוצעה הפירמידה הלפלסיאנית כפירוק של תמונה לשם דחיסתה – עבודתם של Burt ו-Adelson. הרעיון פשוט - בניית הפירמידה לתמונה, ואז ביצוע קוונטיזציה לשכבותיה כך שערכים קרובים לאפס מאופסים, וכל ערך אחר מומר לערכי ייצוג קלים לדחיסה. ניתן אף לחשוב על קוונטיזר שונה לכל שכבה.

בציורים הבאים אנו מראים כיצד נבנית הפירמידה הלפלסיאנית וכיצד ממנה חוזרים לתמונת המקור. לצורך דחיסה הבניה נותרת כמתואר, והשחזור זהה למעט הכנסת קוונטיזציה וקידוד אנטרופיה לכל שכבה $P_L(k)$ בטרם השימוש בה.



בניית הפירמידה הלפלסיאנית והקשר שלה לפירמידה הגאוסית.

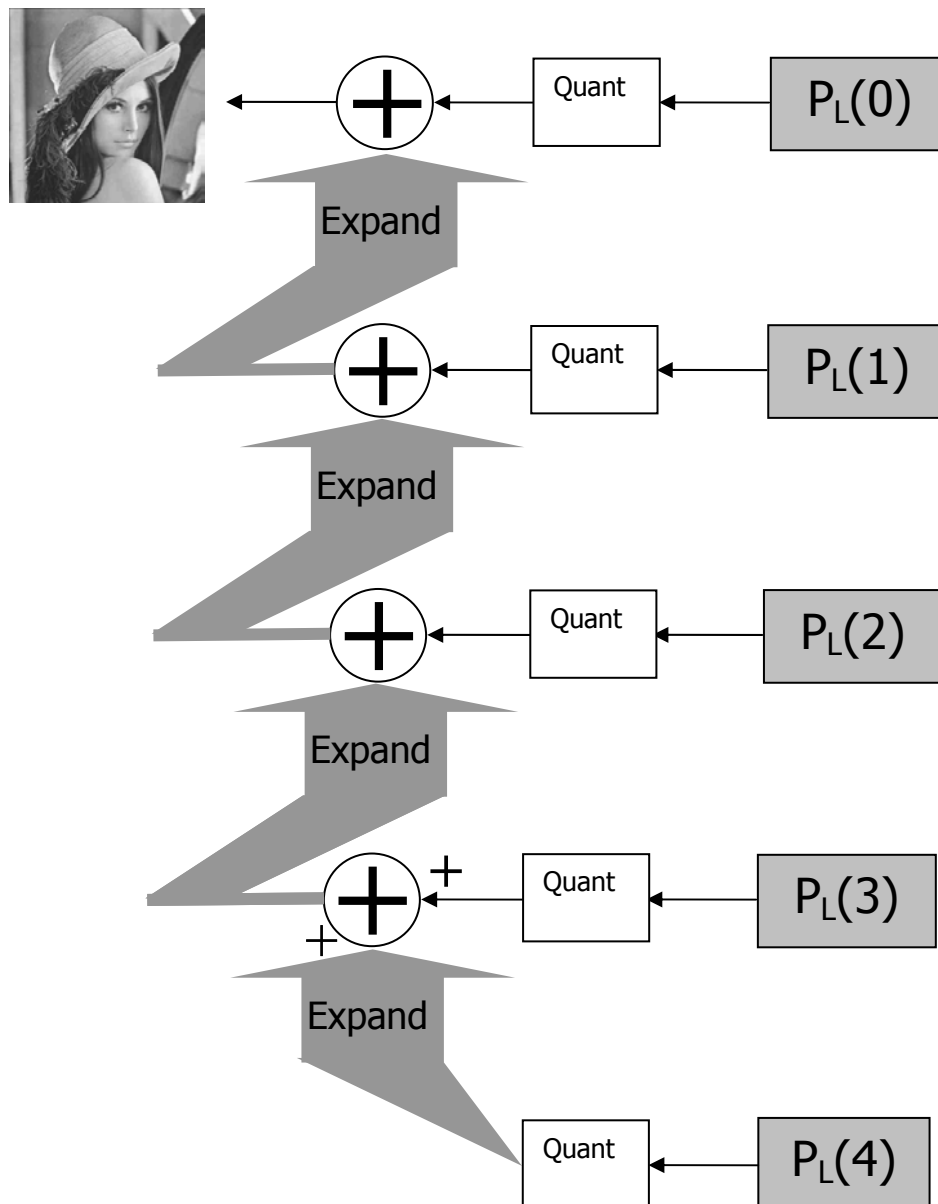
בפועל ביצועי גישה זו טובים במקצת מאלו של הפירמידה הגאוסית. הסיבה לכך היא הבאה - בפירמידה הלפלסיאנית נוסחת העדכון של תמונת המוצא שונה במקצת. זו תהיה:

$$\hat{P}_G\{k-1\} = \text{Expand}\{\hat{P}_G\{k\}\} + Q\{P_L\{k\}\}$$

כלומר, שכבת השגיאה מהפירמידה הלפלסיאנית עוברת קוונטיזציה וזו מצורפת לגרסה המוגדלת של התמונה הקודמת כפי שהוחזקה במפענח. משוואה זו ניתנת לרישום גם כ-

$$\begin{aligned}\hat{P}_G\{k-1\} &= \text{Expand}\{\hat{P}_G\{k\}\} + Q\{P_L\{k\}\} = \\ &= \text{Expand}\{\hat{P}_G\{k\}\} + Q\{P_G\{k-1\} - \text{Expand}\{P_G\{k\}\}\}\end{aligned}$$

כשהשתמשנו בהגדרת שכבות הפירמידה הלפלסיאנית.



בניית התמונה מהפירמידה הלפלסיאנית.

נזכור כי המשוואה המקבילה עבור הפירמידה הגאוסיאנית היא

$$\hat{P}_G\{k-1\} = Q\{P_G\{k-1\} - \text{Expand}\{\hat{P}_G\{k\}\}\} + \text{Expand}\{\hat{P}_G\{k\}\}$$

רואים כי ישנו הבדל עדין מאוד בין שתי הגישות – זו של הפירמידה הלא-גאוסיאנית מבצעת קוונטיזציה על שגיאת החיזוי האמיתית בין שכבות הפירמידה הגאוסית, בעוד שבפירמידה הגאוסית הקוונטיזציה היא על השגיאה כשהתמונה המפוענחת מעורבת. לכן השגיאה גדולה יותר והביצועים ירודים יותר.

בציור הבא אנו רואים תוצאות שתי גישות הפירמידה כשהפעם פעולת ה-expand הינה אינטרפולציה ביליניארית, ומכאן השיפור המשמעותי בהשוואה לתוצאות שהוצגו קודם לכן.



Quantization step=10. Compression ratio=2.9355



Quantization step=10. Compression ratio=3.0378

דחיסה עם הפירמידה הגאוסית (מקור משמאל ותוצאה מימין) עם פעולת expand מבוססת אינטרפולציה ביליניארית. מולה, למטה, תוצאה מקבילה של הפירמידה הלא-גאוסיאנית.

עיבוד תמונות בעזרת מחשב (236327)

הרצאה מס' 12

התמרות – גישה אלגברית

כבר פגשנו במסגרת דיון קודם את התמרת הפורייה (רציף ודיסקרטי), ואנו שבים במסגרת פרק זה לנושא זה בשל חשיבותו הרבה. הרעיון הבסיסי שילווה אותנו לכל אורכו של הדיון כאן הוא שבהינתן תמונה בגודל N על M כמטריצת ערכי רמות אפור, מבלי ששמנו לב נכפה עלינו ייצוג מאוד מסוים של התמונה מעל הבסיס הטריגונומלי $e_{[j,k]}$ (תמונה שכולה אפס למעט האיבר ה- $[j,k]$ השווה 1). התמונה מתוארת כצירוף ליניארי של פונקציות בסיס טריגונומליות אלה, ולמה שבחירה זו תחשב טובה יותר מאחרות? מדוע לדבוק בה, למעט פשטותה? אנו נראה כי אמנם קיימות התמרות לבסיסים אחרים, עימם ניתן לפשט בעיות, לחסוך בחישובים ועוד.

עוד עובדה שתרחף מעל רוב פרק זה כאקסיומה היא שאנו מעוניינים לעבוד עם התמרות ליניאריות בלבד, בהן התמונה תיוצג כצירוף ליניארי של תמונות בסיס כלשהו. ברור כי בחירה זו מצמצמת באופן דרסטי את מקבץ האפשרויות, אך מאידך, מבטיחה לנו טיפול מתמטי קל יותר. אנו נתחיל את הדיון בבניית הקשר בין בחירת בסיס אלטרנטיבי והתמרה ליניארית.

התמרות ליניאריות - בסיס

נתחיל את הדיון באותות חד ממדיים באורך N . התמרה ליניארית קונוונציונלית הינה פעולה הניתנת לתיאור כמכפלה במטריצה. ההתמרה $\underline{U}$ נתונה ע"י:

$$\underline{U} = A\underline{V}$$

כשאנו מניחים ש- A מטריצה ריבועית בגודל $[N \times N]$. במקרה זה נקבל כי וקטור ההתמרה זהה בגודלו לוקטור המקור. אנו נדרוש כי A תהיה הפיכה על מנת

לאפשר התמרה הפוכה מ- $\underline{u}$ ל- $\underline{v}$. במקרים רבים אנו נלך צעד נוסף במגמת הפשטת ההתמרה ונשתמש רק בהתמרות שביצוען קדימה-אחורה קל עד טריוויאלי - התמרות יוניטריות, עבורן מתקיים:

$$\underline{A}^{-1} = \underline{A}^H = [\underline{A}^T]^*$$

הזכרנו כי התמרה מהווה ייצוג עפ"י בסיס חדש. מהו הקשר בין בסיס חדש זה והמטריצה הנבחרת $\underline{A}$? נניח כי נתון לנו סט אותות בסיס $\underline{b}_k$ איתו אנו רוצים לעבוד. נניח בשלב זה כי מספר פונקציות הבסיס הוא בדיוק N (כמספר הדגימות בוקטור המקור $\underline{v}$), וכן נניח כי בסיס זה פורס את המרחב. לכן נייצג את $\underline{v}$ כצירוף ליניארי של פונקציות בסיס אלה ונקבל:

$$\underline{v} = \sum_{k=1}^N u_k \underline{b}_k = \underline{B} \underline{u}$$

כאשר $\underline{B}$ היא מטריצה המכילה את $\underline{b}_k$ כעמודותיה, והווקטור $\underline{u}$ הוא מקדמי הצירוף הליניארי. כיוון שהבסיס שלם (פורס את המרחב) ובעל N אלמנטים, קיים רק צירוף ליניארי אחד שייתן שיויון. צירוף זה מתקבל ע"י:

$$\underline{u} = \underline{B}^{-1} \underline{v} = \underline{A} \underline{v}$$

ומכיוון שהבסיס שלם נקבל כי המטריצה $\underline{B}$ לא סינגולרית והפיכה. למעשה, קיבלנו כי ההתמרה שיש לבצע על $\underline{v}$ לקבלת $\underline{u}$ היא ההיפוך של $\underline{B}$, כלומר $\underline{B}$ מייצג את ההתמרה ההפוכה. קשר פשוט יותר יתקבל אם נעשה שימוש בוקטורים $\underline{b}_k$ אורתונורמליים, כלומר כאלה שמקיימים

$$\forall k, j \quad \underline{b}_k^T \underline{b}_j = \delta(k - j)$$

אז יתקבל כי המטריצה $\underline{B}$ יוניטרית, ולכן

$$\underline{u} = \underline{B}^{-1} \underline{v} = \underline{B}^H \underline{v} = \underline{A} \underline{v}$$

ויצא כי שורותיה של המטריצה $\underline{A}$ מכילות (לאחר פעולת צמוד קומפלקסי) את הוקטורים המהווים את הבסיס החדש.

להתמרה יוניטרית תכונה חשובה אותה אנו מכירים בשם "משפט פרסבל" עבור המקרה הפרטי של התמרת פורייה. אם נחשוב לרגע על התמרה יוניטרית כפעולה גיאומטרית המקבלת וקטור ויוצרת ממנו וקטור חדש, אזי, התמרה יוניטרית לא משנה את עוצמתו (הנמדדת בנורמה) של הוקטור בתהליך, אלא פועלת כסיבוב ברב מימד. קל לראות זאת אלגברית ע"י הקשר:

$$\underline{U} = A\underline{V} \Rightarrow \|\underline{U}\|^2 = \|A\underline{V}\|^2 = \underline{V}^H A^H A \underline{V} \underset{A^H A = I}{=} \underline{V}^H \underline{V} = \|\underline{V}\|^2$$

במקרים מסוימים יש עניין בהתמרות בהן יש יותר פונקציות בסיס מאשר נקודות בתמונה. בסיסים אלו נקראים Over-Complete. במקרה כזה, המטריצה B המכילה את וקטורי בסיס אלה כעמודותיה, ולכן יש לה יותר עמודות משורות (מספר העמודות L גדול מ-N), ולכן, הווקטור $\underline{U}$ יהיה ארוך יותר מ- $\underline{V}$. במקרה כזה הקשר המתאר את ההתמרה ההפוכה הוא:

$$\underline{V} = \sum_{k=1}^L u_k \underline{b}_k = B\underline{U}$$

ההתמרה - תהליך חישוב $\underline{U}$ מתוך $\underline{V}$ - מורכבת יותר הפעם. כיוון שזו מערכת ליניארית בת N משוואות ויותר מ-N נעלמים (איברי $\underline{U}$), ישנם אינסוף אפשרויות של פתרונות $\underline{U}$, ונשאלת השאלה את מי מהם לבחור. נוכל להעדיף את ה- $\underline{U}$ בעל הנורמה המינימאלית, $\underline{U}$ בעל חלקות מרבית בין איבריו, $\underline{U}$ הדליל ביותר האפשרי ועוד. נקודה מעניינת היא שלמרבית המקרים לא ניתן יהיה להציג את בניית $\underline{U}$ מתוך B ו- $\underline{V}$ כפעולה ליניארית, דהיינו מכפלה במטריצה.

נעבור כעת לתמונות בגודל N על M ונדון בהתמרות המתוארות כמטריצה ריבועית. התמרתה של תמונה כזו תניב MN ערכים. ראינו תהליך בו התחלנו עם התמרה כלשהי, עברנו להתמרות ליניאריות, ופישטנו הלאה ע"י שימוש בהתמרות יוניטריות. הפשטה אפשרית אחרת הנוגעת לדו-ממד היא השימוש בהתמרות פרידות (ספרביליות). בהתמרות אלו, במקום לעסוק ב-MN תמונות בסיס דו-מימדיות, אנו נניח כי יש לנו שתי משפחות של וקטורים חד-מימדיים:

$$\{\underline{c}_k\}_{k=1}^M \in \mathbb{R}^M, \quad \{\underline{d}_k\}_{k=1}^N \in \mathbb{R}^N$$

כלומר, M וקטורים באורך M כל אחד למרחב העמודות, ו-N וקטורים אחרים באורך N כל אחד לייצוג מרחב השורות. משני סטים אלו נבנה הבסיס הדו-מימדי לפי:

$$\underline{b}_{[j,k]} = \underline{c}_j \underline{d}_k^T \Rightarrow \underline{b}_k = \underline{c}_j \otimes \underline{d}_k$$

הביטוי הראשון בונה את $\underline{b}_k$ כתמונה בגודל M על N ע"י הכפלת וקטור עמודה בווקטור שורה. הביטוי השני השקול לו עושה שימוש בסידור לקסיקוגרפי לפי

עמודות, ויוצר את $\underline{b}_k$ ע"י מכפלת קרונקר. את מלוא הבסיס B ניתן לייצר ע"י הפעולה:

$$B = C \otimes D^T$$

נראה דוגמה פשוטה להבהרת הרעיון הזה. להתמרתן של תמונות בגודל 3 שורות על 4 עמודות נתונים לנו שני סטים של בסיסים חד-ממדיים באורך 3 ו-4 בהתאמה. בסיסים אלו מתוארים כעמודות המטריצות C ו-D:

$$C = \begin{bmatrix} 1 & 1 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \quad D^T = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

נבנה בעזרת שני בסיסים אלו בסיס לתמונות דו-ממדיות בגודל 3 (שורות) על 4 (עמודות). יהיו לנו 12 תמונות בסיס בנות 12 אלמנטים כל אחת. לדוגמה, וקטור בסיס ראשון יהיה:

$$\underline{b}_1^T = [1 \ -1 \ 0] \otimes [1 \ 0 \ 1 \ 0] = [1 \ -1 \ 0 \ 0 \ 0 \ 0 \ 1 \ -1 \ 0 \ 0 \ 0 \ 0]$$

באופן דומה, בניית מלוא הבסיס B תיתן:

$$B = C \otimes D^T = \begin{bmatrix} 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & -1 & 1 \\ 1 & 1 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 1 & -1 \end{bmatrix}$$

מדוע בסיס פריד (ספרבילי) נחשב פשוט יותר? בשל התכונה הבאה (המתוארת על ההתמרה ההפוכה):

$$\underline{v} = B\underline{u} = [C \otimes D^T] \underline{u} = [C \cdot U \cdot D^T]_{cs}$$

משוואה זו אומרת שבמקום לכפול את הווקטור $\underline{u}$ במטריצת ההתמרה המלאה, ההתמרה ניתנת להשגה ע"י פעולה על התמונה במבנה דו-מימדי U, כשכופלים משמאל במטריצה C ומימין במטריצה D (עם שחלוף). הפעלת C משמאל

מבצעת התמרה חד-ממדית על העמודות של U , ולאחר מכן הפעלת D מימין משלימה בביצוע התמרה חד-ממדית על השורות של התוצאה.

מהן תמונות הבסיס בהן אנו עושים שימוש בביטוי מהצורה $C \cdot U \cdot D^T$? בהנחה שזוהי ההתמרה ההפוכה, קל לקבל את משפחת תמונות הבסיס ע"י בחירת תמונת מקדמים U כאפסים עם '1' בשורה k ובעמודה n . זה ייתן

$$C \cdot U \cdot D^T = \underline{c}(:,k) \underline{d}(j,:) = \underline{c}_k \underline{d}_j^T$$

בדיוק כפי שתיארנו. השוואת כמות המכפלות תחשוף מדוע זה כה אטרקטיבי: בגישה הישירה עלינו לכפול מטריצה בגודל $[MN \times MN]$ בוקטור בגודל MN . לכן, סך המכפלות הוא M^2N^2 . בגישה הספרבילית, הפעלת C ב- U כרוכה ב- M^2N מכפלות (מטריצת התוצאה היא בגודל M על N , וכל איבר בה חושב ע"י M מכפלות), והכפלה נוספת של התוצאה ב- D דורשת N^2M מכפלות נוספות (מטריצת התוצאה היא בגודל M על N וכל איבר בה דורש N מכפלות). היחס בין כמויות החישובים הוא:

$$\text{Ratio} = \frac{M^2N^2}{NM(M+N)} = \frac{MN}{M+N}$$

לדוגמה, עבור $M=N$, יחס זה גודל כמו $0.5N$ כלומר עולה ליניארית עם N .

עד כה עסקנו בביצוע ההתמרה ההפוכה מ- $\underline{U}$ ל- $\underline{V}$. מה באשר להתמרה הישירה עבור המקרה הספרבילי? נעשה שימוש בתכונה הבאה למכפלת קרונקר ונקבל את הקשר הבא:

$$(C \otimes D)^{-1} = (C^{-1} \otimes D^{-1}) \Rightarrow B^{-1} \underline{V} = (C \otimes D)^{-1} \underline{V} = (C^{-1} \otimes D^{-1}) \underline{V} = C^{-1} \underline{V} D^{-1} \Big|_{CS}$$

כלומר, קיבלנו כי גם ההתמרה הישירה היא ספרבילית, ולכן גם בה קיים אותו מרכיב של חסכון בחישובים. אם בנוסף להיותה ספרבילית, ההתמרות C ו- D יוניטריות, נקבל כי גם פעולת ההיפוך פשוטה, וההתמרה הכוללת קלה מאוד לביצוע. במקרה זה, ההתמרה הכוללת גם היא יוניטרית. ראינו אם כך סיבות מצוינות להעדיף התמרות ריבועיות, יוניטריות ופרידות. ואמנם, אלה הן ההתמרות הפופולריות - התמרת ה- DFT , התמרת ה- DCT , התמרת $HAAR$, והתמרת ה- $Wavelet$. אבל, כפי שנאמר כבר בסעיף זה, ישנם מקרים בהם נרצה התמרה כללית יותר בשל סיבות ייחודיות.

התמרת ה-DFT הדו-ממדית ותכונותיה

במסגרת דיון קודם כבר פגשנו את התמרת ה-DFT הדו-ממדית. במסגרת סעיף זה נחזור ונציגה, והפעם באוריינטציה של התמרה יוניטרית, ספרבילית דו-ממדית. ייחודה וחסרונה המרכזי של התמרה זו, בהשוואה להתמרות אחרות, היא עובדת היותה מרוכבת. ברוב מוחלט של המקרים נעסוק בתמונות מעל הממשיים, והמעבר לערכי התמרה מרוכבים יהווה עבורנו מטרד.

התמרת ה-DFT הדו-ממדית צומחת מתוך התמרת DFT חד-ממדית. התמרה זו מוגדרת באופן הבא: לוקטור $v[m]$ מעל התמך $\Omega_v = \{m \mid 0 \leq m \leq M-1\}$, התמרת ה-DFT החד-ממדית נתונה ע"י

$$V[k] = F_{1D-DFT} \{v[m]\} = \frac{1}{\sqrt{M}} \sum_{m=0}^{M-1} v[m] \exp\left\{-j2\pi \frac{mk}{M}\right\}$$

עבור התמך $\Omega_v = \{k \mid 0 \leq k \leq M-1\}$. ההתמרה ההפוכה מוגדרת ע"י

$$v[m] = F_{1D-DFT}^{-1} \{F[k]\} = \frac{1}{\sqrt{M}} \sum_{k=0}^{M-1} F[k] \exp\left\{j2\pi \frac{mk}{M}\right\}$$

עבור תמונה $f[m,n]$ מעל התמך $\Omega_f = \{[m,n] \mid 0 \leq m \leq M-1, 0 \leq n \leq N-1\}$ התמרת ה-DFT הדו-ממדית נתונה ע"י ביצוע התמרה על שורותיה של התמונה, ולאחר מכן על עמודותיה. סדר זה יכול להתחלף ללא שינוי, כיוון שהפעולה המבוצעת פרידה. לכן, בשלב ראשון מתקבל:

$$F[m, \ell] = F_{1D-DFT} \{f[m,n]\} = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} f[m,n] \exp\left\{-j2\pi \left(\frac{n\ell}{N}\right)\right\}$$

ובשלב שני מתקבל:

$$\begin{aligned} F[k, \ell] &= F_{1D-DFT} \{F[m, \ell]\} = \frac{1}{\sqrt{M}} \sum_{m=0}^{M-1} F[m, \ell] \exp\left\{-j2\pi \left(\frac{mk}{M}\right)\right\} = \\ &= F_{2D-DFT} \{f[m,n]\} = \frac{1}{\sqrt{MN}} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} f[m,n] \exp\left\{-j2\pi \left(\frac{mk}{M} + \frac{n\ell}{N}\right)\right\} \end{aligned}$$

באופן דומה, ניתן לראות כי ביצוע ההתמרה ההפוכה בשני שלבים של התמרה הפוכה חד-ממדית מניבים את הקשר הבא:

$$f[m,n] = F_{DFT}^{-1} \{F[k, \ell]\} = \frac{1}{\sqrt{NM}} \sum_{k=0}^{M-1} \sum_{\ell=0}^{N-1} F[k, \ell] \exp\left\{j2\pi \left(\frac{mk}{M} + \frac{n\ell}{N}\right)\right\}$$

כאשר התמך במישור תדר נתון ע"י $\Omega_F = \{[k, \ell] \mid 0 \leq k \leq M-1, 0 \leq \ell \leq N-1\}$, כך שלעניינינו, יוצא כי רק MN ערכים מגדירים במלואה את ההתמרה. כל שמבדיל בין ההתמרות קדימה ואחורה הוא סימנו של המעריך המרוכב. ההגדרות המוצעות שונות במקצת מאילו המופיעות בדיון הקודם, והבדל זה הינו בקבוע לפני ההתמרה אשר מבטיח כי מדובר בהתמרה יוניטרית.

כשדנים ב-DFT מקובל להשתמש בסימון $W_N^k = \exp\{-j2\pi k/N\}$. אז ההתמרה וההתמרה ההפוכה נתונות ע"י ייצוג מטריצי מהצורה (נתייחס להתמרה חד-ממדית):

$$\begin{bmatrix} V[0] \\ F[1] \\ \vdots \\ F[M-1] \end{bmatrix} = \frac{1}{\sqrt{M}} \begin{bmatrix} W_M^0 & W_M^0 & \dots & W_M^0 \\ W_M^0 & W_M^1 & \dots & W_M^{M-1} \\ \vdots & \vdots & \ddots & \vdots \\ W_M^0 & W_M^{M-1} & \dots & W_M^{(M-1)^2} \end{bmatrix} \begin{bmatrix} v[0] \\ v[1] \\ \vdots \\ v[M-1] \end{bmatrix} \Rightarrow \underline{F} = \mathbf{W}_M \underline{f}$$

מטריצת ההתמרה $\mathbf{W}$ הינה מטריצה סימטרית ויוניטרית, ולכן, היפוכה קיים ופשוט (נתון ע"י הצמוד בלבד בשל הסימטריות). ההתמרה הדו-ממדית ניתנת אף היא לתיאור מטריצי וקטורי. למעשה ישנן שתי אפשרויות לייצגה. בדרך הראשונה, כשאנו מתייחסים ל- f כמטריצת הפיקסלים בגודל M שורות על N עמודות, הכפלתה משמאל במטריצה $\mathbf{W}_M$ תביא לביצוע התמרה על עמודותיה של התמונה. הכפלת התוצאה מימין במטריצה $\mathbf{W}_N$ משלימה ע"י ביצוע התמרה על שורות התמונה, וכך קיבלנו את F - תמונת ההתמרה הדו-ממדית. לכן:

$$F = \mathbf{W}_M f \mathbf{W}_N$$

בגישה השנייה, נמיר את התמונה למבנה וקטורי ע"י סידור לקסיקוגרפי לפי עמודות ונקבל וקטור $\underline{f}$ באורך MN . וקטור זה יוכפל באופן הבא:

$$\underline{F} = \mathbf{W}_{2D-DFT[M,N]} \underline{f} = [\mathbf{W}_M \otimes \mathbf{W}_N] \underline{f}$$

מטריצה זו בגודל MN על MN מבצעת את פעולת ההתמרה במלואה על הווקטור f מניבה את תמונת ההתמרה כשגם היא במבנה וקטור ע"י סידור לקסיקוגרפי לפי עמודות. ברור כי המטריצה הכוללת מהווה אף היא מטריצה יוניטרית (ניתן להוכיח זאת ע"י שימוש בתכונות מכפלת קרונקר).

כבר ראינו בעבר מספר תכונות חשובות של ה-2D-DFT - זוהי התמרה ליניארית, יוניטרית (ולכן משפט פרסבל!!) וספרבילית. ההתמרה מחזורית

במובן זה שאם יוצבו אינדקסים מחוץ לתחום ההגדרה $\Omega_F = \{[k, \ell] \mid 0 \leq k \leq M-1, 0 \leq \ell \leq N-1\}$, התוצאה תחזור על עצמה. יש להתמרה זו יישום מהיר הכרוך ב-

$$MN \log_2 N + NM \log_2 M = (NM) \log_2 (NM)$$

פעולות, במקום ביצוע ישיר של ההתמרה הכרוך ב- $M^2 N^2$ פעולות. כבר ראינו כי ישנה סימטריה נוחה בין המקדמים במקרה בו ההתמרה נעשית על וקטור ממשי.

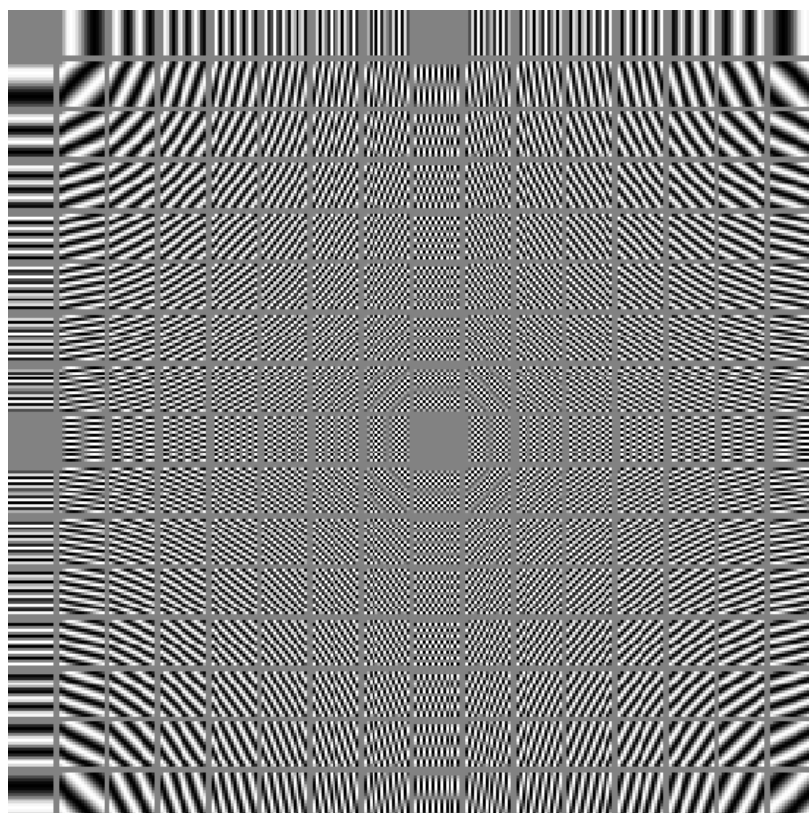
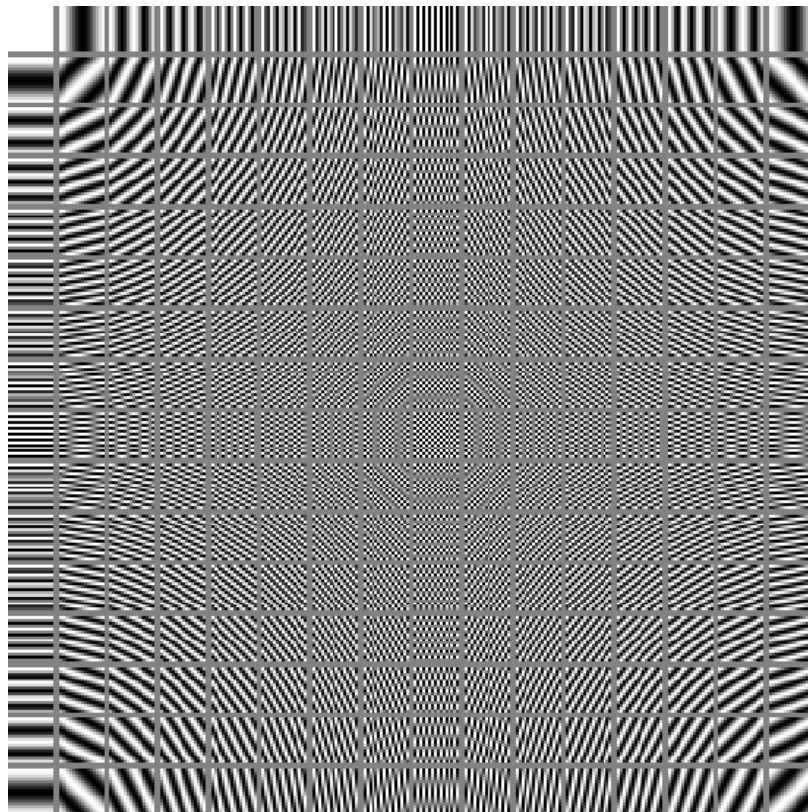
התכונה החשובה מכל היא תכונת הקונבולוציה, לפיה בהינתן שתי תמונות $f_1[m, n]$ ו- $f_2[m, n]$ בגודל זהה M על N , תוצאת ה-DFT הדו-מימדי של הקונבולוציה הציקלית ביניהם זהה לתוצאת המכפלה של התמרותיהן. כלומר, אם:

$$z[m, n] = \sum_{k=0}^{M-1} \sum_{j=0}^{N-1} f_1[k, j] f_2[(m-k) \bmod M, (n-j) \bmod N]$$

$$Z[k, \ell] = F_1[k, \ell] F_2[k, \ell] \text{ אזי}$$

כיצד נראים וקטורי הבסיס של ההתמרה הזו? התמרת ה-DFT הדו-מימדית מציעה שימוש ב- MN תמונות בסיס בגודל M על N כל אחת, לצורך ייצוגה של תמונה נתונה. על מנת לקבל את תמונת הבסיס ה- k , עלינו לבנות וקטור מלאכותי $\underline{f}$ (בתדר) בו כל הערכים 0 למעט האיבר ה- k שיהיה 1, ולבצע התמרת פורייה הפוכה עליו, לקבלת $\underline{f}$. מכאן נובע באופן פשוט כי עמודותיה של המטריצה $\mathbf{W}_{2D-DFT[M, N]}^H = \mathbf{W}_M^H \otimes \mathbf{W}_N^H$ הם וקטורי הבסיס של ההתמרה.

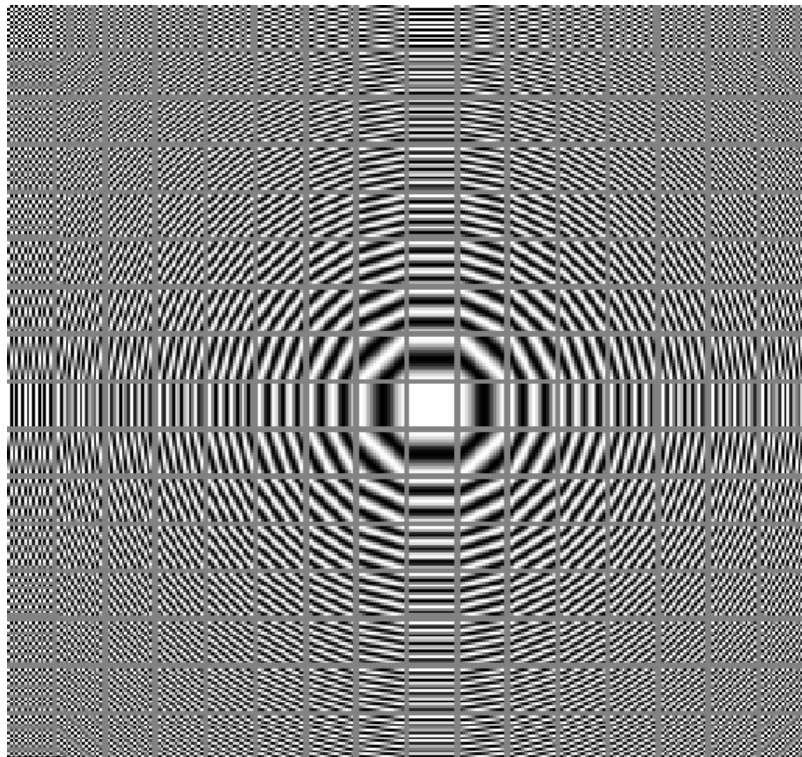
נציג את תמונות הבסיס של התמרת ה-2D-DFT לתמונות בגודל 16 על 16. תהינה 25 תמונות בסיס, אך כיוון שכל תמונת בסיס כזו היא מרוכבת, יש להציג אותה כשתי תמונות עבור החלק הממשי והחלק המדומה. הציור הבא מראה את משפחת הבסיסים הזו.



וקטורי הבסיס של התמרת ה-2D-DFT. למעלה - החלק הממשי,
למטה - החלק המדומה

ניתן לראות כי תמונת הבסיס הראשונה קבועה, וממנה התדר הולך ומתחזק כשהוא הרמוני (פונקציית קוסינוס או סינוס) טהור, ועם מגוון כיוונים. מעבר לחצי (אחרי האינדקס 9) מתקבלת ירידה שוב.

נוח לראות את תמונות הבסיס לאחר פעולת הזזה למרכז, כך שהתדרים עולים כשמרחקים באופן רדיאלי מהמרכז – הציור השני נותן ראייה זו, כשזו מבוצעת על הממשיים בלבד לצורך המחשה. כפי שניתן לראות, סביב נקודת המרכז התדר עולה ובמגוון זוויות לפי כיוון ההתקדמות. באופן זה כל שפה בתמונה יכולה להיות מיוצגת.



וקטורי הבסיס של התמרת ה-2D-DFT לאחר פעולת הזזה למרכז (FFTSHIFT). למעלה - החלק הממשי, למטה - החלק המדומה

הערכים המרוכבים הנכפים עלינו מהווים בעיה. בידיעה שנעסוק בתמונות מעל הממשיים, האם יש דרך לעקוף מגבלה זו? מסתבר שכן! עבור אותות חד-ממדיים, בהינתן אות ממשי כלשהו הנדרש להתמרה, נוכל לשכפלו לאורך כפול ע"י תמונת מראה ולקבל אות סימטרי וממשי. התמרת ה-DFT על אות זה צפויה לתת תוצאה ממשית וסימטרית. לקיחת חצייה הראשון ייתן ייצוג מלא – כך נבנית התמרת ה-DCT.

התמרת ה-DCT הדו-ממדית ותכונותיה

התמרת ה-DCT החד-ממדית לוקטור $v[m]$ באורך M וההתמרה ההפוכה לה נתונות ע"י שני הקשרים הבאים:

$$V[k] = \begin{cases} \sqrt{\frac{1}{M}} \sum_{m=0}^{M-1} v[m] \cos\left[\frac{\pi(2m+1)k}{2M}\right] & k = 0 \\ \sqrt{\frac{2}{M}} \sum_{m=0}^{M-1} v[m] \cos\left[\frac{\pi(2m+1)k}{2M}\right] & k \neq 0 \end{cases}$$

$$v[m] = \sqrt{\frac{1}{M}} V[0] + \sqrt{\frac{2}{M}} \sum_{k=1}^{M-1} V[k] \cos\left[\frac{\pi(2m+1)k}{2M}\right]$$

גם הפעם מדובר בהתמרה ליניארית אך ללא ערכים מרוכבים. ניתן, בדומה ל-DFT, להציע מבנה מטריצי לתיאור ההתמרה, וקל לראות מה יהיה תוכנה של מטריצה מעין זו, אשר תסומן ב- $\mathbf{C}_M$. מטריצה זו אינה סימטרית, אך היא ממשית ויוניטרית. לכן, היפוכה נתון פשוט ע"י פעולת השחלוף (ללא צמוד מרוכב).

בדומה ל-DFT, התמרה דו-ממדית לתמונה f בגודל M על N תיעשה ע"י ביצוע התמרה לשורות ולעמודות, כלומר ביצוע פעולה פרידה. לכן:

$$\mathbf{F} = \mathbf{C}_M \mathbf{f} \mathbf{C}_N$$

או בגישה של תמונה במבנה וקטורי ע"י סידור לקסיקוגרפי לפי עמודות:

$$\underline{\mathbf{F}} = \mathbf{C}_{2D-DCT[M,N]} \underline{\mathbf{f}} = [\mathbf{C}_M \otimes \mathbf{C}_N] \underline{\mathbf{f}}$$

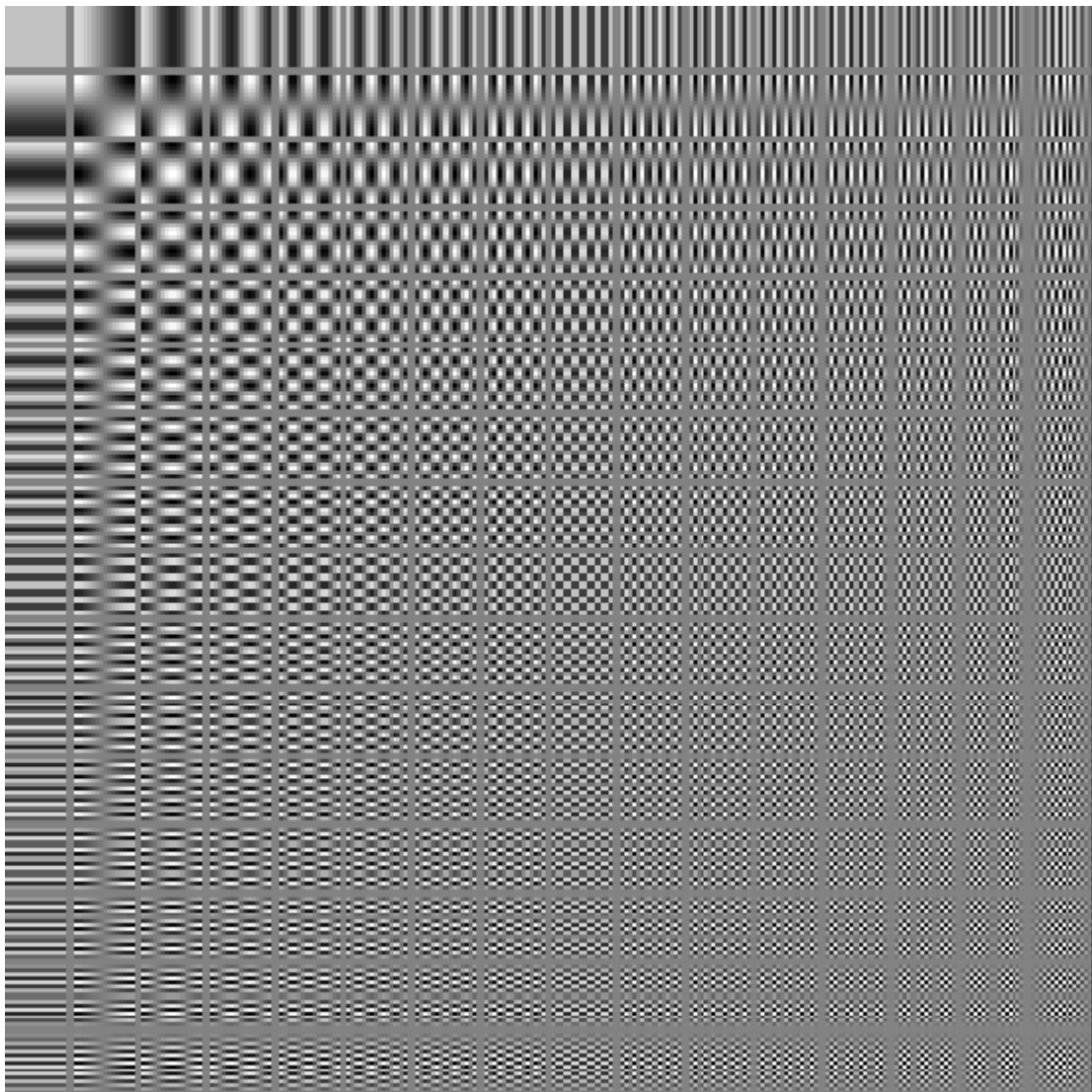
ומטריצה זו הינה שוב יוניטרית, המייצגת התמרה פרידה. נסקור בקצרה וללא הוכחות את תכונותיה של התמרה זו.

1. **ליניאריות, יוניטריות, וספרביליות:** פעולת ה-2D-DCT המתוארת ע"י המטריצה $\mathbf{W}_{2D-DCT[M,N]}$ מהווה התמרה ליניארית, ריבועית, יוניטרית, ופרידה.

כל תכונות אלה קלות לאבחנה מתוך כל שנאמר עד כה.

2. **יישום מהיר:** התמרת ה-DCT קשורה באופן הדוק לפעולת ה-DFT. כתוצאה מכך, גם ביצוע ה-DCT הדו-מימדי על תמונה בגודל M על N יכול להיעשות באופן מהיר ע"י שימוש בפעולות FFT. סך הפעולות הדרוש זהה לביצוע התמרת DFT דו-מימדי בגודל זהה.

3. **קונבולוציה:** בשל הקשר ההדוק בין ה-DCT ל-DFT, ניתן להראות כי ניתן בעזרת DCT לבצע קונבולוציה מהירה יותר, בדומה לדרך בה הדבר מוצע ב-DFT. עם זאת, הקשרים מורכבים יותר, ואין זה הצד החזק של ה-DCT.
4. **וקטורי הבסיס להתמרה:** הציור הבא מראה את וקטורי הבסיס של התמרת ה-DCT הדו-מימדית לתמונות בגודל 16 על 16. ניכר הדמיון הרב להתמרת ה-DFT.



תמונות הבסיס של התמרת ה-DCT לתמונות בגודל 16 על 16.

אנו נחזור להתמרה זו כשנדון בדחיסת תמונות – מסתבר כי התמרה זו מתאימה מאוד לצרכי דחיסה.

התמרת Hadamard הדו-ממדית ותכונותיה

התחלנו את הדיון בהתמרת ה-DFT, וראינו כי היא כופה עלינו חשבונות עם מספרים מרוכבים. התמרת ה-DCT הקלה בעיה זו בהציעה התמרה יוניטרית ממשית. כעת נמשיך קו זה של הפשטה ונציע התמרה בה אין כלל מכפלות במספרים כלשהם - כל איברי המטריצה הממירה יהיו 1 או -1. זוהי התמרת Hadamard, המוכרת גם בשם התמרת Hadamard-Walsh.

כמקודם, נתחיל בתיאור ההתמרה החד-מימדית לוקטור באורך M, ונניח הפעם כי M מהווה חזקה שלמה של 2 (דרישה זו מוכרת לנו גם בהקשר ל-DFT, שם היא חשובה לשם יישום יעיל של ה-FFT). בניגוד לדרך בה נקטנו בדיון בהתמרות קודמות, הפעם נבנה ישירות את מטריצת ההמרה H. התהליך הוא רקורסיבי - נתחיל במטריצה בסיסית בגודל 2 על 2:

$$H_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} +1 & +1 \\ +1 & -1 \end{bmatrix}$$

התמרה זו תקפה לוקטורים באורך 2, וניכר כי היא יוניטרית. לשם בניית התמרה לאורך 4 נבנה את H_2 ע"י:

$$H_2 = H_1 \otimes H_1 = \frac{1}{2} \begin{bmatrix} +1 & +1 & +1 & +1 \\ +1 & -1 & +1 & -1 \\ +1 & +1 & -1 & -1 \\ +1 & -1 & -1 & +1 \end{bmatrix}$$

בשל דרך בנייתה, ברור כי התמרה זו יוניטרית. התהליך יכול להמשיך שוב ושוב, ומתקבל כי:

$$H_n = H_{n-1} \otimes H_1$$

והתוצאה היא מטריצה המכילה +1 ו-(-1), שהינה יוניטרית וסימטרית. בשל כך, ההתמרה ההפוכה נתונה ע"י אותה מטריצה ממש.

ברור כי התמרה זו לתמונות בגודל M על N (שוב, בהנחה כי אלו הם חזקות שלמות של 2) מתקבלת ע"י מכפלת קרונקר בין שתי מטריצות התמרה חד-מימדיות. עובדה זו מבטיחה שההתמרה הכוללת לתמונה תהווה מטריצה סימטרית, יוניטרית, בעלת איברים +1 ו-(-1), ולמעשה, זהה לחלוטין להתמרה לוקטור חד-מימדי באורך NM. דבר זה אינו נכון ל-DCT ו-DFT!

כשמסתכלים על שורה מסוימת במטריצת ההמרה H, מספר החלפות הסימן מקביל לתדר. מספר החלפות הסימן מוכר בשם Sequency, וניתן לראות כי שורות המטריצה H אינן מסודרות בסדר Sequency מונוטוני. מקובל לעיתים לקחת את מטריצת ההמרה שנבנתה ולסדר את שורותיה כך שאמנם יתקבל Sequency גדל מונוטונית.

דוגמה: נראה מהי מטריצת Hadamard לוקטורים באורך 8, וכיצד לסדרה. המטריצה נתונה ע"י:

$$H_3 = \frac{1}{2\sqrt{2}} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \\ +1 & -1 & +1 & -1 & +1 & -1 & +1 & -1 \\ +1 & +1 & -1 & -1 & +1 & +1 & -1 & -1 \\ +1 & -1 & -1 & +1 & +1 & -1 & -1 & +1 \\ +1 & +1 & +1 & +1 & -1 & -1 & -1 & -1 \\ +1 & -1 & +1 & -1 & -1 & +1 & -1 & +1 \\ +1 & +1 & -1 & -1 & -1 & -1 & +1 & +1 \\ +1 & -1 & -1 & +1 & -1 & +1 & +1 & -1 \end{bmatrix}$$

ה- Sequency לפי שורות הוא - 0, 3, 4, 1, 6, 2, ו- 5. המטריצה לאחר סידורה תהיה:

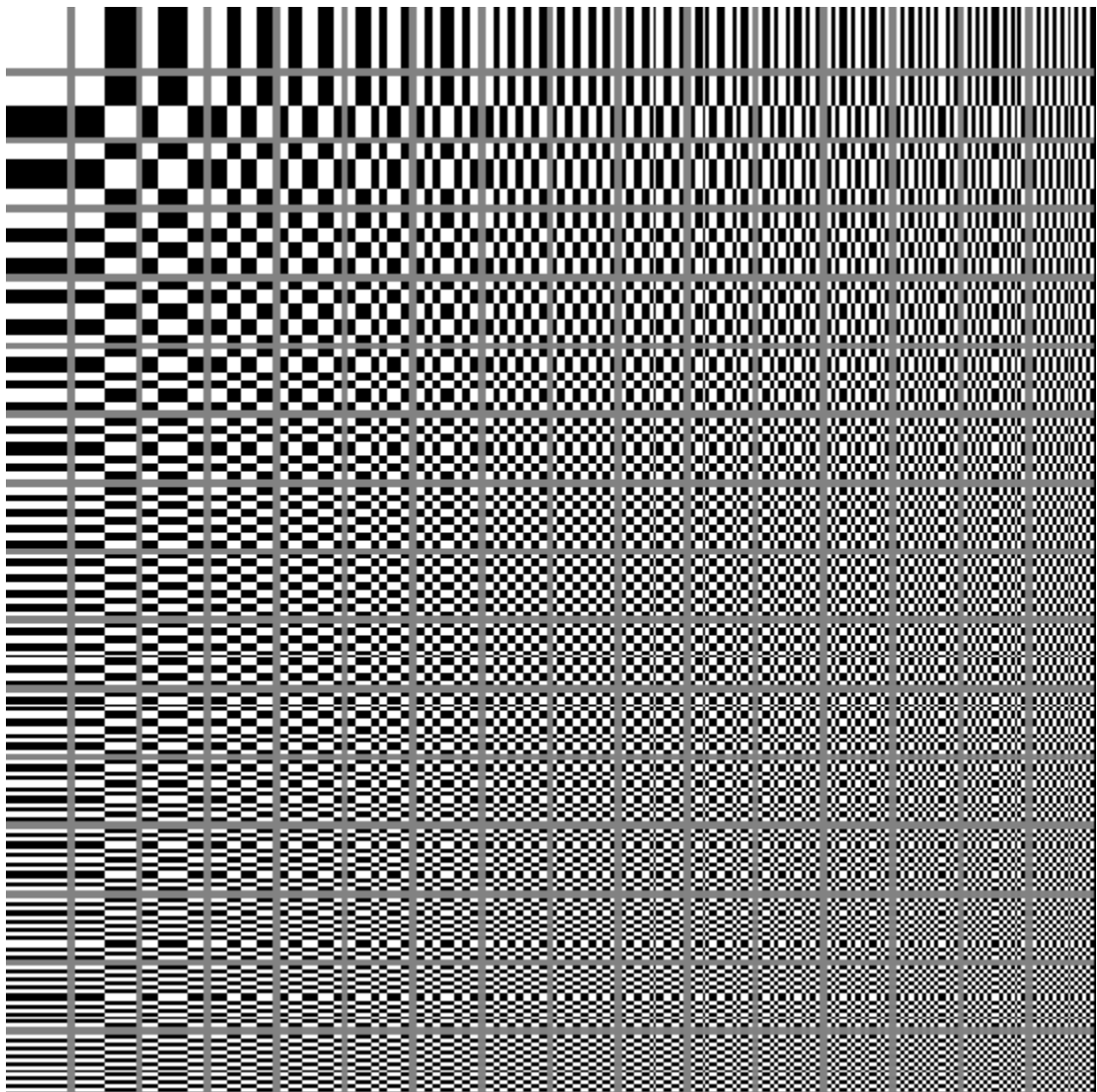
$$\begin{bmatrix} +1 & +1 & +1 & +1 & +1 & +1 & +1 & +1 \\ +1 & +1 & +1 & +1 & -1 & -1 & -1 & -1 \\ +1 & +1 & -1 & -1 & -1 & -1 & +1 & +1 \\ +1 & +1 & -1 & -1 & +1 & +1 & -1 & -1 \\ +1 & -1 & -1 & +1 & +1 & -1 & -1 & +1 \\ +1 & -1 & -1 & +1 & -1 & +1 & +1 & -1 \\ +1 & -1 & +1 & -1 & -1 & +1 & -1 & +1 \\ +1 & -1 & +1 & -1 & +1 & -1 & +1 & -1 \end{bmatrix}$$

נסקור בקצרה וללא הוכחות את תכונותיה של התמרה זו.

1. **ליניאריות, יוניטריות, וספרביליות:** פעולת ה- 2D-Hadamard המתוארת ע"י המטריצה $H_{2D-Had.[M,N]}$ מהווה התמרה ליניארית, ריבועית, יוניטרית, ופרידה. כל תכונות אלה קלות לאבחנה מתוך כל שנאמר עד כה. יתרה מזו, התמרה דו-ממדית זו זהה בדיוק להתמרת Hadamard חד-ממדית על וקטורים באורך MN. המטריצה $H_{2D-Had.[M,N]}$ סימטרית ומורכבת מאיברי +1 ו-1 בלבד.

2. **יישום מהיר:** גם להתמרה זו יישום מהיר, בדומה למבנה FFT, אך מהיר אף יותר בשל שימוש בחיבורים בלבד.

3. **וקטורי הבסיס להתמרה:** הציור הבא מראה את וקטורי הבסיס של התמרת Hadamard הדו-ממדית לתמונות בגודל 16 על 16. ניכר הדמיון הרב להתמרת ה-DCT. תכונה מעניינת היא שאם לוקחים את בסיס ה-DCT לאחר פעולת sign (כלומר שליליים ל-1- וחיוביים ל-1+), מקבלים התמרה קרובה מאוד להתמרת Hadamard, עובדה המצביעה על הקשר והדמיון הקרוב ביניהן.



תמונות הבסיס של התמרת Hadamard לתמונות בגודל 16 על 16.

סיכום

אנו נעצור את הדיון כאן ונביא את המשכו – הדיון בהתמרת ה-wavelet בנפרד. אנו נראה כי ההתמרות שהוצגו כאן יכולות להוות בסיס נאות לדחיסת תמונות ולהביא לביצועים טובים מאלו שהושגו בגישת החיזוי.

עיבוד תמונות בעזרת מחשב (236327)

הרצאה מס' 13

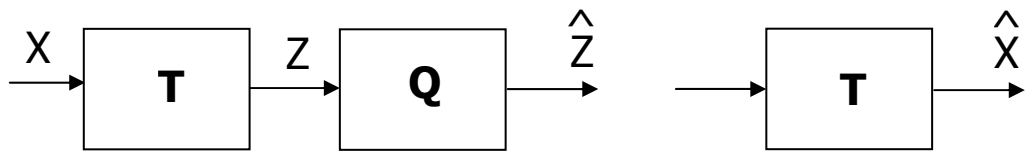
דחיסת תמונות – גישות התמרה

דחיסה בגישת ההתמרה - יסודות

כבר ראינו כי התמרת הפורייה מועילה מאוד בביצוע קונבולוציה, בניתוח מערכות ליניאריות וקבועות במקום, ואף ראינו את תרומתה בשחזור תמונות. מסתבר כי להתמרה זו, או התמרות דומות לה מקום חשוב גם בדחיסה. הרעיון הבסיסי פשוט – בהינתן סט של פיקסלים, נחליפם בערכי ההתמרה, ואת אלה נקודד ונשדר. אם ההתמרה תיבחר בצורה נאותה, נקבל כי קידוד ערכי ההתמרה יעיל יותר מקידוד הפיקסלים במקור.

לכן, כתחליף לגישת החיזוי מוצעת גישה אחרת - גישת ההתמרה. גם גישה זו מנסה להשיג את שתי התכונות שהוזכרו – הוצאת קשר בין הערכים לקידוד, ומיצוי האנטרופיה - אך בדרך שונה. ניקח את התמונה הנדרשת - נניח כי היא בגודל M שורות על N עמודות - ונבצע עליה התמרה. נניח כי אנו משתמשים בהתמרה ליניארית ריבועית (כלומר - בהינתן סט של K פיקסלים, ההתמרה הינה הכפלה במטריצה בגודל K -על- K , ומתקבל לפיכך מוצא עם K איברים כבמקור.

אם ההתמרה (המטריצה בגודל K -על- K שמיישמת את ההתמרה) תיבחר בחוכמה, נקבל כי מקדמי ההתמרה יהיו חסרי קשר סטטיסטי ביניהם, ובנוסף בעלי אנטרופיה נמוכה. נוכל להעביר כל אלמנט כסקלר לקידוד אנטרופיה, ולהשיג כתוצאה מכך דחיסה. על מנת להכניס שגיאות לתהליך ולשפר את יחס הדחיסה, נוכל לקחת את ערכי ההתמרה ולהעבירם דרך קוונטיזר בטרם העברתם לקידוד אנטרופיה. הציור הבא מראה סכימת בלוקים של מקודד ומפענח הבנויים על עקרון זה.



סכימת דחיסה עם שגיאות המבוססת על התמרה. מצד שמאל המשדר ומצד ימין המקלט.

נותרה השאלה המרכזית - מיהי ההתמרה אשר תגשים יעד זה של ביטול הקשר בין הפיקסלים והקטנת האנטרופיה שלהם.

למעשה, תפיסת ההתמרה לא שונה בתכלית מגישת החיזוי. נניח כי בידנו וקטור באורך 8 אלמנטים הדורש דחיסה. גישת החיזוי מחליפה כל ערך בחיזוי סיבתי של שכניו ושגיאת החיזוי היא זו המשודרת. נוכל להציע התמרה שבמקביל תחשב משמנות האלמנטים את חיזויים על סמך שכניהם, ע"י ההכפלה בהתמרה

$$\begin{bmatrix} 1 & -0.5 & 0 & 0 & 0 & 0 & 0 & -0.5 \\ -0.5 & 1 & -0.5 & 0 & 0 & 0 & 0 & 0 \\ 0 & -0.5 & 1 & -0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & -0.5 & 1 & -0.5 & 0 & 0 & 0 \\ 0 & 0 & 0 & -0.5 & 1 & -0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & -0.5 & 1 & -0.5 & 0 \\ 0 & 0 & 0 & 0 & 0 & -0.5 & 1 & -0.5 \\ -0.5 & 0 & 0 & 0 & 0 & 0 & -0.5 & 1 \end{bmatrix}$$

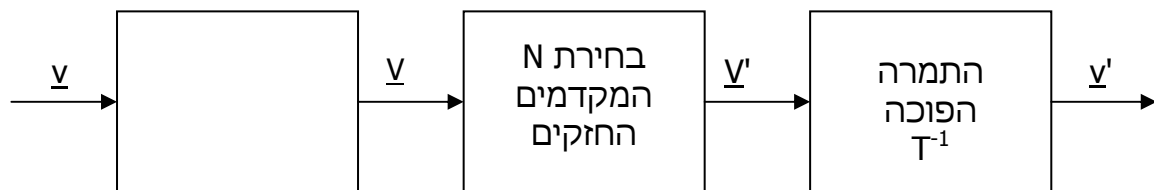
הבעיה עם בחירת התמרה כזו היא שהיא סינגולרית. פירוש הדבר שביצוע התמרה הפוכה אינה אפשרית. אפילו לו בחרנו מטריצה אחרת שאינה סינגולרית, זו בדרך כלל תהיה מטריצה בעלת condition-number (יחס ערכים עצמיים גדול ביותר לקטן ביותר) גבוה, וזוהי ערובה להגבר מטריד של שגיאות בתהליך ההיפוך.

תפיסת ההתמרה מוציאה את העוקץ מבעיות נומריות אלו ע"י ביצוע סידרתי וסיבתי של ההתמרה והיפוכה. במונחים של אלגברה ליניארית נאמר כי ההתמרה בשיטות אלו נעשית ע"י מטריצה משולשת תחתונה (כך מושגת הסיבתיים), והיישום מערבב כניסה ומוצא כדי למנוע אובדן יציבות. כעת נחפש גישה אחרת.

התמרה אופטימלית לתמונה

נציג אם כך את הבעיה הבאה - נתון לנו סט של תמונות ורצוננו לתכנן את ההתמרה הטובה ביותר עבורן. ברור כי מיד עולה השאלה - מהו קריטריון הטיב בו נשתמש? נשתמש בתכונת הייצוג הקומפקטי ע"י חלק מבסיסי ההתמרה כדרך לקביעת ההתמרה הטובה ביותר, וניכר שיש לה קשר לדחיסה.

הרעיון הוא הבא - נקבל וקטור (אשר יכול להיות תמונה לאחר סידור לקסיקוגרפי) $\underline{v}$ שאורכו M , ונבצע עליו התמרה T , לקבלת $\underline{v'}$. לאחר ההתמרה ניקח את M איברי הוקטור $\underline{v'}$ ומתוכם נבחר את N (ברור כי $0 < N \leq M$) האיברים הראשונים (בהנחה שהם יהיו החזקים), ונאפס את האחרים. פעולה זו תסומן ב- $I_N X$ - כלומר, הוקטור מוכפל במטריצה בגודל M על M הדומה למטריצת יחידה I , למעט העובדה שרק N איברים ראשונים על האלכסון הם 1. את התוצאה שתסומן ב- $\underline{v'}$ נזין להתמרה הפוכה של T , ונקבל וקטור המתיימר להיות קירוב של $\underline{v}$, ולכן יסומן ב- $\underline{v'}$. הציור הבא מראה תהליך זה כסכימת בלוקים. התמרה מוצלחת תיחשב התמרה שלכל N שנבחר תיתן כי $\underline{v}$ ו- $\underline{v'}$ קרובים ככל האפשר.



מערכת דחיסה אלמנטרית הזורקת אלמנטי התמרה חלשים

בעיה אחרת לכאורה שונה היא הבעיה הבאה - נתייחס לסכימה בציור הנ"ל ונדרוש כי לאחר איפוס $M-N$ האיברים החלשים, וקטור התוצאה $\underline{v'}$ יהיה בעל האנרגיה המרבית האפשרית. ברור כי אם T תיבחר להיות יוניטרית מטריצה שעמודותיה אורתונורמליות, אנרגיה זו תהיה קטנה מאנרגיית וקטור המקור $\underline{v}$, ולכן דרישה זו דומה במקצת לדרישה הקודמת כיוון ש- T תנסה להיות כזו שתרכז את האנרגיה של הוקטורים המוזנים לה במקדמים הראשונים של ההתמרה.

נציג שתי בעיות אלו באופן פורמאלי - סט התמונות (למעשה וקטורים) עבור מתוכנת ההתמרה האופטימאלית יסומן ב- $\{\underline{v}_1, \underline{v}_2, \dots, \underline{v}_L\}$. עבור N כלשהו נקבל כי:

$$\underline{V}' = I_N T \underline{V} \quad , \quad \underline{v}' = T^{-1} I_N T \underline{v}$$

לכן, הבעיה הראשונה אותה הגדרנו תהיה:

$$\text{Minimize}_T \quad \varepsilon_1^2\{T\} = \sum_{k=1}^L \left\| T^{-1} I_N T \underline{v}_k - \underline{v}_k \right\|^2$$

כשמשמעות הביטוי היא שלכל אחד מהוקטורים עבורם נרצה את ההתמרה, על ההפרש בין כניסת המערכת למוצאה להיות קטן ככל האפשר. עבור הבעיה השנייה שהגדרנו נקבל את בעיית האופטימיזציה הבאה:

$$\text{Maximize}_T \quad \varepsilon_2^2\{T\} = \sum_{k=1}^L \left\| I_N T \underline{v}_k \right\|^2$$

כשכאן אנו דורשים אנרגיה מרבית לאחר ההתמרה והקיצוץ.

התקבלו שתי בעיות אופטימיזציה מאוד מעניינות אשר לא נפתור במסגרת זו! מסתבר כי הפתרון לשתי הבעיות זהה. ההתמרה האופטימלית בשני המובנים הנ"ל היא המטריצה היוניטרית המלכסנת את המטריצה הבאה:

$$R = \sum_{k=1}^L \underline{v}_k \underline{v}_k^H$$

כאשר יש לסדר את עמודותיה לפי סדר יורד של הערכים העצמיים. לכן, קיבלנו התמרה יוניטרית חדשה התלויה בתמונות עליהן היא מבוצעת. כאשר אנסמבל התמונות אינסופי ומייצג את חוק פילוג התמונות, התמרה זו קרויה התמרת KLT - על שמם של Karhunen ו-Loeve אשר הציעו אותה לראשונה. למעשה, כאשר מדובר באנסמבל כזה, המטריצה R היא מטריצת האוטוקורלציה של הוקטור האקראי $\underline{v}$.

האם התמרה זו מעשית? מסתבר כי במקרים רבים לא! הצורך לחשב התמרה לתמונות מהווה מעמסה כבדה ומטרידה. בדחיסת תמונות, שם נעשה שימוש ניכר בהתמרות, גישה זו גם אינה אפשרית כיוון שיש לשדר למקלט גם את ההתמרה בה נעשה שימוש, וזה הורס את יחס הדחיסה. שאלה טבעית המתעוררת אם כן היא האם למשפחות מסוימות של תמונות ניתן לקרב את ה-KLT ע"י התמרה פשוטה ומוכרת? למזלנו התשובה לכך היא חיובית. לתמונות

טבעיות, עבור בלוקים קטנים יחסית (עד 16 על 16 פיקסלים), התמרת ה-DCT מהווה קירוב טוב מאוד להתמרת ה-KLT.

תכונה מעניינת של התמרת ה-DCT שאותה לא הזכרנו היא שמטריצת התמרת ה-DCT החד-מימדית מלכסנת במדויק מטריצות תלת-אלכסוניות, בעלות המבנה הבא:

$$Q = \begin{bmatrix} 1-a & -a & 0 & & \\ -a & 1 & -a & 0 & \\ & \ddots & \ddots & \ddots & \\ & 0 & -a & 1 & -a \\ & & 0 & -a & 1-a \end{bmatrix}$$

מבנה זה קרוי מטריצה תלת-אלכסונית (יש רק שלושה אלכסונים פעילים) בעלת מבנה טופליץ (איברי כל אלכסון שווים). בשל תכונה זו, ניתן להראות כי מטריצות טופליץ מרקוביות הבנויות באופן הבא:

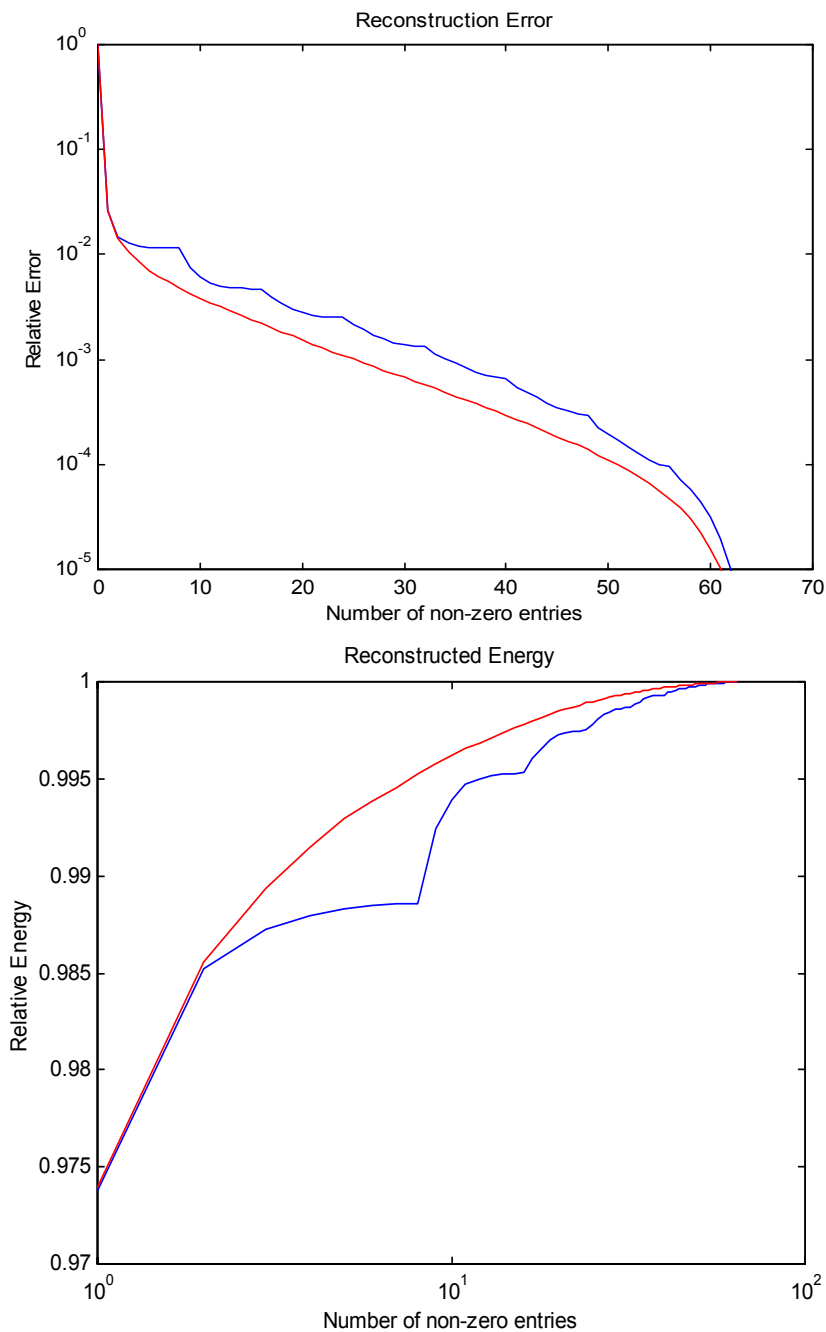
$$R = \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{M-1} \\ \rho & 1 & \rho & \rho^2 & \dots \\ & \ddots & \ddots & \ddots & \\ & \rho^2 & \rho & 1 & \rho \\ \rho^{M-1} & & \rho^2 & \rho & 1 \end{bmatrix}$$

מתלכסנות ע"י התמרת ה-DCT בקירוב טוב אם $\rho \rightarrow 1$. הסיבה לכך היא שעבור $\rho \rightarrow 1$ מתקבל כי היפוכה של R היא בדיוק טוב Q, ולכן ברור שה-DCT מהווה מטריצה מלכסנת. תכונה זו חיונית כיוון שמטריצת טופליץ מרקובית מתארת התנהגות אופיינית של אותות בטבע, ולכסונה חשוב לצורך ייצוג נאות או דחיסה.

מסתבר כי לבלוקים קטנים של פיקסלים ישנו מבנה מקרובי למטריצה R. משמעות תופעה זו – אם נחשב את הקורלציה בין פיקסלים שכנים, ההתנהגות תהיה מרקובית, בה הקשר של פיקסל לשכן כלשהו פרופורציונית באופן מערכי דועך למרחקם.

בשל האמור לעיל, ה-DCT מהווה התמרה פופולארית מאוד בעיבוד תמונה בכלל ובדחיסת תמונות בפרט. בציור הבא מובאת השוואה בין ה-DCT ל-KLT. השוואה זו מתייחסת ל-1024 בלוקים של תמונות בגודל 8 על 8

שנלקחו מהתמונה Lena. בהשוואה זו מוצגים שני קריטריוני הטיב אותם הצגנו, וניכר כי על-פי שניהם ה-KLT טוב יותר, אך גם ניכרת קירבה גדולה בין השניים.



השוואת ה-KLT (אדום) ל-DCT (כחול) עבור 1024 בלוקים בגודל 8 על 8 פיקסלים. למעלה: השוואת שגיאת השחזור (קריטריון ראשון), למטה: השוואת מידת האנרגיה היחסית (קריטריון שני).

דחיסה בשיטת ההתמרה - JPEG

ראינו כי התמרת ה-KLT היא ההתמרה היוניטרית הטובה ביותר במובן שאם נזרוק חלק ממקדמי ההתמרה, השגיאה בין תמונת המקור לתמונה המשוחזרת לאחר התמרה הפוכה תהיה מינימלית. זריקת חלק מהאיברים כמו ביצוע של קוונטיזציה גסה מאוד עבור האיברים הנזרקים. יתרה מזו - הזכרנו כי ה-KLT מלכסנת את מטריצת האוטו-קורלציה של המידע בתמונה, ומסתבר (ללא הוכחה) כי לכסון זה כמוהו כמו ביטול הקשר הסטטיסטי בין הפיקסלים השונים.

ברור כי אסור לשכוח את החיסרון המרכזי של התמרת ה-KLT - היותה תלויה בתמונה בה אנו משתמשים. בשל תכונה זו לא ניתן להשתמש ב-KLT לדחיסה. אבל ראינו כי התמרת ה-DCT מהווה קירוב טוב מאוד ל-KLT, והדבר נכון במיוחד לתמונות קטנות (16 על 16 ומטה פיקסלים). לכן, כתחליף ל-KLT ניתן לקחת את התמונה, לחלק אותה לבלוקים קטנים (גודל אופייני לשימוש הוא 8 על 8), לבצע על כל בלוק כזה התמרת DCT דו-מימדית. התמרה זו יכולה להיעשות ע"י אלגוריתם מהיר המבוסס על ה-FFT.

לאחר ההתמרה, כל מקדם התמרה מועבר דרך קוונטיזר. מתוך 64 הערכים המתקבלים (בשל ההתמרה על 8 על 8), הקרובים לראשית הם החשובים בהיותם מייצגים מידע נמוך תדר. עבורם תיעשה קוונטיזציה עדינה (שקול לאמירה כי להם תינתן שגיאה קטנה יחסית). לערכים הרחוקים מהראשית תבוצע קוונטיזציה גסה יותר, ויתכן שאף ערכים אלו ייזרקו. גישה מקובלת היא לבצע קוונטיזציה יוניפורמית לערכי ההתמרה עם גודל מדרגה התלוי במיקום המקדם - קטן עבור מקדם הקרוב לראשית, וגדול עבור מקדם הרחוק מהראשית. גם כאן ניתן להשתמש בקוונטיזר בעל מדרגות ברוחב משתנה (מקס-לוייד) או קוונטיזציה הנבנית מתוך שיקולי מערכת הראיה.

האלגוריתם שתואר הוא אלגוריתם ה-JPEG הידוע ביותר הסטנדרט לדחיסת תמונות. אלגוריתם זה מבצע דחיסה ע"י התמרת DCT דו-מימדית וקוונטיזציה אחידה תלוית מקדם.

דוגמה: נדחוס את התמונה Lena ע"י אלגוריתם בסיסי מבוסס התמרה. בשלב ראשון כל בלוק של 8 על 8 מותמר ע"י DCT דו-מימדי. כל בלוק של 8 על 8 עובר קוונטיזציה ל- 8 סיביות באיבר ה- $(0,0)$, ל- 3 סיביות לאיברים $(0,*)$ ו- $(*,0)$, (סה"כ 14 איברים כאלה המייצגים את השורה הראשונה והעמודה הראשונה), ול- 2 סיביות לאיברים $(1,1)$, $(1,2)$, $(2,1)$ ו- $(2,2)$. לכן, לכל בלוק שבמקור יוצג ע"י $8 \times 64 = 512$ סיביות אנו משתמשים ב- 68 סיביות, כלומר יחס דחיסה של 1 ל- 8 בקירוב. גודל המדרגה נקבע ע"י חישוב השונות של כל מקדם התמרה ע"י מיצוע הבלוקים בכל התמונה. התקבלה התמונה הבאה, המופיעה בציור הבא. השגיאה בין שתי התמונות (מחושבת ע"י סכום ריבועי ההפרשים מחולק במספר הפיקסלים ולאחר שורש) היא 13.14 (ביחידות של רמות אפור).

מעיון בתמונה המתקבלת ניכרת תופעה האופיינית לאלגוריתמי דחיסה הפועלים בבלוקים - אפקט הבלוקיות. גבולות הבלוקים ניכרים וצורמים - תופעה שנובעת מכך שהפעולות שבוצעו זרקו מידע גדול מהבלוק, ולא נעשה כל ניסיון לשמור על רצף בין הבלוקים. קיימים אלגוריתמי דחיסה מוצלחים יותר אשר מטפלים יפה בבעיה ומצליחים להשיג יחס דחיסה של 1 ל- 10 בלי שעין אנושית תוכל להבחין בהבדל. קיימים גם אלגוריתמים הנובעים מתוך תורת השחזור שנדונה בפרק קודם אשר יודעים לקחת תמונה עם אפקט קלקול זה ולתקנה תוך כפיית חלקות בין הבלוקים ושימוש שפות באופן כללי.



תמונת מקור וגרסתה הדחוסה בגישת ההתמרה כמתואר בדוגמה

לשם השוואה, נפעיל את אלגוריתם ה- JPEG כפי שמיושם ב- Matlab. הפעלת אלגוריתם זה נעשית ע"י פקודת שמירת התמונה (imwrite) באופן הבא - בהינתן תמונה כמערך ב- Matlab הקרוי A (ונתון כ- uint8), הפקודה

```
imwrite(A,'temp.jpg','quality',70)
```

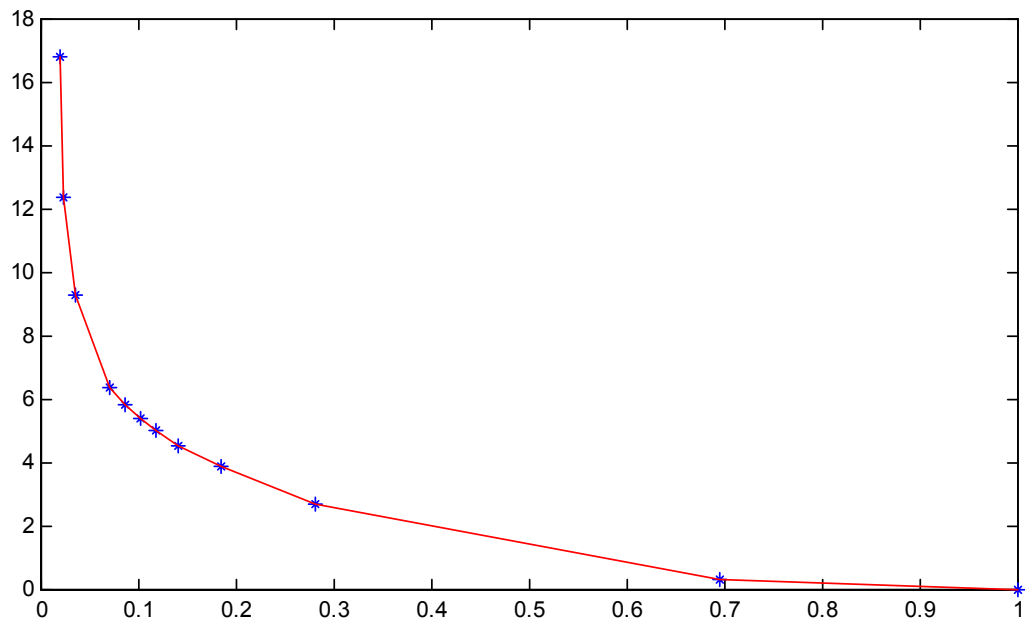
שומרת תמונה זו לקובץ בשם temp.jpg (השם מסגיר איזה פורמט התמונה מקבלת - במקרה זה דחיסת JPEG). באלגוריתם ה- JPEG, לא קובעים מראש את יחס הדחיסה אלא את מידת אובדן המידע. עבור איכות של 100% מקבלים דחיסה ללא אובדן, ועבור ערך נמוך יותר מתירים אובדן. כאשר ערך זה מאוד נמוך יתקבל יחס דחיסה מצוין על חשבון שגיאה גדולה בין התמונה המקורית לתמונה לאחר פריסה.

הציור הבא מראה תוצאות של דחיסת ה- JPEG. הגרף התקבל עבור מגוון ערכי האיכות (quality), וכנגד כל ערך כזה מוצג שורש השגיאה הריבועית הממוצעת כפונקציה של נפחו היחסי של הקובץ ביחס למקור. רואים כי עבור התמונה המקורית בנפח יחידה מקבלת שגיאה 0, ירידה לנפח של 0.7 מהמקור נותנת שגיאה זעומה של כ- 0.3 רמות אפור. תוצאה זו מעניינת כיוון שהיא חושפת תכונה שלילית של קידוד מבוסס התמרה - יכולנו לקבל נפח של כ- 0.5 ללא שגיאה כלל, לו עשינו שימוש בדחיסת אנטרופיה. לעומת זאת, בהתמרה, מתקבלת שגיאה גם לנפחים גדולים מזה. עבור נפח של כ- 0.1 מתקבלת שגיאה של 5.4 - הרבה יותר טוב מהשגיאה אותה קיבלנו בדוגמה לעיל.

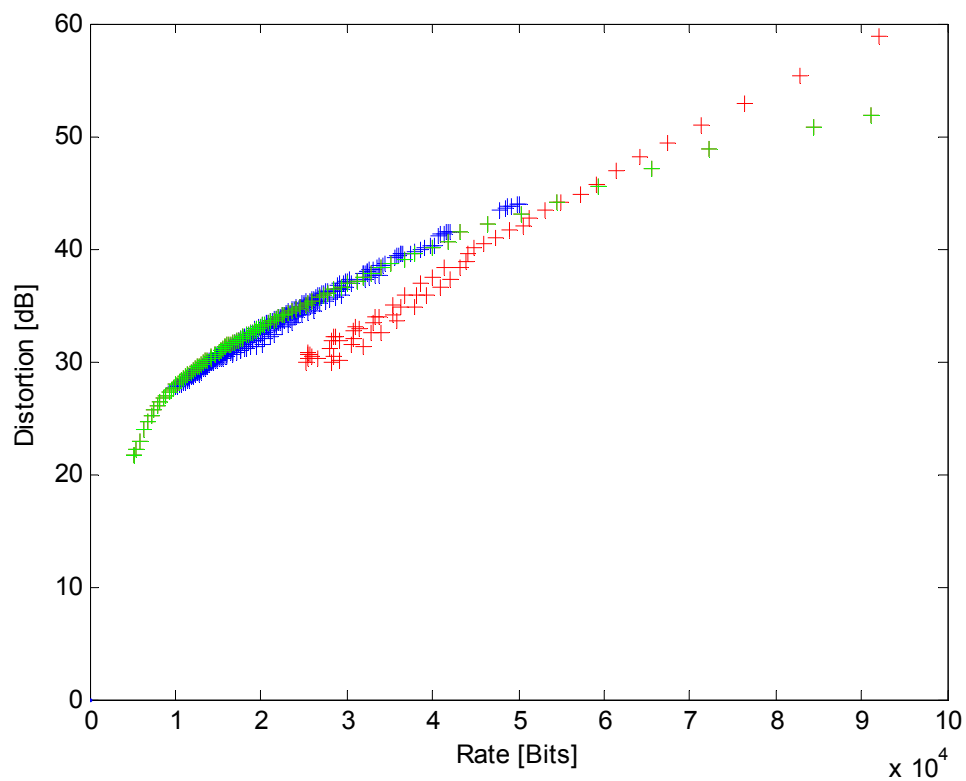
למעשה, נוכל להציג את אותו גרף בדיוק עם שימוש ב- PSNR כפי שענינו קודם בדיון בשיטות חיזוי. הציור הבא מראה השוואה של ה- JPEG לשיטות ה- DPCM ושיטת סורק-זאבי. כפי שניתן לראות, הביצועים פחות או יותר שקולים בין סורק-זאבי ובין ה- JPEG (אך יש לזכור - בשיטת סורק-זאבי חישבנו אנטרופיה ולא ממש ביצענו קידוד אשר עלול לבזבז סיביות בפועל).

עוד תופעה מעניינת היא ששימוש ב- JPEG באיכויות גבוהות נחות מסכימת ה- DPCM. מסתבר כי תופעה זו הייתה ידועה והטרידה את הועדה שקבעה את JPEG כסטנדרט, ולכן נבחר אלגוריתם נוסף הקרוי JPEG-LS כאלגוריתם ה- JPEG הרשמי לדחיסה ללא אובדן או כמעט ללא אובדן. אלגוריתם

זה פותח ע"י שלושה חוקרים ממעבדות HP (כולם בוגרי הטכניון) והוא מבוסס על DPCM, אם כי שיטת קידוד האנטרופיה שם חכמה לאין ערוך.



ציר אנכי מייצג את נפחו היחסי של קובץ התמונה, והציר האנכי את שורש השגיאה הריבועית הממוצעת. כל זאת עבור אלגוריתם JPEG כפי שמיושם ב-Matlab.



השוואת JPEG לשיטת סורק-זאבי - עקומות R(D). אדום - DPCM, כחול - סורק-זאבי, וירוק - JPEG

התמונה הבאה מראה מספר תמונות אופייניות של ה-JPEG והתופעות הצפויות להתקבל בדחיסה עמוקה – אפקט הבלוקיות. במהלך העשור האחרון הוצעו אלגוריתמים רבים שמציעים post-processing על תמונות JPEG לשם "מחיקת" תופעת הבלוקיות. גישות אלו בפועל "מרימות" את עקומת ה- $R(D)$ באזור הקצבים הנמוכים ע"י שיפור ה-SNR.

כיצד ה-JPEG בפועל עושה את הקוונטיזציה והקידוד? הדבר אינו פשוט כלל, ומשם מושג עיקר הכוח בביצועי הדחיסה. נראה כאן את העקרונות בלבד, משום שבעקר זהו נושא טכני (ומשעמם מבחינה מדעית).



תוצאות ה-JPEG: שמאל למעלה - מקור, ימין למעלה - יחס דחיסה 7.12 ושגיאה 4.5, למטה משמאל - יחס דחיסה 11.65 ושגיאה 5.84, למטה בצד ימין - יחס דחיסה 52.27 ושגיאה 16.8

בידינו בלוקים בגודל 8 על 8 פיקסלים שעברו התמרת DCT. הצעד הראשון יהיה קוונטיזציה של מקדמים אלו, וברור כי לא נכון לבצע קוונטיזציה זו בצורה זהה לכל המקדמים. חוקרים של מערכת הראיה האנושית בדקו מהו סף

הרגישות של מערכת הראיה שלנו לכל אחד ממקדמי ההתמרה – יוצרים תמונה ובה מזיזים מקדם מסוים מעלה מטה עד למצב שהצופה מתחיל להבחין בהבדל. לבלוקים של 8 על 8 פיקסלים נמצאה הטבלה הבאה כערכי הסף היחסיים:

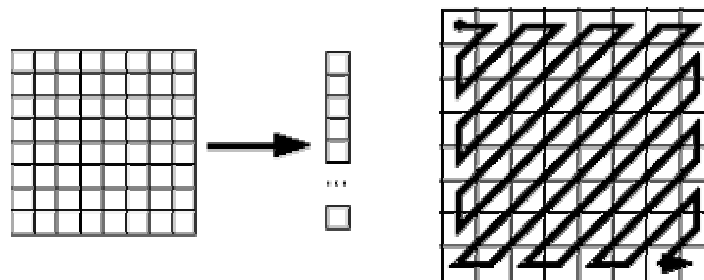
16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

פירוש הדבר הוא, למשל, שמקדמים הקרובים לפינה השמאלית העליונה חשובים יותר, ואילו בפינה הימנית התחתונה פחות חשובים. מכאן ואילך נקבע את מדרגות הקוונטיזציה בכל מקדם להיות פרופורציוניים למספרים אלו. כיוון שאנו רוצים יחסי דחיסה שונים, נוכל להכפיל מטריצת מספרים אלו במקדם מתקן – קטן ליחסי דחיסה נמוכים וגדול ליחסי דחיסה גבוהים – ולבצע הקוונטיזציה.

הצעד הבא הוא דחיסת אנטרופיה של מקדמי התמרה לאחר הקוונטיזציה. ב-JPEG מקובל לשדר את האיבר ה-(1,1) בנפרד ולבצע DPCM (ללא אובדן) על מקדמים אלו. הרעיון הוא לנצל קשרים בין ערכים סמוכים, ולבצע קוד הופמן על ההפרשים.

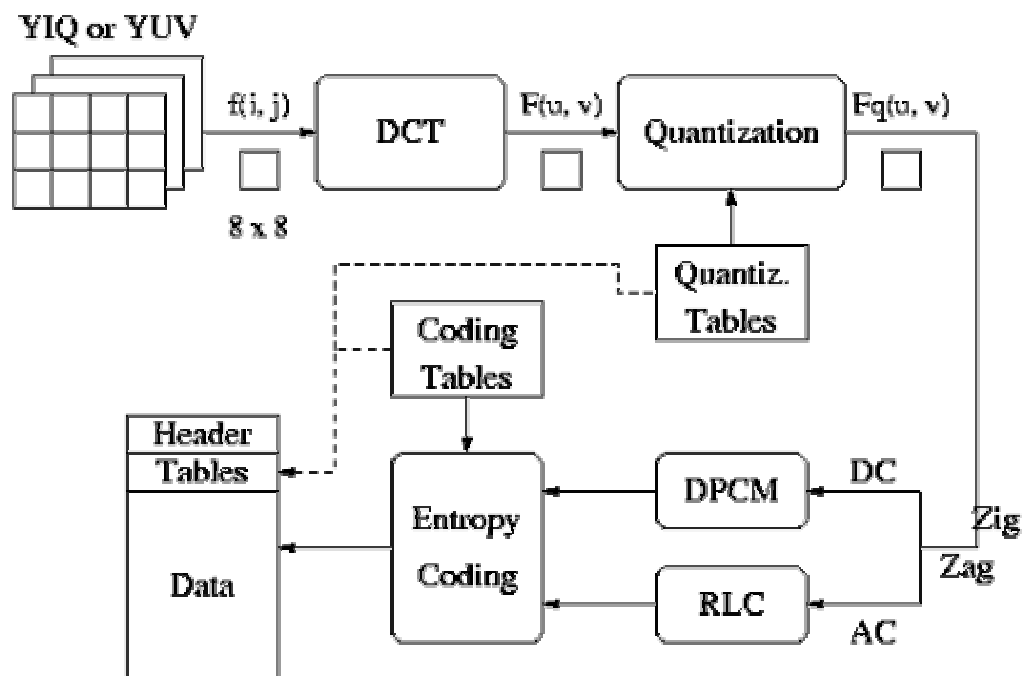
באשר ליתר המקדמים, אלא מאורגנים בסריקת זיג-זג כמתואר בצוור

הבא:



סריקת זיג-זג של בלוק מקדמי ההתמרה ב-JPEG.

מסתבר כי הווקטור המתקבל כולל אפסים רבים ורצוי לנצלם. הגישה שיושמה בסטנדרט ה-JPEG היא שיטת run-length encoding בה הווקטור מומר לצמדי מספרים – אורך קפיצה וערך מתקבל. מתחילים באיבר הראשון בווקטור זה באורך 63 אלמנטים, קובעים את הערך שישנו במקדם הראשון ומספר האיברים למקדם השונה מאפס הבא. במקדם הבא השונה מאפס מתעדים את ערכו ומספר האיברים שיש לקפוץ לשלם הגעה לערך השונה מאפס הבא, וכך הלאה. קוד הופמן מיוחד שתוכנן מראש לשני סוגי מספרים אלו (ערך וקפיצה) מעבירים לסיביות בשיטה של קוד בעל אורך מילה משתנה, וישנו קוד מיוחד השמור לאמירה "סוף הבלוק" כדי לא לספר את כולו אם היתר אפסים.



סכימת בלוקים של דוחס ה-JPEG.

האם זהו האלגוריתם המושלם? מה שגוי בשיטת ה-JPEG שיכול להיות הפוטנציאל לשיפור בשיטות עתידיות? שאלה זו הטרידה חוקרים רבים שחיפשו דרכים לקבלת אלגוריתם דחיסה טובים יותר. בעיה אחת שמוסכמת הינה גישת הקירוב הליניארי. גם אם נניח כי יש להשתמש בהתמרה ליניארית כדרך לביצוע הדחיסה, מדוע בתכנון ההתמרה האופטימלית (KLT) או חיקוייה (DCT) או הדמרד) יש לבחור את המקדמים הראשונים ולא את החזקים באשר הם? שינוי גישה זו קרוי קירוב לא ליניארי כי המקומות בהם נלקחים המקדמים הינו פונקציה לא ליניארית של הכניסה.

מרכיב אחר שאותר כמגביל הינו עניין הרזולוציה בתמונה. הבחירה בבלוקים בגודל קבוע (8 על 8 פיקסלים) מתעלם לחלוטין מתוכן התמונה. בנתחי תמונה בהן ישנה התנהגות חלקה יותר, ראוי לבחור בלוקים גדולים יותר. לכן נראה כי גישה בה גודל הבלוק יהיה אדפטיבי לתוכן באיזשהו אופן תוכל להועיל. למעשה, אם להיות הוגנים, במקום מסוים בתמונה יש תכנים ממגוון "אורכי-גל" – מידע המשתנה איטי ועל גביו מידע המשתנה מהר. נראה כי בכל מקום בתמונה נכון היה להפעיל התמרה מולטי-רזולוציונית הרואה את כל אלה במשותף, כאילו הותמרו בלוק בגדלים שונים באותו מקום. כל זה הוביל לצעד הבא – התמרות ה-wavelet.

עיבוד תמונות בעזרת מחשב (236327)

הרצאה מס' 14

התמרת wavelet ודחיסה בעזרתה

התמרת Haar ומבוא ל-Wavelet

מבין כל המטריצות היוניטריות שיכולות לשמש אותנו כהתמרה, ראינו בתחילה מטריצה בה הערכים מרוכבים (DFT), שיפרנו ע"י מציאת התמרה ממשית (DCT), והקלנו את המורכבות החישובית לאין ערוך ע"י שימוש במטריצה בה האיברים הם $+1$ ו- -1 . האם ניתן להמשיך במגמת הפשטה זו? התשובה היא כן! נציע כעת התמרה אשר גם בה הערכים $+1$ ו- -1 , אך מרבית איברי מטריצת ההתמרה יהיו אפס, ובכך נשיג התמרה הדורשת פחות חישובים. התמרה זו היא התמרת Haar. נתחיל בהצגתה של התמרה זו לחד מימד ונראה כיצד ממנה נובעת התמרה דו-ממדית.

חשוב להבהיר כי יתרונה של התמרת Haar אינו מתמצא בהיותה בעלת מטריצת התמרה דלילה (בה יש הרבה אפסים). התמרה זו מהווה מקרה פרטי פשוט של משפחה רחבה של התמרות הקרויות Wavelet, אשר נמצאו להיות יעילות בשל מגוון סיבות לטיפול באותות טבעיים. במסגרת זו נציג את רק את היסודות לתחום רחב ועשיר זה.

כמקודם, נתחיל בתיאור ההתמרה החד-ממדית לווקטור באורך M זוגי. נתחיל במטריצה בסיסית בגודל 2 על 2. מטריצה זו זהה להתמרת Hadamard:

$$D_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \Rightarrow D_1^{-1} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

התמרה זו תקיפה לוקטורים באורך 2, וניכר כי היא יוניטרית. משמעות התמרה זו היא שבהינתן שני מספרים a ו- b הבונים את הווקטור הנכנס להתמרה, ההתמרה תהיה:

$$\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} a+b \\ a-b \end{bmatrix} = \begin{bmatrix} c \\ d \end{bmatrix} \Rightarrow \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} c \\ d \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} c+d \\ c-d \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix}$$

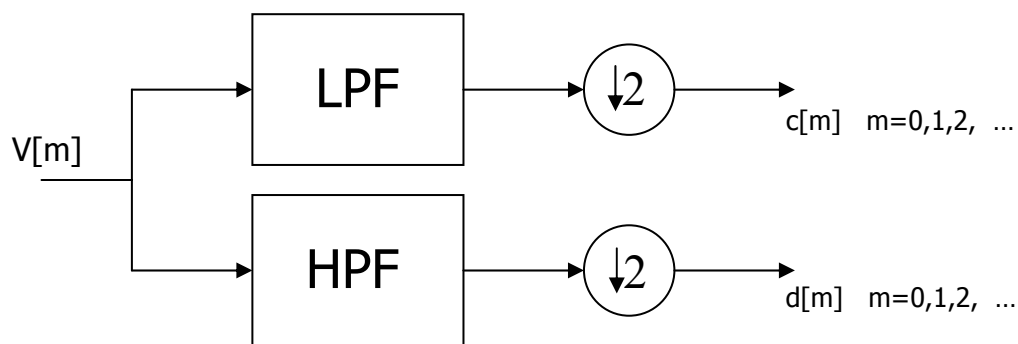
כלומר, צמד המספרים הוחלף בהתמרה בסכומם ובהפרשם (ובהכפלה בקבוע הנובע מרצוננו בהתמרה יוניטרית). ההתמרה ההפוכה ביצעה אותו דבר בדיוק.

בהינתן סידרה בת M (מספר זוגי!) איברים, נוכל לחלקה לזוגות סמוכים, ולהמיר כל צמד בסכומם והפרשם כפי שבוצע לצמד הקודם. אם נסמן את סידרת הכניסה ב- $v[m]$ ($m \geq 0$), נקבל את סידרת המספרים הבאים:

$$c[m/2] = \frac{1}{\sqrt{2}} [v[m] + v[m+1]] \quad m = 0, 2, 4, \dots, M-1$$

$$d[m/2] = \frac{1}{\sqrt{2}} [v[m] - v[m+1]] \quad m = 0, 2, 4, \dots, M-1$$

וכך התקבלה סידרת מספרים בת M איברים המייצגת באופן מלא את סידרת המקור. למעשה, המערכת המתוארת בציור הבא שקולה לתהליך שתואר.



מערכת סינון ודילול ליצירת הסדרות c ו d .

במערכת זו, ערוץ ה- LPF מחשב ממוצעים מקומיים לפי:

$$c[m] = \frac{1}{\sqrt{2}} [v[m] + v[m+1]] \quad m = 0, 1, 2, 3, 4, \dots, M-1$$

אך כיוון שאנו נזקקים רק לזוגיים (או אי-זוגיים) שבהם, ישנה פעולת דילול ביחס 2, שזורקת את כל התוצאות במקומות הזוגיים (או אי-זוגיים – נדרשת עקביות). באופן דומה, ערוץ ה- HPF מחשב הפרשים מקומיים לפי:

$$d[m] = \frac{1}{\sqrt{2}} [v[m] - v[m+1]] \quad m = 1, 2, 3, 4, \dots, M-1$$

ושוב דילול הוא זה שמסיר מחצית המספרים. על מנת לקבל את הסדרות c ו d ולחזור למקור $v[m]$, עלינו לקחת כל צמד של ערכים מתאימים, ולבצע עליהם

סיכום והפרש, כמקודם, ולשלב את הערכים לפי הסדר הנכון. ניתן לראות כי סידרת הפעולות הבאה משיגה אותו אפקט:

1. קח'י את שתי הסדרות c ו-d והכנס'י אפס בין כל שתי ערכים שלהן (כך שאיבריהן הזוגיים אפסים).

2. העבר'י את הסדרה המוגדלת c דרך LPF זהה לזה בו נעשה שימוש קודם.

3. העבר'י את הסדרה המוגדלת d דרך HPF זהה לזה בו נעשה שימוש קודם.

4. חסר'י את התוצאות, וקבל את v.

נמחיש תהליך זה דרך דוגמה: ניקח את הסדרה הבאה ונפעיל עליה את התהליכים שתוארו.

$$\underline{v}=[1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7 \ 8 \ 9 \ 10 \ 11 \ 12 \ 13 \ 14 \ 15 \ 16]$$

הסדרות c ו-d תתקבלנה ע"י

$$\begin{aligned} c &= \frac{1}{\sqrt{2}} [(1+2) \ (3+4) \ (5+6) \ (7+8) \ (9+10) \ (11+12) \ (13+14) \ (15+16)] = \\ &= \frac{1}{\sqrt{2}} [3 \ 7 \ 11 \ 15 \ 19 \ 23 \ 27 \ 31] \end{aligned}$$

$$\begin{aligned} d &= \frac{1}{\sqrt{2}} [(1-2) \ (3-4) \ (5-6) \ (7-8) \ (9-10) \ (11-12) \ (13-14) \ (15-16)] = \\ &= \frac{1}{\sqrt{2}} [-1 \ -1 \ -1 \ -1 \ -1 \ -1 \ -1 \ -1] \end{aligned}$$

בסדרות אלו לקחנו כל זוג של מספרים והמרנו אותו בסכום (ב-c) ובהפרש (ב-d) כשאנו זורקים את האיברים הזוגיים בתהליך הסינון המלא. נבצע כעת את תהליך השחזור המוצע. נתחיל בלקיחת הסדרות ומילואן באפסים כשמתחילים באפס:

$$c = \frac{1}{\sqrt{2}} [0 \ 3 \ 0 \ 7 \ 0 \ 11 \ 0 \ 15 \ 0 \ 19 \ 0 \ 23 \ 0 \ 27 \ 0 \ 31 \ 0]$$

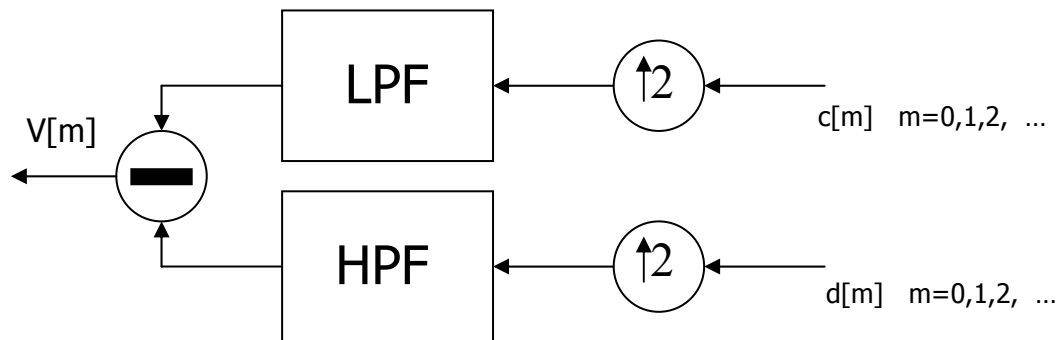
$$d = \frac{1}{\sqrt{2}} [0 \ -1 \ 0 \ -1 \ 0 \ -1 \ 0 \ -1 \ 0 \ -1 \ 0 \ -1 \ 0 \ -1 \ 0 \ -1 \ 0]$$

הפעלת LPF ו-HPF על שתי סדרות אלו תיתן:

$$\begin{aligned} \text{LPF}\{c\} &= \frac{1}{2} [(0+3)(3+0)(0+7)(7+0) \ \dots \ (0+31)(31+0)] = \\ &= \frac{1}{2} [3 \ 3 \ 7 \ 7 \ 11 \ 11 \ 15 \ 15 \ 19 \ 19 \ 23 \ 23 \ 27 \ 27 \ 31 \ 31] \end{aligned}$$

$$\begin{aligned} \text{HPF}\{d\} &= \frac{1}{2} [(0+1)(-1-0)(0+1)(-1-0)(0+1) \ \dots] = \\ &= \frac{1}{2} [+1 \ -1 \ +1 \ -1 \ +1 \ -1 \ +1 \ \dots] \end{aligned}$$

וחיסור שתי סדרות אלו ייתן את סידרת המקור בדיוק. גם את תהליך השחזור ניתן לתאר כסכימת בלוקים הדומה לציור הנ"ל. סכימת השחזור מתוארת בציור הבא.



מערכת אינטרפולציה וסינון לשחזור v מהסדרות c ו d .

כיוון שבכל התהליך הנ"ל בוצעו פעולות ליניאריות על איברי $v[m]$, נוכל לייצגן כמכפלת הווקטור $\underline{v}$ במטריצה. עבור וקטור בעל ארבעה איברים, מטריצה זו תהיה הבאה:

$$D_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

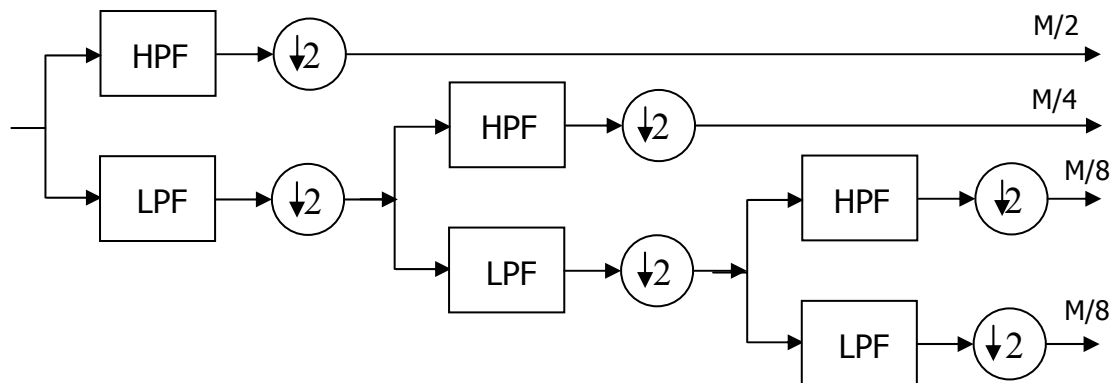
כך ששתי השורות הראשונות בונות את הממוצעים (LPF) ושתי השורות האחרונות בונות את ההפרשים (HPF). במטריצה זו רק חצי מהאיברים שונים מאפס. באופן דומה, לווקטור בעל שמונה איברים נקבל:

$$D_3 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix}$$

וקל לראות כי מטריצה זו יוניטרית, ובה רק רבע מהאיברים שונים מאפס. חשוב להבהיר - זו עדיין לא התמרת Haar.

עד כה התמקדנו במרכיב החשוב הראשון של ההתמרה - ההפרדה לסדרות LPF ו-HPF מדוללות. נציג כעת את המוטיב השני המרכזי והוא נושא המולטי-רזולוציה.

הרעיון בעיקרו פשוט: את סידרת ההפרשים d נותיר ללא שינוי, ועל סידרת הממוצעים c נבצע אותו תהליך כמקודם, כלומר בניית שתי סדרות של ממוצעים והפרשים עבודה. ברור כי כל סידרה כזו כבר תהיה באורך רבע מהאורך המקורי, וסך כל האיברים יסתכם ל- M . תהליך זה יחזור רקורסיבית כל פעם על סידרת ה-LPF, עד לקבלת סדרות באורך 1. התוצאה היא התמרת Haar. סכימת בלוקים של דרך בנייתה מתוארת בציור הבא.



התמרת Haar כפעולות סינון ודילול

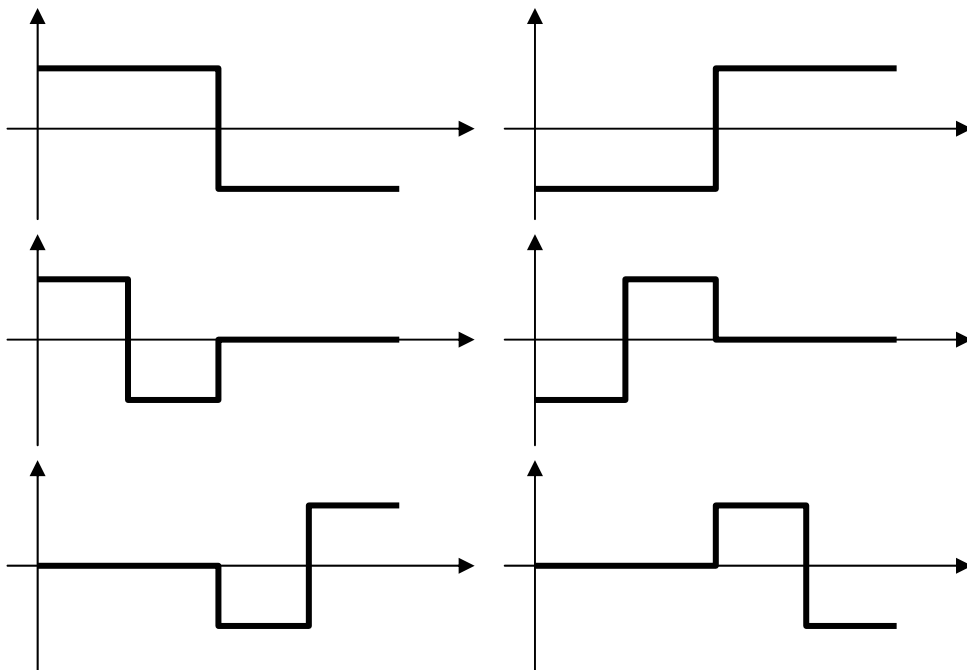
לדוגמה, התמרת וקטור בעל 4 אלמנטים יתבצע ע"י המטריצה הבאה:

$$W_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 & 0 \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ \sqrt{2} & -\sqrt{2} & 0 & 0 \\ 0 & 0 & \sqrt{2} & -\sqrt{2} \end{bmatrix}$$

המטריצה הימנית במכפלה היא המטריצה שנכתבה קודם. מטריצה זו מוכפלת במטריצה שמותירה את שתי השורות התחתונות ללא שינוי, ומבצעת התמרה על שני האיברים הראשונים כפי שתואר קודם. ניתן באופן דומה לבנות את המטריצה המתייחסת להתמרת וקטור באורך 8 אלמנטים:

$$W_3 = \frac{1}{\sqrt{8}} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ \sqrt{2} & \sqrt{2} & -\sqrt{2} & -\sqrt{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sqrt{2} & \sqrt{2} & -\sqrt{2} & -\sqrt{2} \\ 2 & -2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & -2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & -2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & -2 \end{bmatrix}$$

תכונה מעניינת וחשובה של מטריצת ההתמרה היא ששורותיה (למעט הראשונה) מהוות תולדה של אותו גל המעוות ע"י הזזה ומתיחה. גלים אלו נקראים "גלונים" וזהו התרגום לשם Wavelet. הציור הבא מראה גלים אלו כגרפים.



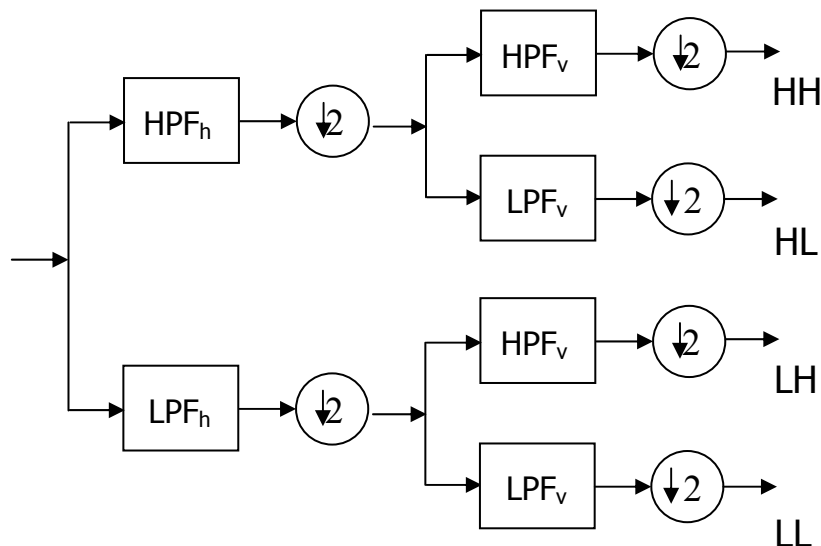
פונקציות הבסיס של התמרת Haar החד-מימדית כפונקציות הנובעות מאותו מקור עם הזזות וכיווץ\מתיחה.

אם נסתכל לרגע במטריצה W שהתקבלה, נזהה את אלמנט ה-DC שהינו ממוצע האות. זהו המקדם הראשון בהתמרה. כל יתר השורות מתארות פעולת "גזירה" או פעולת מסנן העברת גבוהים אך במרווחים שונים. פעולות אלו מתחילות במרווח גדול (מתייחס לתדרים נמוכים יחסית) ועד למרווח קטן של 2

אלמנטים (תדרים גבוהים). כמות הנקודות השמורה בכל פס תדר פרופורציונית הפוך למרווח הנ"ל. לכן יש לנו כאן פירוק תדרי-מקומי עם תפיסה מולטי-רזולוציונית.

באשר לדו-מימד, כמו בכל המקרים הקודמים, ההתמרה על תמונה בגודל M על N מתקבלת ע"י הפעלתה באופן ספרבילי. זוהי התמרה יוניטרית קלה במיוחד הניתנת לביצוע ע"י $O\{MN\}$ פעולות (בניגוד לכל ההתמרות הקודמות הדורשות $O\{MN \cdot \log\{MN\}\}$). ביצוע ההתמרה בדרך יעילה נעשה ע"י מערכת הסינונים והדילולים שתוארה קודם לכן.

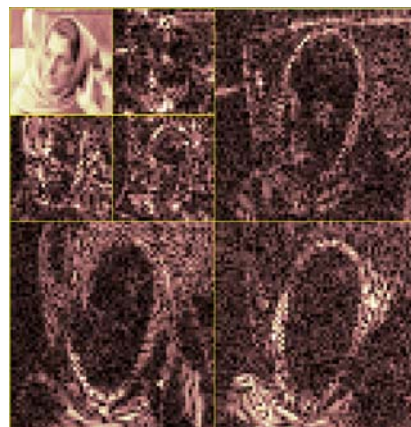
בבלוק זה, לאחר ביצוע סינון LPF ו-HPF על שורות (ולכן הסימון h- Horizontal), כל מוצא עובר גם סינון LPF ו-HPF על עמודות. אפקטיבית, מתקבל כי תחום התדר מתחלק לארבעה אזורים שווים המסומנים ב- LH, HL, LL ו- HH. באזור ה- LL המתייחס לתדרים הנמוכים מתבצעת שוב חלוקה לארבעה אזורים וחוזר שוב ושוב. הציור הבא מראה כיצד בסופו של דבר נראית החלוקה להתמרה בת שלושה שלבים (המתייחסת לתמונות בגודל 8 על 8).



LL	LH	LH	LH
HL	HH		
HL		HH	
HL			HH

פסי התדר בהתמרת Haar

אם נתייחס לבלוק LH הגדול, בלוק זה נבנה ע"י סינון תמונת המקור במסנן LPF אופקי ו- HPF אנכי. כיוון שכעת התוצאה ממלאת רק רבע תחום בתדר, היא מדוללת ביחס 2:1 בכל ציר (כלומר צורכת רבע מכמות הפיקסלים בתמונת המקור). לכן, לתמונה של 8 על 8, בלוק זה מוחזק במטריצה של 4 על 4. באופן דומה, בלוק ה- LH הפנימי יותר מוחזק ע"י 2 על 2, וכו'.

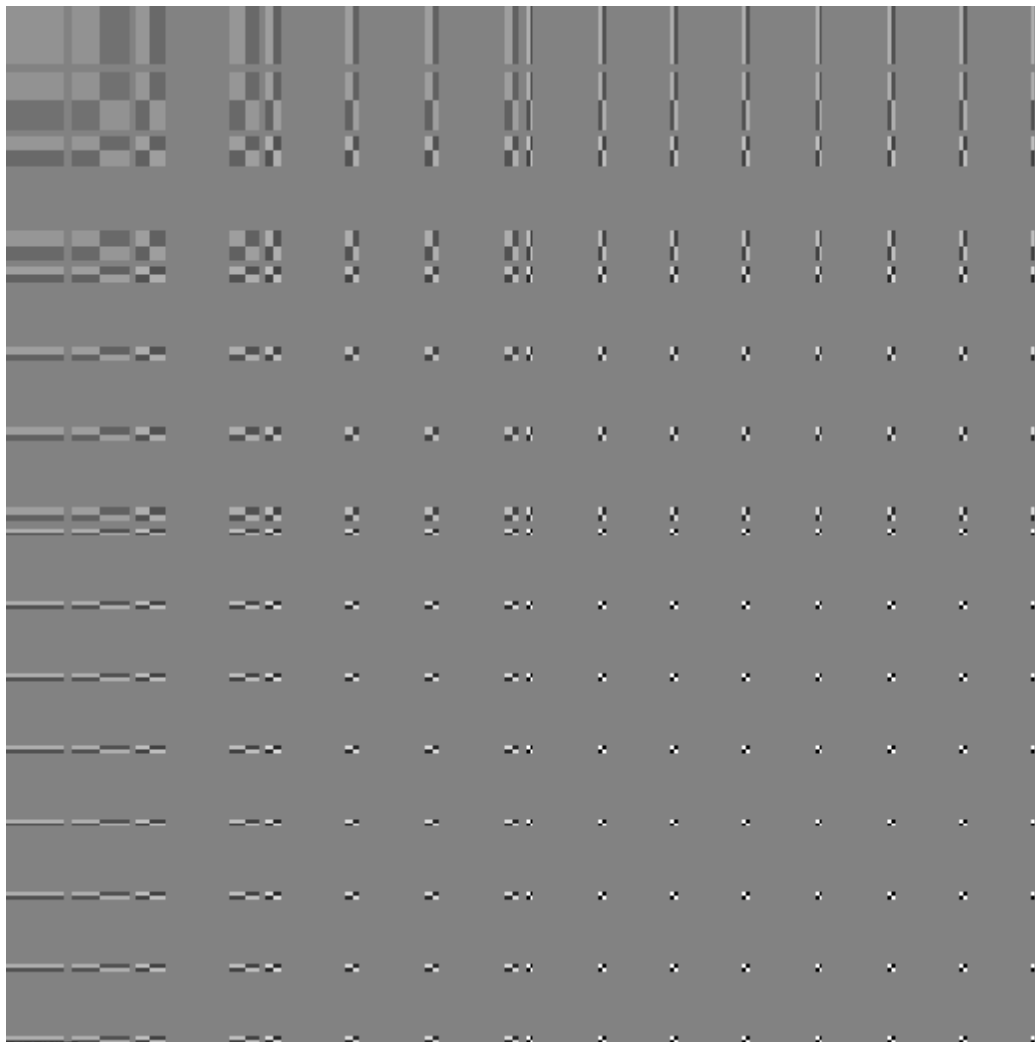


פירוק אופייני ע"י Wavelet.

פונקציות הבסיס של התמרה דו-מימדית זו מתוארים בציור הבא. ניכר כי פונקציות אלה שונות בתכלית מכל ההתמרות אותן פגשנו קודם. הייחוד בפונקציות בסיס אלה הוא עובדת היותן דלילות, כלומר תמונות בהן מעט

איברים שונים מאפס. לדבר זה קשר ישיר להיות פונקציות הבסיס מקומיות. למעשה, אנו רואים כאן גלונים (Wavelet) הדומים זה לזה עד כדי מיקום וגודל.

בכל האמור עד כה התמקדנו בהתמרת Haar. התמרה זו מהווה מקרה פרטי של משפחת התמרות ה-Wavelet. בהתמרת Wavelet כללית מוחלף המסנן [1-1] אשר שימש ל-HPF במסנן אחר, ארוך יותר בדרך כלל, וכזה שיכול לבצע הפרדה טובה יותר של פסי התדר. מסנן זה הנקרא "פונקצית האם" יוצר את כל מאפייני ההתמרה, וממנו נגזרות כל פונקציות הבסיס של ההתמרה. תכנון מסנן HPF כזה שייתן כי ההתמרה הכוללת היא אורתו-נורמלית אינה בעיה קלה כלל, וישנן פונקציות מסוימות ידועות היכולות להוות פתרון.



פונקציות הבסיס של התמרת Haar לתמונות בגודל 16 על 16

לסיכום, התמרת Haar שונה באופייה מההתמרות הקודמות בהן עסקנו. זוהי התמרה יוניטרית וממשית. התמרה זו ניתנת לביצוע מהיר (ב- $O\{M\}$ פעולות) בשל היותה דלילה. בנוסף לקלות ביצועה הנומרי, יש לה תכונה אטרקטיבית חשובה שלא נידונה כאן - היכולת לטפל באותות המשנים את אופיים הספקטרלי במקום באופן נאות.

למה התמרת Wavelet טובה לתיאור תמונות?

ישנן מספר גישות להצדקת בחירת התמרת ה-wavelet לשימוש בתמונות. הבסיסית שבהן באה מתורת הקירובים, בה נדונה אפשרות של שחזור מקירוב לא ליניארי – נתחיל עם תיאור בסיסי של הצדקה זו.

כשהצגנו את התמרת ה-DCT כבחירה מוצלחת לייצוג תמונות, עשינו זאת כשאנו נשענים על דמיונה הרב להתמרת KLT. מתוך ההבנה שהתמרת KLT היא ההתמרה הטובה ביותר במובן של ריכוז אנרגיית האות למקבץ מקדמים קטן, הסקנו כי שימוש ב-DCT צריך להיות מוצלח בתיאור תוכן התמונה.

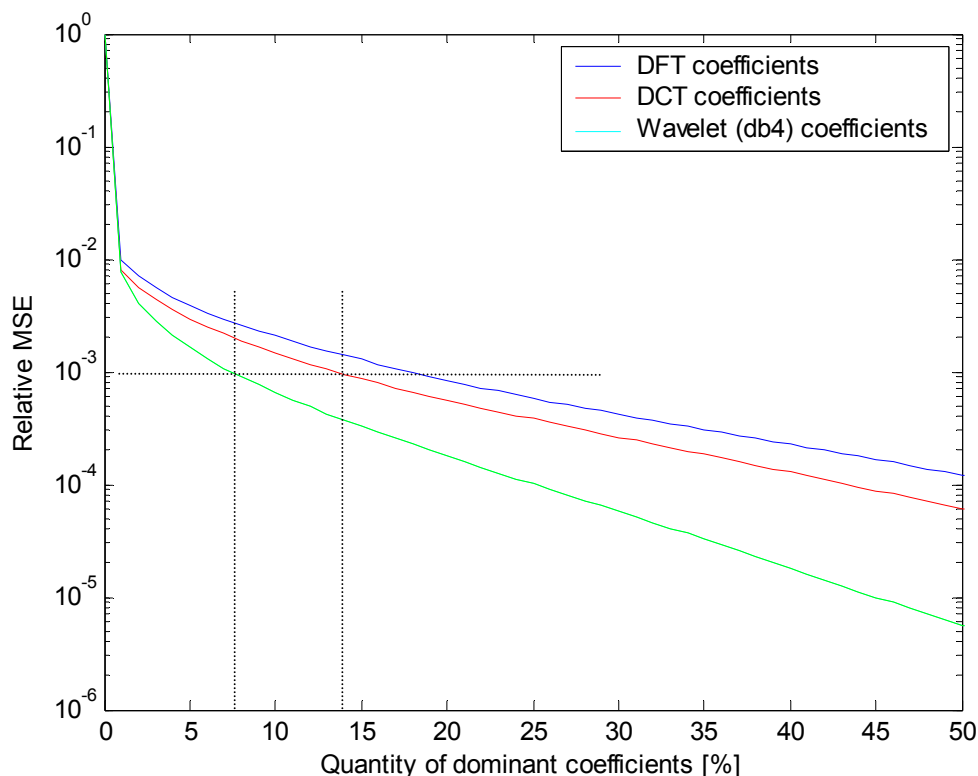
כל שזה נכון אבל הכל היה מבוסס על תפיסה ליניארית לפיה אנו מבצעים את ההתמרה, קוצצים את זנב ווקטור המקדמים (אוסף המקדמים בסוף הווקטור) ומבצעים התמרה הפוכה לשחזור האות. תהליך זה ליניארי מקצה לקצה אמנם כך תיארנו אותו כמשוואה

$$\hat{x} = T^{-1} I_n T x$$

כש-T היא ההתמרה הנדונה, ו- I_n מטריצה אלכסונית בה n איברי האלכסון הראשונים הם '1' והיתר '0'.

מדוע שנבחר את המקדמים הראשונים? מדוע שלא נבחר את המקדמים הדומיננטיים? בגישה זו, שאינה ליניארית, ניקח את מקדמי ההתמרה ונבחר את n הגבוהים שבהם ומהם נבנה שחזור. ברור כי התוצאה תהיה טובה יותר במובן של שגיאת השחזור כי לקחנו מקדמים חשובים יותר. מאידך, ברור שלייצוג מקדמים אלו נדרש לא רק לערכיהם אלא גם לכתובתם, ולכן ליותר סיביות ייצוג בסך הכל. לכן יש לשאול האם גישה אלטרנטיבית זו יעילה יותר בסופו של דבר.

קשה לתכנן התמרה שתהיה אופטימלית (כמו KLT) בקירוב לא ליניארי כמתואר לעיל. עם זאת, נוכל לבחון בפועל את ביצועי ההתמרות המוכרות לנו – DFT, DCT, והתמרת wavelet – נבצע את ההתמרה, ניקח מספר משתנה של מקדמים דומיננטיים, ונבצע שחזור. ההגרף הבא מראה את התוצאה המתקבלת עבור התמונה lena, וניכר כי מקדמי התמרת ה-wavelet עשו כאן עבודה טובה יותר. לשם דוגמה, להגעה לשגיאה יחסית של 10^{-3} נדרשים כמעט כמחצית ממקדמי ה-wavelet בהשוואה למקדמי DCT.



שחזור תמונה ממספר מקדמים חלקי – בחירת מקדמים דומיננטיים.

הנה תוכנית ה-Matlab שיצרה גרף זה:

```
X=imread('lena.jpg'); X=double(X);
% FFT
Xf=fft2(X); Xfs=flipud(sort(abs(Xf(:)))));
T=Xfs(round(512^2/100:512^2/100:512^2));
for k=1:1:100,
    Xf0=Xf.*(abs(Xf)>T(k));
    Erf(k)=mean(mean(abs(iff2(Xf0)-X).^2));
end;
```

```

figure(1); semilogy(0:1:100, [1, Erf/mean(mean(X.^2))]);
% DCT
Xf=dct2(X); Xfs=flipud(sort(abs(Xf(:))));
T=Xfs(round(512^2/100:512^2/100:512^2));
for k=1:1:100,
    Xf0=Xf.*(abs(Xf)>T(k));
    Erd(k)=mean(mean(abs(idct2(Xf0)-X).^2));
end;
hold on; semilogy(0:1:100, [1, Erd/mean(mean(X.^2))], 'r');
%Wavelet
[Xf,t]=wavedec2(X,9,'db4');
Xfs=flipud(sort(abs(Xf(:))));
T=Xfs(round(512^2/100:512^2/100:512^2));
for k=1:1:100,
    Xf0=Xf.*(abs(Xf)>T(k));
    Erw(k)=mean(mean(abs(waverec2(Xf0,t,'db4')-X).^2));
end;
semilogy(0:1:100, [1, Erw/mean(mean(X.^2))], 'c');
axis([0 50 1e-6 1]);

```

בתורת הקירובים (Approximation theory) ישנו דיון מעמיק בשאלה שתיאור אמפירי שלה נתנו לעיל. הגישה היא מתמטית צרופה והיא הולכת לפי הקו הבא: בהנחה שהתמונה הנדונה באה ממשפחת האותות החלקים למקוטעין (עם תיאור מתמטי מדויק של תכונה זו), כיצד ידעו המקדמים עבור גודל תמונה הולך לאינסוף? מוכח כי מקדמי הפורייה דועכים לפי שורש מספר המקדמים בעוד מקדמי ה-wavelet דועכים באופן פרופורציוני ישיר למספר המקדמים, ומכאן שזה ייצוג יעיל יותר.

שאלה מעניינת שלא נעסוק בה כאן היא האם זוהי ההתמרה הטובה ביותר שניתן להשיג? מסתבר כי התשובה היא שלילית! ההתמרה האופטימאלית צריכה לתת דעיכה לפי ריבוע מספר המקדמים, ומסתבר אף כי ניתן לתכנן התמרה כזו, אך אנו לא ניכנס לנושא זה שהינו חדש למדי ועם תוצאות שאינן מבוססות דיין בינתיים.

נתנו סיבה טובה לבחירת התמרת ה-wavelet ובה גם ביסוס מתמטי טוב. נראה עתה סיבה אחרת פשוטה יותר ואינטואיטיבית שלמעשה מסבירה את התוצאה לעיל.

כשאנו מפעילים התמרת DCT בבלוקים באלגוריתם ה-JPEG, אנו בוחרים בלוקים בגודל 8-על-8 פיקסלים. כתוצאה מכך אנו מפסידים את הקורלציה בין גבולות הבלוקים המהווים 25% מכלל שטח התמונה. לו הגדלנו את גודל הבלוק היינו מנצלים יותר מקורלציות אלו. לעומת זאת, עם הגדלת גודל הבלוק הקירבה בין ה-DCT ל-KLT הייתה נופלת, ואיתה יעילות הייצוג. זאת מכיוון שלתוכן התמונה בבלוקים גדולים של פיקסלים לא נכונה ההנחה הגורפת של מודל מרקובי מסדר ראשון.

לו לקחנו את תוכן התמונה ופיצלנו אותו לפסי תדר (כמו בבניית הפירמידה הלפלסיאנית) לתדרים הנמוכים הנחת המרקוביות הייתה ממשיכה להיות נכונה גם לבלוקים גדולים, ואילו לפסי התדר הגבוהים הנחה זו הייתה נשברת. כך יכולנו לנצל הקורלציות יותר טוב כשאנו מפעילים גודל בלוק אפקטיבי שונה לתכולה תדרית שונה – זהו בדיוק האפקט שהתמרת ה-wavelet נותנת.

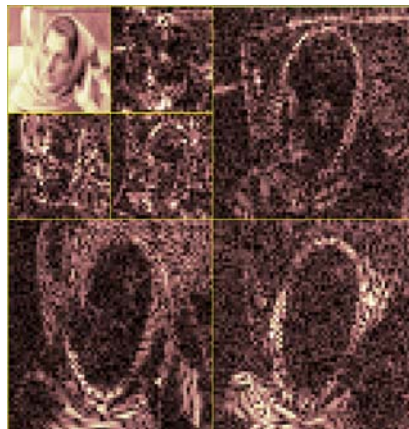
דחיסת תמונות מבוססת Wavelet

אלגוריתם ה-JPEG יוצב בשנות ה-80 והפך לסטנדרט פופולארי בו השימוש נפוץ במצלמות דיגיטאליות, אגירת תמונות באינטרנט, ועוד. במהלך שנות ה-90 פותחו במקביל אלגוריתמי דחיסה חדשים מבוססי wavelet אשר הראו ביצועים טובים לאין ערוך מאלו של ה-JPEG. ועדות התקינה ש-"קידשו" את ה-JPEG הבינו כי עליהן להציע סטנדרט חדש, ואמנם ב-2001 הוצע אלגוריתם דחיסה חדש מבוסס wavelet הקרוי JPEG2000. נראה את העקרונות הבסיסיים בדחיסה מבוססת wavelet.

הבסיס צריך להזכיר לנו מאוד את דחיסת התמונה מבוססת הפירמידה הלפלסיאנית. למעשה התמרת ה-wavelet בונה פירמידה של רזולוציות בה הקודקוד (הרזולוציה הגסה ביותר) היא תמונה של ממש אך מאוד קטנה. משם אנו יורדים בשכבות הפירמידה ומקבלים שלשות של תמונות בגודל זהה בהם

ישנם פרטים – ניתן להתייחס אליהן כאל שגיאות חיזוי של התמונה הקטנה לו הוגדלה. שלושת השכבות נותנות את השגיאות האופקית, האנכית, והאלכסונית, לפי מסנני ה-wavelet (LH, HL, ו-HH). אנו יורדים בשכבות הפירמידה ומקבלים עוד שלשות כאלה, ובסיום משלימים כמות זהה בדיוק של מקדמי ייצוג כמספר הפיקסלים בתמונת המקור. מכאן שגישה זו צפויה להיות הרבה יותר יעילה משיטת הפירמידה הלפלסיאנית כי אין כאן יתירות.

לכן, גישה דחיסה מתבקשת היא קוונטיזציה של מקדמי ההתמרה ממש כמקודם, ואז בפריסת התמונה מבצעים התמרה הפוכה על מקדמי התמרה פגומים במקצת. זהו הרעיון הבסיסי והוא פועל יפה – טוב יותר מ-JPEG.



פירוק אופייני ע"י Wavelet.

נקודה חשובה שלא טרחנו להבהיר כאן במיוחד, לכל אורך הדיון על דחיסה היא הנקודה הבאה: בדחיסת תמונות ומידע בכלל ישנו מרחק ניכר בין רעיון בסיסי מוצלח לבין יישום מוצלח. כשנכנסנו לתוך נבכי שיטת ה-JPEG גילינו המון "טריקים" ורעיונות שכל מהותם היא סחיטת סיביות ומיצוי כל יתרות אפשרית – כל זה מעבר לרעיון השלדי של התמרה וקוונטיזציה. כך גם הדבר כאן ב-JPEG2000. קוונטיזציה על מקדמי ה-wavelet הינה רעיון ראוי, אך אין די בו. אם נסתכל בתמונה של מקדמי ההתמרה נגלה כי יש קשר בין השכבות ברזולוציות השונות ובתוך אותה רזולוציה. ניצול קשר זה בתהליך הדחיסה יועיל היטב לאלגוריתם הדחיסה.

גישה שזכתה לתשומת לב גבוהה היא שיטה הקרויה EZW בה עושים שימוש בשיטה הקרויה zero-tree. נתאר אותה בקצרה על מנת להמחיש כיצד סוחטים את מקדמי התמרת ה-wavelet היטב. יוצרי ה-EZW עמדו על האבחנות הבסיסיות הבאות:

- מקדמי ההתמרה נוטים להיות חזקים בפסי התדר הנמוכים, ומראים באופן כללי דעיכה עם העלייה לתדרים גבוהים (שכבות הרזולוציה העדינה בפירמידה – השלשות בהן יותר פיקסלים).
 - כמו כן ברור כי מקדמים גדולים האם אלו שציריכם להיות מטופלים היטב – מהקטנים ניתן להיפטר.
- שני רעיונות בסיסיים עומדים בבסיס שיטת ה-zero-tree. הראשון הוא קוונטיזציה פרוגרסיבית, והאחר הוא קישור מקדמים במבנה עץ. רעיונות אלו משתמשים באבחנות הנ"ל לסחיטה מוצלחת של מקדמי ה-wavelet.

רעיון הפרוגרסיביות הינו רעיון בסיסי הקיים גם בשיטות דחיסה אחרות, אך לא העלנו אותו עד כה. גם אחרי דחיסה מתקבל קובץ לא קטן ושידורו יכול להיארך למשך שניות\דקות. בהינתן הסיביות שנקלטו עד כה, נרצה לבנות תמונה חלקית על סמך המידע הקיים. יהיה זה נפלא אם אחרי מקבץ חלקי מאוד נוכל כבר לקבל רושם כללי על התמונה המתקבלת. כיצד נוהגות השיטות השונות בהיבט זה?

- בגישת ה-DPCM נוכל לבנות את התמונה שורה אחר שורה, ולהראות את השורות שהגיעו עד כה. גישה זו לא מוצלחת מההיבט הפרוגרסיבי.
- בשיטת סורק-זאבי הבעיה דומה אך התמונה תיבנה בעקלתון מבוסס סריקת ההילברט – גם לא מוצלח.
- השיטות הפירמידליות נוחות מאוד. כבר אחרי מקבץ קטן של פיקסלים נוכל לראות תמונה מלאה אם כי חסרת פרטים עדינים – זוהי בדיוק הפרוגרסיביות במהותה.
- שיטת ה-JPEG במקורה לא אמורה להיות פרוגרסיבית. אנו שולחים כל בלוק בנפרד. עם זאת, בהינתן כל בלוקי התמונה לאחר קוונטיזציה בדוחס, נוכל לארגן את הקובץ הדחוס כך שכל מקדמי ה-DC יקובצו תחילה. כך תתקבל במקלט פרוגרסיביות נעימה בה כל בלוק מיוצג ע"י ממוצעו. לאחר מכן נשלח את המקדם השני בכל בלוק וכל בלוק

במפענח יקבל עידון-מה של תוכנו. כך יכול התהליך להמשיך, כשכל שנדרש הוא ארגון מקדמי ההתמרה בדרך מתאימה.

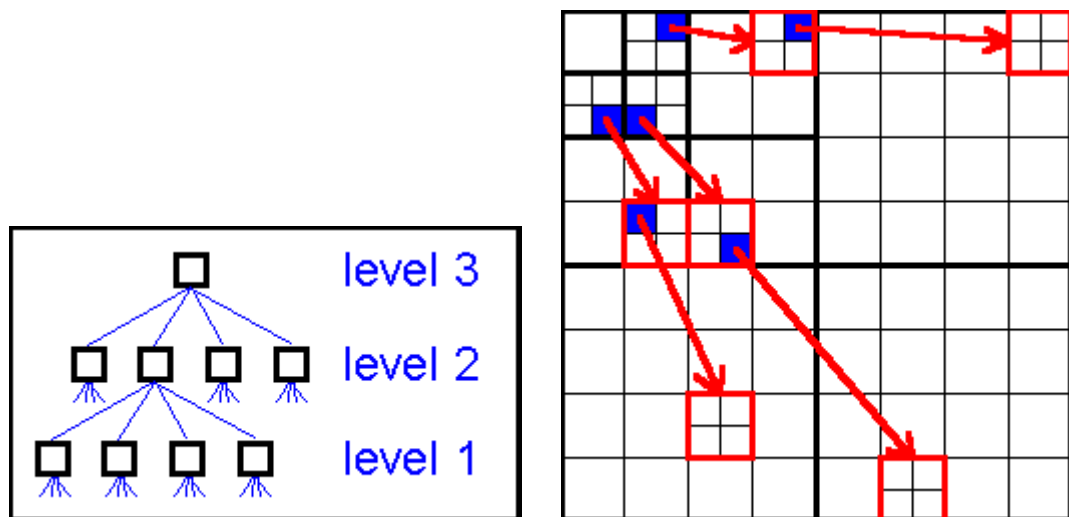
- שיטת ה-JPEG2000 מביאה פרוגרסיביות באופן טבעי בשל היותה בעלת מבנה פירמידלי. נתחיל מקודקוד הפירמידה בה הרזולוציה גסה, ונעדר עם פרטים במורד הפירמידה.

הפרוגרסיביות שקיימת ב-EZW מוסיפה נדבך נוסף של פרוגרסיביות מעבר לניצול הברור של המבנה הפירמידלי. נקבע סף כלשהו ונבחן מיהם מקדמי ההתמרה שגבוהים (בערך מוחלט) מסף זה. רק לאלה נשלח מידע מתוך הנחה שהיתר אפסים. המידע שיישלח יהיה חלקי – קוונטיזציה גסה של המקדם ולא יותר. במקלט כבר ניתן יהיה להציג תמונה על סמך מידע זה שיהיה גם ברזולוציה וגם בדיוק הערכים.

בשלב הבא כל מקדם שקיבל הקצאת סיביות יוחלף בשגיאת הערך שהוקצה לו, כך ששגיאה זו תהווה את הערך החדש לטיפול – באופן זה לא ננציח שגיאות אלא נחזור ונטפל בהן אם הן גדולות. שוב נקבע סף – קטן יותר – נבחן את המקדמים הדומיננטיים לפי סף זה, ונשדר מידע עליהם, וחוזר חלילה. באופן זה אם נמשיך, נגיע לסדרת סיביות גדולה המביאה לייצוג מלא של התמונה במקדמי ההתמרה, וכל תת-קבוצה של סיביות מתחילת הקובץ ועד לנקודה כלשהי מתארת דחיסה ביחס מסויים – כל שכל עקומת ה- $R(D)$ נבנתה לתמונה הנ"ל.

עד כה דנו בעניין הקוונטיזציה והפרוגרסיביות שבה. הבעיה הברורה שתעלה היא שבכל שלב בתהליך הנ"ל נקבל מספר קטן של מקדמי wavelet הפזורים בתמונה ושעליהן עלינו לתת מידע. אם לא נפעל נכון נזדקק לא רק למידע על כיצד לעדכן מקדמים אלו, אלא גם לספר על מיקומם וזהו עול שיעלה בריבוי סיביות – כאן נכנס מבנה העץ שהוזכר קודם. מסתבר כי ישנה לרוב – לא תמיד – חוקיות נוחה לפיה אם נארגן את מקדמי ההתמרה במבנה עץ quad-tree כמתואר בציור הבא, תימצא חוקיות נוחה לפיה אם בשורש ישנו מקדם חלש, מרבית הסיכויים שגם "ילדיו" בעץ יהיו קטנים. חוקיות זו מאפשרת לקודד את מיקומי המקדמים באופן יעיל.

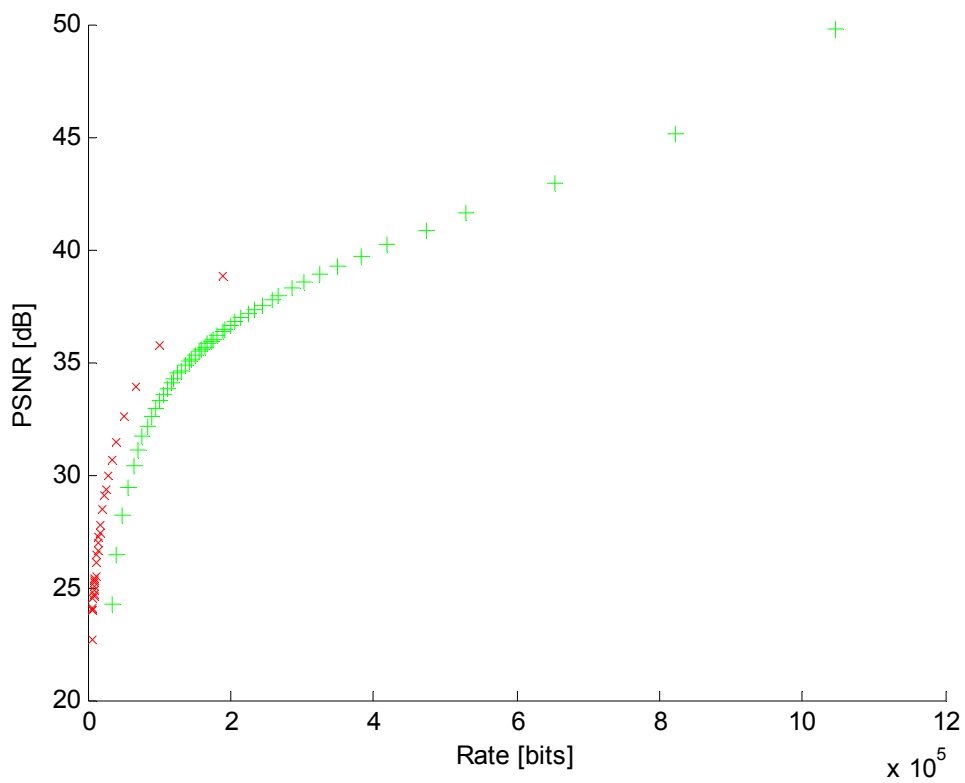
אנו נעצור את התיאור כאן – למעשה JPEG2000 לא עושה שימוש בגישה זו כי נמצאה גישה אחרת פשוטה יותר להשגת אותו יעד, אך כאמור, אלו פרטים טכניים משעממים מבחינה מדעית.



מבנה ה-zero-tree לקידוד יעיל של מיקומי המקדמים הדורשים טיפול

הגרף הבא מראה את ביצועי אלגוריתם ה-JPEG2000 בהשוואה לאלה המושגים ע"י ה-JPEG. לשם בדיקת ה-JPEG2000 השתמשנו בתוכנת ifranview היכולה לקבל ולייצר תמונות בפורמט JP2. לקחנו את התמונה Lena ושמרנו אותה ברמות איכות שונות בפורמט JPEG2000 ובחנו את הביצועים על עקומת ה- $R(D)$ כשקצב נמדד לפי גודל הקובץ והאיכות ע"י המרת התמונה לפורמט TIF, קריאתה ל-Matlab וחישוב ה-PSNR. רואים כי האלגוריתם החדש עולה על ה-JPEG הישן. ישנו שיפור של כ-3dB בין האלגוריתמים, אך חשוב מזה, ישנם גדלי קובץ אליהם JPEG לא יכול להגיע ו-JPEG2000 משיג באיכויות ירודות.

הציור הבא מראה מספר צמדים אופייניים של תמונות בעלי כמות סיביות דומה בין JPEG ל-JPEG2000, וניכר השיפור הדרמטי שבפורמט החדש.



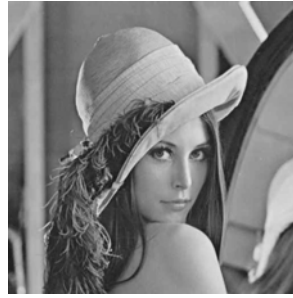
עקומת קצב העיוות המשווה בין JPEG לבין JPEG2000.



תמונת JPEG באיכות 86 ובנפח 48KB ולעומתה תמונת JPEG2000 עם אותו נפח.



תמונת JPEG באיכות 70 ובנפח 29.5KB ולעומתה תמונת JPEG2000 עם אותו נפח.



תמונת JPEG באיכות 40 ובנפח 18.1KB ולעומתה תמונת JPEG2000 עם אותו נפח.



תמונת JPEG באיכות 18 ובנפח 10.7KB ולעומתה תמונת JPEG2000 עם אותו נפח.



תמונת JPEG באיכות 9 ובנפח 6.6KB ולעומתה תמונת JPEG2000 עם אותו נפח.

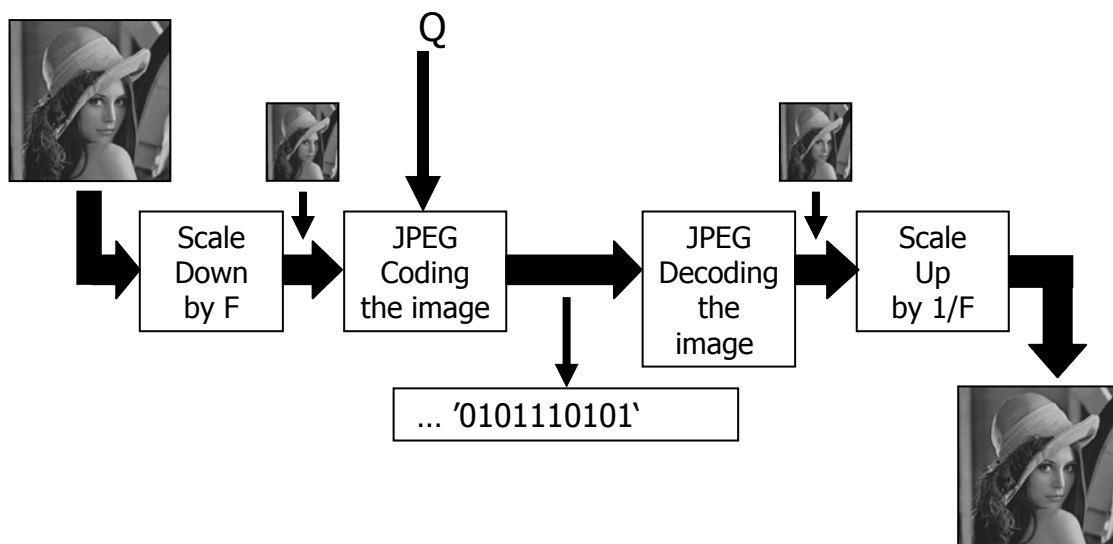


תמונת JPEG באיכות 1 ובנפח 2.8KB ולעומתה תמונת JPEG2000 עם אותו נפח.

השפעת Scale על ביצועי דחיסה ב-JPEG

נחזור לרגע לתוצאה מעניינת שהתקבלה בהשוואה בין JPEG ל-JPEG2000 – עובדת היות ה-JPEG2000 מסוגל להגיע לקבצים קטנים מאוד בעוד JPEG עצמו חסום. האומנם אין מוצא פשוט ממגבלה זו? אנו נראה כי קיים מוצא פשוט ויפה אשר ייתן זווית ראייה מעניינת על יתרונו של ה-JPEG2000.

אנו נציע את השימוש במערכת הבאה, בה התמונה מוקטנת, נדחסת באלגוריתם ה-JPEG, ולאחר פריסתה מוגדלת חזרה ע"י אינטרפולציה כלשהי. הרעיון הבסיסי הוא שבחלוקה של התמונה לבלוקים בגודל 8-על-8 פיקסלים, אפילו אם נרצה להקצות סיבית אחת לבלוק (למעשה המינימום האפשרי גדול מזה), גודל התמונה יכתיב כמות סיביות ככמות הבלוקים ואין למטה מזה מכיוון שב-JPEG כל בלוק מקבל טיפול זהה. בהקטנה אנו מקטינים את כמות הבלוקים ולכן פורצים את המחסום הנ"ל.



סכימת בלוקים של תהליך הדחיסה-פריסה המשופר

הציורים הבאים מראים צמדים של תמונות במתקבלות האחת ב-JPEG רגיל, והאחרת בהקטנה 2:1 לפני הדחיסה והגדלה לאחר הפריסה. אנו רואים תוצאות טובות לאין ערוך בגישה המשופרת. אנו גם רואים כי ניתן לרדת לגודל של קבצים מתחת למינימום האפשרי ב-JPEG. איך מנגנון זה משתווה ל-JPEG2000? האם ניתן להשתמש בו גם בסטנדרט זה לשם השגת שיפור נוסף?



תוצאת JPEG באיכות 5 (0.162 bits per pixel) וגרסתה התואמת בהקטנה 2:1.

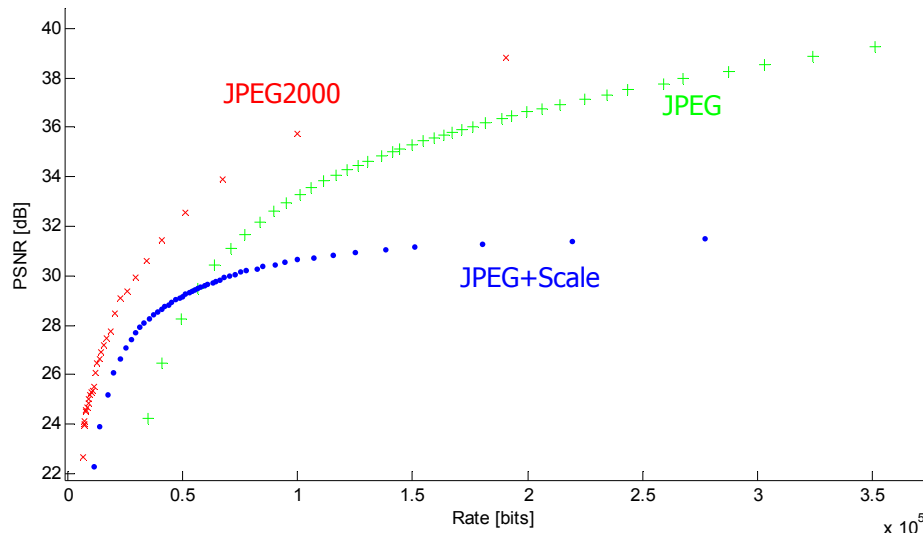


תוצאת JPEG באיכות 1 (0.127 bits per pixel) וגרסתה התואמת בהקטנה 2:1.



תוצאת JPEG עם הקטנה 2:1 - 0.078 bits per pixel – אין תוצאה תואמת ב-JPEG

הנה תוצאות ה-JPEG, ה-JPEG2000 והשיטה בה ישנו scale על גבי צירי קצב-עיוות. אנו רואים כי השיטה המוצעת מצליחה ליצר קבצים קטנים יותר כמתואר, אנו גם רואים כי התוצאות טובות מאלה של ה-JPEG אך אינן מתחרות ב-JPEG2000. ברור גם כי אם מתבקשת איכות מוצא גבוהה, שיטה זו נופלת בנקודה כלשהי מה-JPEG בשל אובדן המידע שבהקטנה.



עקומת קצב העיוות המשווה בין JPEG, JPEG2000 ו-JPEG עם הקטנה.

ישנה תופעה מעניינת המתרחשת כאן ושאליה לא התכוונו. בהגדלת התמונה במוצא הפריסה אנו מטשטשים את גבולות הבלוקים וכתוצאה מכך מוחקים הרבה מתופעת הבלוקיות שכל-כך מאפיינת את ה-JPEG. למעשה קיימות שיטות מתקדמות שנועדו לתת ריכוך של תופעת הבלוקיות ושיטה זו מהווה מתחרה מעניין.

האם גישה זו תעבוד ב-JPEG2000? מסתבר שלא! הסיבה היא שבשיטות מבוססות wavelet המשחק עם גודל התמונה קיים כמרכיב יסודי באלגוריתם הדחיסה ומנוצל עד תום. נאמר זאת אחרת – מדוע שיהיה פקטור הקטנה אחד זהה לכל שטח התמונה? נכון היה לתת גודל בלוק אדפטיבי לכל אזור לפי תוכנו. יתרה מזו – נכון היה לתת מספר גדלי בלוק לכל אזור כך שתכנים שונים המצויים באותו אזור יטופלו כל אחד בצורה מיטבית – את כל זה עושה התמרת ה-wavelet ללא שתכנון זאת במפורש.