

МАТЕМАТИЧКА ГИМНАЗИЈА

Примена нумеричких трансформација у мултимедији

Матурски рад из математике

Аутор:
Василије Хаџи-Пурић

Ментор:
проф. Вукашин Брковић

Београд, 2026.

Садржај

1	Увод у дигиталне мултимедијалне сигнале	1
1.1	Дискретизација сигнала	2
1.2	Узорковање и Најквистова теорема	2
1.3	Квантизација	3
1.4	Фреквенцијске репрезентације	3
1.5	Значај нумеричких трансформација	3
2	Брза Фуријеова трансформација (FFT)	5
2.1	Математичка основа DFT	5
2.2	Рачунска сложеност DFT	7
2.3	Идеја FFT алгоритма	7
2.4	Cooley–Tukey алгоритам	7
2.5	Пример имплементације FFT алгоритма	7
2.6	Примене FFT алгоритма у мултимедији	9
2.6.1	Аудио филтрирање	9
2.6.2	Спектрограми	9
2.6.3	Редукција шума	9
2.7	Конкретан пример FFT анализе сигнала	9
3	Дискретна косинусна трансформација (DCT)	12
3.1	Веза DCT и DFT	12
3.2	Типови DCT	13
3.3	Улога DCT у JPEG компресији	14
3.4	Енергетска компактност и квантизација	15
3.5	Нумеричка стабилност	15
3.6	Кодирање блокова 8×8	15
3.7	Конкретан пример DCT компресије	16
4	Дискретне вејвлет трансформације (DWT)	18
4.1	Вишескална анализа	18
4.2	Наг вејвлет	18
4.3	Daubechies вејвлети	19
4.4	Примена у JPEG-2000	20

4.5	Робустност и локализација	20
4.6	Конкретан пример DWT декомпозиције	21
5	Упоредна анализа трансформација	23
5.1	Рачунска сложеност	23
5.2	Квалитет реконструкције	23
5.3	Ефикасност компресије	24
5.4	Нумеричка стабилност	24
5.5	Општи закључак упоредне анализе	25
6	Примери из реалних мултимедијалних стандарда	26
6.1	JPEG: DCT и квантизација	26
6.2	MP3: MDCT и психоакустика	26
6.3	MPEG видео компресија: компензација покрета и DCT	27
6.4	JPEG-2000 (DWT): вејвлет компресија	27
7	Закључак	28
	Захвалница	29

1 Увод у дигиталне мултимедијалне сигнале

Савремени свет заснива се на огромној количини дигиталних података. Фотографије, музика, видео записи, интернет комуникација и системи вештачке интелигенције функционишу захваљујући дигиталној обради сигнала. Под појмом дигитални сигнал подразумева се низ података који представља неку физичку величину, као што су звук, осветљеност слике или јачина електромагнетног таласа [1].

Дигитална мултимедија обухвата слике, звук, видео и друге облике информација који се преносе, складиште или обрађују помоћу рачунара. Да би рачунар могао да обрађује ове сигнале, неопходно је да они буду представљени у дискретном облику, односно као коначан низ бројева.

На пример, дигитална фотографија састоји се од великог броја пиксела, при чему сваки пиксел има своју нумеричку вредност осветљености или боје. Слично томе, дигитални аудио сигнал представља се као низ вредности које описују промену звучног таласа током времена.

Основна идеја савремене обраде сигнала јесте да се сложени сигнали могу представити помоћу једноставнијих математичких компоненти. Уместо директног посматрања сигнала у временском или просторном домену, често је корисније анализирати његову структуру у фреквенцијском домену.

Фреквенцијски домен описује од којих је „таласних компоненти“ сигнал састављен. На пример, музички тон може се разложити на основну фреквенцију и више хармонијских компоненти. На сличан начин, слике се могу представити комбинацијом спорих и брзих промена осветљености.

Овакав приступ омогућава ефикаснију компресију, уклањање шума, анализу и реконструкцију сигнала. Управо зато су математичке трансформације постале основа савремених мултимедијалних система.

У овом раду анализирају се три важне трансформације:

- FFT (Брза Фуријеова трансформација),
- DCT (Дискретна косинусна трансформација),
- DWT (Дискретна вејвлет трансформација).

FFT омогућава ефикасну анализу фреквенција у сигналу и широко се користи у аудио обради и телекомуникацијама. DCT представља основу JPEG компресије слика, јер омогућава концентрацију енергије у малом броју коефицијената. DWT омогућава вишескалну анализу сигнала и посебно је погодна за обраду слика са локалним структурама и ивицама.

У наставку рада биће објашњени основни математички принципи ових трансформација, њихове особине, алгоритми и примена у савременим мултимедијалним системима.

1.1 Дискретизација сигнала

Сигнали који се јављају у природи најчешће су аналогни, што значи да се њихова вредност мења непрекидно током времена. На пример, звучни талас који човек чује или електрични напон у микрофону могу имати бесконачно много различитих вредности у било ком временском тренутку.

Математички, аналогни сигнал описује се функцијом $x(t)$, где променљива t представља време. Пошто рачунари не могу директно да обрађују континуалне функције, неопходно је да се сигнал претвори у дискретан облик погодан за нумеричку обраду.

Овај поступак назива се дискретизација сигнала. Уместо непрекидног посматрања сигнала, вредности се мере само у одређеним временским тренуцима који су међусобно равномерно удаљени.

Дискретизован сигнал записује се у облику:

$$x[n] = x(nT_s),$$

где је T_s период узорковања, односно временско растојање између два суседна узорка сигнала.

Променљива n представља редни број узорка, док вредност $x[n]$ описује сигнал у одређеном тренутку времена. На овај начин континуални сигнал се претвара у низ бројева који се могу чувати у меморији и обрађивати различитим алгоритмима.

На пример, код дигиталног аудио записа микрофон периодично мери вредност звучног таласа и претвара га у низ дискретних узорака. Што је период узорковања мањи, добија се већи број узорака и вернија репрезентација оригиналног сигнала.

1.2 Узорковање и Најквистова теорема

Након дискретизације сигнала поставља се важно питање: колико често је потребно мерити вредност сигнала да би се он могао верно реконструисати?

Поступак периодичног мерења вредности аналогног сигнала назива се узорковање. Резултат узорковања је низ дискретних вредности које представљају сигнал у одређеним временским тренуцима.

Број узорака који се узима у једној секунди назива се фреквенција узорковања и означава се са f_s . Она се најчешће мери у херцима (Hz).

Ако је фреквенција узорковања сувише мала, део информација о сигналу се губи и долази до појаве изобличења познатог као алијасинг. Тада се различите фреквенције погрешно интерпретирају као ниже фреквенције, што доводи до нетачне реконструкције сигнала.

Основни резултат теорије узорковања представља Nyquist–Shannonова теорема.

Теорема 1.1 (Најквистова теорема). *Ако је сигнал ограниченог опсега фреквенција $f_{\max}$, онда је тачна реконструкција могућа ако је фреквенција узорковања*

$$f_s \geq 2f_{\max}.$$

Вредност $2f_{\max}$ назива се Најквистова фреквенција. Теорема показује да је за верну реконструкцију сигнала неопходно узорковати сигнал најмање два пута брже од његове највеће фреквенцијске компоненте.

На пример, човеково ухо чује звукове приближно до 20 kHz. Због тога се код аудио компакт-дискова користи фреквенција узорковања 44.1 kHz, што је нешто више од двоструке вредности максималне чујне фреквенције.

Најквистова теорема представља један од најважнијих резултата дигиталне обраде сигнала и чини математичку основу савремених аудио, видео и телекомуникационих система [2].

1.3 Квантизација

Након узорковања, вредности сигнала се заокружују на најближи елемент коначног скупа дискретних нивоа. Ово уводи квантизациону грешку, али је неизбежан корак у дигиталном систему.

1.4 Фреквенцијске репрезентације

Многи сигнали садрже структуру која је тешко уочљива у временском домену, али постаје јасна у фреквенцијском домену. На пример, аудио сигнал у временском домену изгледа као талас који се мења из тренутка у тренутак. Међутим, када се исти сигнал пребаци у фреквенцијски домен, види се од којих тонова је састављен.

Слична идеја важи и за слике. Уместо да посматрамо само појединачне пикселе, трансформације омогућавају да уочимо споре промене осветљености, оштре ивице и ситне детаље. Зато се уводе трансформације које пребацују сигнал у фреквенцијски домен што је веома важно у компресији и анализи мултимедије.

1.5 Значај нумеричких трансформација

Нумеричке трансформације су важне зато што омогућавају другачији приступ анализи сигнала. Уместо да се посматра само ток вредности кроз време или распоред пиксела у слици, трансформације откривају унутрашњу структуру сигнала. Та структура често показује које су фреквенције најзначајније, где се јавља шум и који делови сигнала носе највише информација.

Један од најважнијих примера је компресија слике. Код JPEG формата слика се не памти директно пиксел по пиксел у изворном облику, већ се најпре разлаже помоћу DCT трансформације. После тога се задржавају најважнији коефицијенти, док се мање важни детаљи могу занемарити. На тај начин величина фајла постаје много мања, а слика за људско око остаје скоро иста.

Други пример је обрада звука. Када се аудио сигнал анализира помоћу FFT трансформације, могуће је видети које су фреквенције присутне у сигналу. Ако се у снимку јавља шум на одређеној фреквенцији, та компонента се може ослабити или

уклонити. Због тога су фреквенцијске трансформације основа многих програма за обраду звука.

Трећи пример је анализа слика помоћу DWT трансформације. Вејвлет трансформације омогућавају да се слика разложи на грубу апроксимацију и детаље. То је корисно када је важно очувати ивице, прелазе и локалне структуре, као код медицинских или сателитских снимака.

У овом раду посебно се анализирају FFT, DCT и DWT, јер оне представљају основу великог броја савремених мултимедијалних система.

2 Брза Фуријеова трансформација (FFT)

2.1 Математичка основа DFT

Дискретна Фуријеова трансформација (DFT) дефинише се као:

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-i2\pi nk/N}, \quad k = 0, \dots, N-1.$$

DFT представља основну трансформацију дигиталне обраде сигнала. Њена улога је да сигнал из временског домена пребаци у фреквенцијски домен. На тај начин могуће је анализирати које фреквенције доминирају у сигналу и колики је њихов интензитет.

У формули $x[n]$ представља вредности дискретног сигнала у времену, док $X[k]$ представља спектар сигнала у фреквенцијском домену. Индекс n означава редни број узорка сигнала, а индекс k одговара одређеној фреквенцијској компоненти.

Фактор $e^{-i2\pi nk/N}$ представља комплексни експоненцијални талас. DFT у суштини мери колико је свака фреквенција присутна у посматраном сигналу.

На пример, ако аудио сигнал садржи тонове од 440 Hz и 880 Hz, након примене FFT алгоритма у спектру ће се појавити изражени пикови управо на тим фреквенцијама.

Инверзна трансформација гласи:

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X[k] e^{i2\pi nk/N}.$$

DFT пребацује сигнал у дискретни фреквенцијски спектар.

Теорема 2.1 (Инверзност DFT трансформације). *Нека је дискретна Фуријеова трансформација сигнала $x[n]$ дата формулом*

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-i2\pi nk/N}.$$

Тада се оригинални сигнал може реконструисати формулом

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X[k] e^{i2\pi nk/N}.$$

Доказ. Убацимо израз за $X[k]$ у инверзну формулу:

$$\frac{1}{N} \sum_{k=0}^{N-1} X[k] e^{i2\pi nk/N} = \frac{1}{N} \sum_{k=0}^{N-1} \left(\sum_{m=0}^{N-1} x[m] e^{-i2\pi mk/N} \right) e^{i2\pi nk/N}.$$

Променом редоследа сумирања добијамо

$$\frac{1}{N} \sum_{m=0}^{N-1} x[m] \sum_{k=0}^{N-1} e^{i2\pi(n-m)k/N}.$$

Унутрашња сума једнака је N ако је $n = m$, а једнака је 0 ако је $n \neq m$. Зато у спољашњој суми остаје само члан за $m = n$, па добијамо

$$x[n].$$

Дакле, инверзна DFT формула заиста враћа почетни сигнал. $\square$

Теорема 2.2 (Парсевалова теорема за DFT). *Нека је дискретан сигнал $x = (x_0, x_1, \dots, x_{N-1})$, а његова дискретна Фуријеова трансформација дата формулом*

$$X_k = \sum_{n=0}^{N-1} x_n e^{-2\pi i k n / N}, \quad k = 0, 1, \dots, N-1.$$

Тада важи

$$\sum_{n=0}^{N-1} |x_n|^2 = \frac{1}{N} \sum_{k=0}^{N-1} |X_k|^2.$$

Доказ. Инверзна дискретна Фуријеова трансформација има облик

$$x_n = \frac{1}{N} \sum_{k=0}^{N-1} X_k e^{2\pi i k n / N}.$$

Посматрајмо енергију сигнала у временском домену:

$$\sum_{n=0}^{N-1} |x_n|^2 = \sum_{n=0}^{N-1} x_n \overline{x_n}.$$

Убацавањем инверзне DFT формуле добијамо

$$\sum_{n=0}^{N-1} |x_n|^2 = \frac{1}{N^2} \sum_{n=0}^{N-1} \left(\sum_{k=0}^{N-1} X_k e^{2\pi i k n / N} \right) \left(\sum_{m=0}^{N-1} \overline{X_m} e^{-2\pi i m n / N} \right).$$

Променом редоследа сумирања следи

$$\sum_{n=0}^{N-1} |x_n|^2 = \frac{1}{N^2} \sum_{k=0}^{N-1} \sum_{m=0}^{N-1} X_k \overline{X_m} \sum_{n=0}^{N-1} e^{2\pi i (k-m)n / N}.$$

Користимо ортогоналност комплексних експоненцијалних функција:

$$\sum_{n=0}^{N-1} e^{2\pi i (k-m)n / N} = \begin{cases} N, & k = m, \\ 0, & k \neq m. \end{cases}$$

Зато у двострукој суми остају само чланови за које је $k = m$, па добијамо

$$\sum_{n=0}^{N-1} |x_n|^2 = \frac{1}{N} \sum_{k=0}^{N-1} X_k \overline{X_k} = \frac{1}{N} \sum_{k=0}^{N-1} |X_k|^2.$$

Тиме је теорема доказана. $\square$

2.2 Рачунска сложеност DFT

Директно рачунање DFT захтева:

$$O(N^2) \text{ множења.}$$

Квадратна сложеност $O(N^2)$ значи да број операција веома брзо расте са повећањем броја узорака сигнала. Код савремених мултимедијалних система, где сигнали могу садржати милионе узорака, директно израчунавање DFT постаје сувише споро. За $N = 1,000,000$ овај поступак је практично непримењив.

Из тог разлога развијен је FFT алгоритам, који користи симетрије и периодичност комплексних експоненцијалних функција како би значајно смањио број потребних операција.

2.3 Идеја FFT алгоритма

FFT смањује сложеност на:

$$O(N \log N).$$

Основна идеја је подела на *парне* и *непарне* индексе:

$$X[k] = E[k] + \omega_N^k O[k],$$

$$X[k + N/2] = E[k] - \omega_N^k O[k].$$

где је

$$\omega_N = e^{-2\pi i/N}.$$

У претходним формулама $E[k]$ представља трансформацију парних индекса сигнала, док $O[k]$ представља трансформацију непарних индекса. FFT алгоритам раздваја сигнал на два мања проблема, а затим рекурзивно наставља поступак док не остану трансформације дужине 1.

Оваква подела значајно убрзава рачунање. Уместо квадратне сложености добија се сложеност $O(N \log N)$, што омогућава обраду сигнала у реалном времену.

FFT је данас један од најважнијих алгоритама дигиталне обраде сигнала и користи се у аудио анализи, телекомуникацијама, медицини, астрономији и компресији мултимедије.

2.4 Cooley–Tukey алгоритам

Ако је N степен двојке, трансформација се може применити рекурзивно док не остану трансформације дужине 1.

2.5 Пример имплементације FFT алгоритма

Python

```
def fft(a):
    n = len(a)
    if n <= 1:
        return a
    even = fft(a[0::2])
    odd = fft(a[1::2])
    T = [exp(-2j*pi*k/n) * odd[k] for k in range(n//2)]
    return [even[k] + T[k] for k in range(n//2)] + \
           [even[k] - T[k] for k in range(n//2)]
```

Приказани програм реализује рекурзивну FFT имплементацију. Листа сигнала дели се на елементе са парним и непарним индексима, након чега се FFT примењује засебно на оба дела.

Променљива T представља такозване „twiddle” факторе, односно комплексне експоненцијалне множиоце који се јављају приликом спајања резултата парног и непарног дела сигнала.

Математички, ови фактори имају облик

$$e^{-2\pi i k/n},$$

и представљају комплексне ротације у комплексној равни. Њихова улога је да правилно помере фазе фреквенцијских компоненти тако да се резултати мањих FFT трансформација могу комбиновати у јединствен спектар сигнала.

У сваком кораку алгорита сигнал се раздваја на два дела:

- елементе са парним индексима,
- елементе са непарним индексима.

FFT се затим рекурзивно примењује на оба дела, а twiddle фактори омогућавају њихово правилно спајање у коначни резултат. Без ових множилаца не би било могуће очувати тачне фазне односе између фреквенцијских компоненти сигнала.

Геометријски посматрано, комплексни експоненцијални фактори представљају тачке на јединичној кружници у комплексној равни. Различите вредности параметра k одговарају различитим угловима ротације, што омогућава издвајање појединачних фреквенција сигнала.

Иако је приказана имплементација једноставна и погодна за разумевање основне идеје FFT алгорита, она није оптимална за практичну употребу. Рекурзивни позиви и рад са Python листама успоравају извршавање код великих сигнала.

Због тога се у реалним системима користе високо оптимизоване библиотеке као што су FFTW и NumPy FFT. Ове библиотеке користе напредне технике оптимизације, кеширања и векторизације, као и специјализоване алгоритме прилагођене архитектури процесора.

На тај начин могуће је веома брзо израчунавање FFT трансформације чак и над сигнаlima који садрже милионе узорака, што је од суштинског значаја у мултимедијалној обради, телекомуникацијама и системима за рад у реалном времену.

2.6 Примене FFT алгоритма у мултимедији

2.6.1 Аудио филтрирање

FFT омогућава смањење шума тако што се нежељене фреквенцијске компоненте идентификују и ослабе.

Када се сигнал пребаци у фреквенцијски домен, могуће је уклонити нежељене компоненте сигнала. На пример, шум електричне мреже често се јавља на 50 Hz. FFT омогућава да се та компонента идентификује и ослаби, након чега се сигнал враћа у временски домен применом инверзне трансформације.

2.6.2 Спектрограми

Спектрограм је визуелни приказ промене фреквенцијског садржаја сигнала током времена. Добија се тако што се сигнал подели на кратке временске интервале, а затим се на сваки од тих делова примени FFT.

На овај начин се не добија само информација о томе које су фреквенције присутне у сигналу, већ и када се оне појављују. Због тога је спектрограм веома користан у анализи говора, музике и звучних сигнала чији се садржај мења током времена.

На пример, код људског говора различити гласови имају различиту фреквенцијску структуру. Спектрограм омогућава да се те разлике визуелно прикажу, што је важно у препознавању говора и анализи акустичких сигнала.

2.6.3 Редукција шума

Редукција шума заснована на FFT алгоритму користи чињеницу да се шум и користан сигнал често разликују по фреквенцијском саставу. Када се сигнал пребаци у фреквенцијски домен, могуће је ослабити компоненте које имају малу амплитуду или које припадају познатом опсегу шума.

Међутим, није довољно једноставно уклонити све мале компоненте, јер и неки корисни делови сигнала могу имати малу амплитуду. Зато се у практичним системима редукција шума врши пажљиво, тако да се смањи шум, али да се притом очува природан изглед или звук оригиналног сигнала.

2.7 Конкретан пример FFT анализе сигнала

Посматрајмо сигнал:

$$x(t) = \sin(2\pi 50t) + 0.5 \sin(2\pi 120t).$$

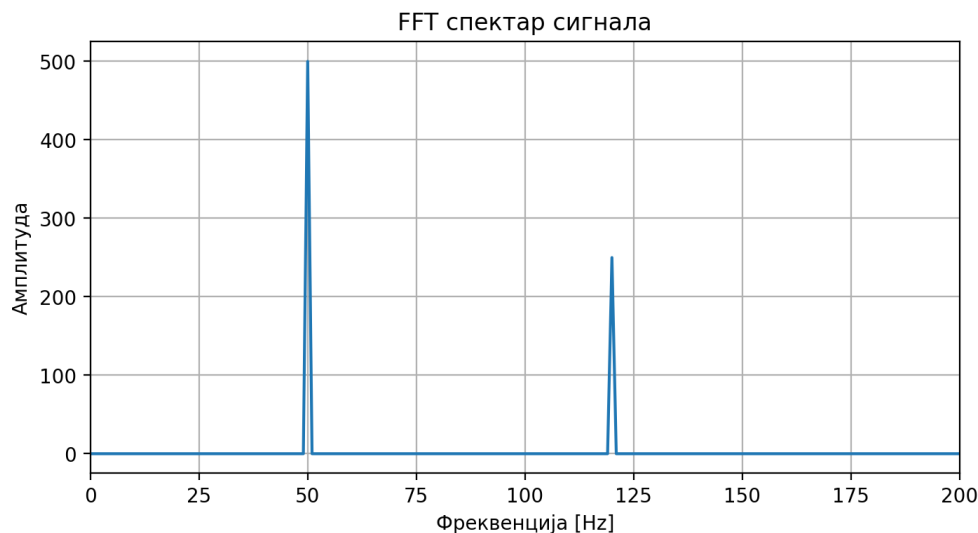
Овај сигнал представља комбинацију две синусоидалне компоненте различитих фреквенција. Прва компонента има фреквенцију 50 Hz и амплитуду 1, док друга компонента има фреквенцију 120 Hz и мању амплитуду 0.5.

У временском домену сигнал изгледа као сложена осцилација и тешко је директно уочити његову структуру. Међутим, након примене FFT трансформације добија се фреквенцијски спектар у коме се јасно појављују два доминантна пика.

Први пик налази се на 50 Hz и има већу амплитуду, док се други пик појављује на 120 Hz и има приближно двоструко мању висину. На овај начин FFT омогућава издвајање фреквенција које чине сигнал.



Слика 2.1. Сигнал $x(t) = \sin(2\pi 50t) + 0.5 \sin(2\pi 120t)$ у временском домену.



Слика 2.2. FFT спектар сигнала са израженим пиковима на 50 Hz и 120 Hz.

Слика 2.1 приказује сигнал у временском домену, где се суперпозиција две синусоидалне компоненте манифестује као сложена осцилација.

Након примене FFT алгоритма добија се спектар приказан на слици 2.2. У спектру се јасно уочавају два доминантна пика који одговарају фреквенцијама од 50 Hz и 120 Hz. Виши пик одговара компоненти веће амплитуде, док је други пик нижи јер друга синусоида има мању амплитуду.

Овај пример показује како FFT омогућава издвајање фреквенцијских компоненти сигнала које нису једноставно уочљиве директним посматрањем сигнала у временском домену.

Овај приступ има ширу примену. На пример, користи се у аудио анализи, препознавању музичких тонова, обради говора и отклањању шума из сигнала.

3 Дискретна косинусна трансформација (DCT)

3.1 Веза DCT и DFT

DCT се може посматрати као DFT реалног и парног проширења сигнала. Основна формула DCT (конкретно DCT-II) је:

$$X[k] = \alpha_k \sum_{n=0}^{N-1} x[n] \cos\left(\frac{\pi(2n+1)k}{2N}\right),$$

где је

$$\alpha_0 = \sqrt{\frac{1}{N}}, \quad \alpha_k = \sqrt{\frac{2}{N}} \text{ за } k > 0.$$

DCT трансформација има сличан циљ као DFT, тј. прелазак у фреквенцијски домен, али користи само косинусне функције. Због тога су резултати реални бројеви, што поједностављује рачунање и смањује меморијске захтеве.

У формули $x[n]$ представља вредности сигнала, док $X[k]$ представља DCT коефицијенте. Мањи индекси одговарају спорим променама сигнала, док већи индекси описују брзе промене и фине детаље.

Код слика већина енергије концентрише се у малом броју нискофреквентних коефицијената. Управо ова особина омогућава ефикасну компресију JPEG формата.

Теорема 3.1 (Ортогоналност косинусних базних функција). *За DCT-II базне функције важи да различите косинусне компоненте носе независне информације о сигналу. Другим речима, за $k \neq l$ важи*

$$\sum_{n=0}^{N-1} \cos\left(\frac{\pi(2n+1)k}{2N}\right) \cos\left(\frac{\pi(2n+1)l}{2N}\right) = 0.$$

Доказ. Користимо тригонометријски идентитет

$$\cos A \cos B = \frac{1}{2} (\cos(A - B) + \cos(A + B)).$$

Нека је

$$A = \frac{\pi(2n+1)k}{2N}, \quad B = \frac{\pi(2n+1)l}{2N}.$$

Тада је посматрана сума једнака

$$\frac{1}{2} \sum_{n=0}^{N-1} \cos\left(\frac{\pi(2n+1)(k-l)}{2N}\right) + \frac{1}{2} \sum_{n=0}^{N-1} \cos\left(\frac{\pi(2n+1)(k+l)}{2N}\right).$$

За $k \neq l$ ове суме представљају збир равномерно распоређених косинусних вредности током целог периода. Због симетрије позитивни и негативни чланови се поништавају, па је укупан збир једнак нули.

Зато су различите DCT компоненте ортогоналне, што значи да свака од њих описује посебан део структуре сигнала. $\square$

3.2 Типови DCT

Дискретна косинусна трансформација постоји у више варијанти. Оне се разликују по начину на који се сигнал симетрично продужава на крајевима интервала, као и по положају тачака у којима се рачунају косинусне базне функције. Најчешће се издвајају четири основна типа: DCT-I, DCT-II, DCT-III и DCT-IV.

У наставку су дате формуле без нормализационих константи. У практичним библиотекама, као што су NumPy и SciPy, могу се користити додатни фактори нормализације да би трансформације биле ортонормалне.

За сигнал $x[0], x[1], \dots, x[N-1]$, DCT-I се дефинише формулом

$$X[k] = x[0] + (-1)^k x[N-1] + 2 \sum_{n=1}^{N-2} x[n] \cos\left(\frac{\pi n k}{N-1}\right), \quad k = 0, 1, \dots, N-1.$$

DCT-I користи симетрију на оба краја сигнала. Због тога се често појављује у теоријским разматрањима ортогоналних система и код проблема у којима су важни гранични услови. У мултимедијалној компресији се ређе користи од DCT-II.

DCT-II има облик

$$X[k] = \sum_{n=0}^{N-1} x[n] \cos\left(\frac{\pi(2n+1)k}{2N}\right), \quad k = 0, 1, \dots, N-1.$$

Ово је најважнији тип дискретне косинусне трансформације у обради слика. DCT-II веома добро концентрише енергију природних слика у мали број нискофреквентних коефицијената. Управо зато JPEG стандард користи DCT-II при компресији блокова слике величине 8×8 .

DCT-III је тесно повезана са DCT-II и у одговарајућој нормализацији представља њену инверзну трансформацију. Њена формула је

$$X[k] = \frac{1}{2} x[0] + \sum_{n=1}^{N-1} x[n] \cos\left(\frac{\pi n(2k+1)}{2N}\right), \quad k = 0, 1, \dots, N-1.$$

Због ове везе, DCT-III се природно јавља у поступку реконструкције сигнала након примене DCT-II. Ако се DCT-II користи за разлагање сигнала на косинусне компоненте, DCT-III се може посматрати као поступак којим се из тих компоненти поново добија сигнал.

DCT-IV се дефинише формулом

$$X[k] = \sum_{n=0}^{N-1} x[n] \cos\left(\frac{\pi(2n+1)(2k+1)}{4N}\right), \quad k = 0, 1, \dots, N-1.$$

Код DCT-IV и улазни и излазни индекси су померени за пола корака. Ова симетрија је важна у обради аудио сигнала, посебно код модификоване дискретне косинусне трансформације, MDCT, која се користи у MP3 и сродним форматима компресије звука.

Дакле, сви типови DCT користе косинусне функције, али се разликују у начину на који третирају крајеве сигнала и положај узорака. У овом раду најважнија је DCT-II, јер представља основу JPEG компресије слика, док је DCT-III важна као њена инверзна варијанта.

3.3 Улога DCT у JPEG компресији

JPEG дели слику на блокове величине 8×8 пиксела и на сваки блок примењује DCT. На тај начин се вредности пиксела, које описују осветљеност у простору, превode у коефицијенте који описују споре и брзе промене осветљености унутар блока.



Слика 3.1. Пример дводимензионе декомпозиције дискретне косинусне трансформације (DCT).

Главна предност DCT трансформације у JPEG компресији јесте то што се већина енергије слике концентрише у малом броју нискофреквентних коефицијената. Високофреквентни коефицијенти, који описују ситне детаље и нагле промене, често имају мале вредности. Због тога се они могу јаче квантизовати или занемарити, а да слика за посматрача остане визуелно прихватљива.

Након примене DCT трансформације већина високофреквентних коефицијената постаје веома мала. Људско око је мање осетљиво на губитак високофреквентних детаља, па се ти коефицијенти могу грубо квантизовати или потпуно уклонити. На

овај начин JPEG постиже велико смањење величине слике уз релативно мали губитак визуелног квалитета.

3.4 Енергетска компактност и квантизација

Коефицијент $X[0]$ представља просечну боју блока (DC компонента), док већи индекси представљају финије детаље (AC компоненте).

DC компонента $X[0]$ садржи просечну осветљеност блока слике. AC компоненте описују промене осветљености и детаље унутар блока.

Што је индекс коефицијента већи, то он представља финије и брже промене слике. Код природних слика већина високофреквентних коефицијената има веома мале вредности, што омогућава њихово уклањање без значајног губитка квалитета.

3.5 Нумеричка стабилност

Нумеричка стабилност представља способност алгоритма да задржи тачност резултата и у присуству заокруживања, квантизације и других нумеричких грешака које се јављају током рачунања.

Код дигиталне обраде сигнала ова особина је веома важна, јер се трансформације често примењују над великим количинама података и реализују на рачунарима ограничене прецизности. Чак и мале грешке могу се акумулирати и довести до видљивог нарушавања квалитета слике или звука.

DCT трансформација показује добру нумеричку стабилност јер користи реалне косинусне функције. За разлику од DFT трансформације, код које се користе комплексни експоненцијални чланови, DCT избегава рад са комплексним бројевима и на тај начин смањује број нумеричких операција и могућност акумулације грешака.

Посебно је важно што DCT концентрише већину енергије сигнала у мали број нискофреквентних коефицијената. Због тога грешке настале квантизацијом високофреквентних компоненти имају релативно мали утицај на визуелни квалитет реконструисане слике.

У JPEG компресији ова особина омогућава да се велики број високофреквентних коефицијената занемари без појаве значајних визуелних изобличења. Иако се при високим степенима компресије могу јавити блок артефакти, DCT у пракси обезбеђује добар однос између стабилности, квалитета реконструкције и ефикасности компресије.

Захваљујући овим особинама, DCT је постала једна од најзначајнијих трансформација у савременој мултимедијалној обради сигнала.

3.6 Кодирање блокова 8×8

Након примене DCT трансформације над блоком слике величине 8×8 , добија се матрица коефицијената која садржи информације о различитим просторним фреквенцијама унутар блока. Да би се постигла ефикасна компресија, над овим коефицијентима примењује се више узастопних корака кодирања.

Први корак је квантизација. Током овог поступка DCT коефицијенти деле се вредностима из квантизационе матрице и затим се заокружују. Високофреквентне компоненте обично добијају грубљу квантизацију, јер је људско око мање осетљиво на ситне детаље слике. На овај начин велики број коефицијената постаје једнак нули.

Након квантизације примењује се зиг-заг скенирање. Коефицијенти се не читају ред по ред, већ дијагонално, од ниских ка високим фреквенцијама. Овакав распоред омогућава да се велики број нула, које се најчешће налазе међу високофреквентним компонентама, групише једна поред друге.

Следећи корак је кодирање дужина серија. Уместо да се свака нула записује појединачно, памти се број узастопних понављања. Ово значајно смањује количину података које је потребно сачувати.

На крају се примењује Huffman кодирање. Чешће вредности добијају краће бинарне кодове, док ређе вредности добијају дуже кодове. На тај начин се додатно смањује величина компресованог фајла.

Комбинација ових корака омогућава JPEG формату да постигне веома висок степен компресије уз релативно мали губитак визуелног квалитета.

3.7 Конкретан пример DCT компресије

Посматрајмо један блок слике величине 8×8 . Такав блок можемо записати као матрицу

$$B = \begin{bmatrix} B[0,0] & B[0,1] & \cdots & B[0,7] \\ B[1,0] & B[1,1] & \cdots & B[1,7] \\ \vdots & \vdots & \ddots & \vdots \\ B[7,0] & B[7,1] & \cdots & B[7,7] \end{bmatrix} \in \mathbb{R}^{8 \times 8}.$$

Сваки елемент $B[x, y]$ представља осветљеност једног пиксела. Пре примене DCT трансформације вредности пиксела се најчешће центрирају око нуле:

$$b[x, y] = B[x, y] - 128.$$

Ово се ради зато што су вредности пиксела обично у интервалу од 0 до 255, док је за трансформацију погодније да буду распоређене око нуле.

На центрирани блок примењује се дводимензиона DCT-II трансформација:

$$C[u, v] = \alpha_u \alpha_v \sum_{x=0}^7 \sum_{y=0}^7 b[x, y] \cos\left(\frac{\pi(2x+1)u}{16}\right) \cos\left(\frac{\pi(2y+1)v}{16}\right),$$

где су $u, v = 0, 1, \dots, 7$, а коефицијенти нормализације су

$$\alpha_0 = \sqrt{\frac{1}{8}}, \quad \alpha_k = \sqrt{\frac{2}{8}}, \quad k > 0.$$

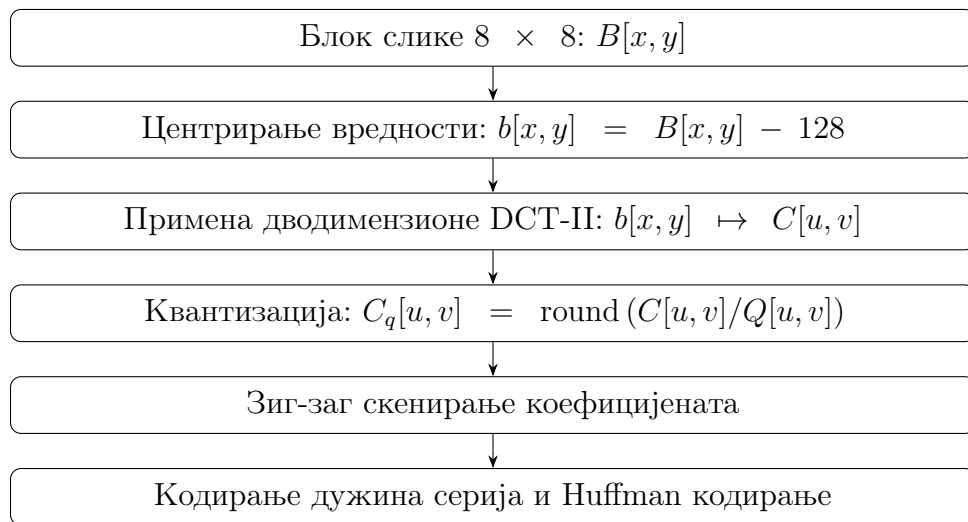
Матрица $C[u, v]$ садржи DCT коефицијенте. Коефицијент $C[0, 0]$ назива се DC компонента и представља просечну осветљеност блока. Остали коефицијенти су AC

компоненте и описују промене осветљености у различитим правцима и на различитим просторним фреквенцијама.

Следећи корак је квантизација:

$$C_q[u, v] = \text{round} \left(\frac{C[u, v]}{Q[u, v]} \right),$$

где је $Q[u, v]$ квантизациона матрица. Веће вредности у матрици Q доводе до јачег заокруживања. Зато се високофреквентни коефицијенти обично јаче квантизују, јер људско око слабије уочава губитак ситних детаља него промену основне структуре слике.



Слика 3.2. Основни ток DCT компресије једног блока слике величине 8×8 .

Зиг-заг скенирање распоређује DCT коефицијенте од нижих ка вишим фреквенцијама. Пошто након квантизације велики број високофреквентних коефицијената постаје нула, овим распоредом се нуле групишу једна поред друге. То омогућава ефикасно кодирање дужина серија.

На пример, ако се након зиг-заг скенирања добије низ

$$(120, -8, 3, 0, 0, 0, 0, 0, \dots),$$

онда није потребно посебно памтити сваку нулу. Довољно је записати колико се нула узастопно појављује.

Математички гледано, компресија је успешна зато што се већина енергије блока налази у малом броју коефицијената:

$$\sum_{(u,v) \in S} |C[u, v]|^2 \approx \sum_{u=0}^7 \sum_{v=0}^7 |C[u, v]|^2,$$

где је S скуп најважнијих нискофреквентних коефицијената.

Ово значи да се оригинални блок може приближно реконструисати и без чувања свих 64 коефицијента. Уместо директног памћења свих вредности пиксела, DCT компресија чува најважније коефицијенте који описују основну структуру блока, док се мање значајни детаљи могу занемарити.

4 Дискретне вејлет трансформације (DWT)

Вејлет трансформације представљају савремени приступ који нуди локализацију и у времену и у фреквенцији, за разлику од FFT и DCT [3].

4.1 Вишескална анализа

Једна од најважнијих особина DWT трансформације јесте могућност вишескалне анализе сигнала. За разлику од класичних трансформација, које сигнал посматрају глобално, вејлет трансформације омогућавају анализу сигнала на различитим размерама.

Основна идеја је да се сигнал раздвоји на:

- апроксимацијске компоненте,
- детаљне компоненте.

Апроксимацијске компоненте описују споре промене сигнала и садрже главну структуру и већину енергије. Детаљне компоненте описују брзе промене, ивице, осцилације и локалне структуре сигнала.

Математички, вејлет трансформација заснива се на скалирајућој функцији $\phi(t)$ и вејлет функцији $\psi(t)$.

Скалирајућа функција описује апроксимацију сигнала:

$$\phi_{j,k}(t) = 2^{j/2} \phi(2^j t - k),$$

док вејлет функција описује детаље:

$$\psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k).$$

У претходним формулама параметар j одређује размеру, односно ниво увећања сигнала, док параметар k одређује положај у времену или простору.

На малим размерама вејлет функције описују фине детаље сигнала, док на већим размерама описују грубље структуре и глобалне промене.

Оваква анализа је посебно корисна код слика и аудио сигнала који садрже локалне промене, јер омогућава истовремену локализацију и у времену и у фреквенцији. Управо због тога DWT има значајну примену у компресији слика, медицинској визуелизацији и анализи природних сигнала.

4.2 Хаар вејлет

Најједноставнији вејлет:

$$\psi(t) = \begin{cases} 1, & 0 \leq t < \frac{1}{2}, \\ -1, & \frac{1}{2} \leq t < 1. \end{cases}$$

Нааг вејвлет представља најједноставнији пример вејвлет функције. Он омогућава раздвајање сигнала на просечне вредности и детаље.

За разлику од FFT и DCT трансформација, које анализирају сигнал глобално, DWT омогућава локализацију и у времену и у фреквенцији. То значи да је могуће одредити не само које су фреквенције присутне, већ и у ком делу сигнала се појављују.

Ова особина је веома важна код слика које садрже ивице, нагле промене и локалне структуре.

4.3 Daubechies вејвлети

Иако је Нааг вејвлет једноставан и погодан за разумевање основних идеја DWT трансформације, у практичним системима често се користе сложенији вејвлети који обезбеђују бољу локализацију и квалитетнију репрезентацију сигнала.

Најпознатију класу таквих функција увела је математичарка Ingrid Daubechies. Daubechies вејвлети представљају фамилију ортогоналних вејвлета који имају компактну подршку и добру глаткоћу.

Компактна подршка значи да је вејвлет различит од нуле само на ограниченом интервалу. Због тога је могуће локално анализирати сигнал без утицаја удаљених делова сигнала.

Математички, Daubechies вејвлети дефинишу се помоћу скалирајуће функције $\phi(t)$ која задовољава релацију:

$$\phi(t) = \sum_{k=0}^{N-1} h_k \sqrt{2} \phi(2t - k),$$

где коефицијенти h_k представљају филтерске коефицијенте вејвлета.

Одговарајућа вејвлет функција $\psi(t)$ дефинише се формулом

$$\psi(t) = \sum_{k=0}^{N-1} g_k \sqrt{2} \phi(2t - k),$$

при чему су коефицијенти g_k повезани са коефицијентима h_k преко ортогоналних услова.

За разлику од Нааг вејвлета, који има нагле прекиде, Daubechies вејвлети су глаткији и боље описују природне сигнале и слике. Због тога производе мање артефаката приликом компресије.

У компресији слика посебно је важно што Daubechies вејвлети добро концентришу енергију сигнала у мали број коефицијената. То омогућава да се велики број мање важних коефицијената занемари, док визуелни квалитет слике остаје висок.

Захваљујући овим особинама, Daubechies вејвлети имају значајну примену у JPEG-2000 стандарду, медицинској визуелизацији, обради сателитских снимака и анализи биомедицинских сигнала [4].

4.4 Примена у JPEG-2000

JPEG-2000 користи DWT уместо DCT, јер вејвлет приступ омогућава боље скалирање резолуције, већу отпорност на шум и постепену реконструкцију слике. Уместо независне обраде малих блокова, слика се разлаже на апроксимације и детаље различитих размера [5].

Овај приступ је важан зато што омогућава да се најпре прикаже груба верзија слике, а затим да се постепено додају детаљи. Таква особина је корисна у системима где је потребно брзо приказивање великих слика или пренос података ограниченом брзином.

За разлику од класичног JPEG формата, код JPEG-2000 не јављају се изражени блок артефакти, јер DWT не користи независне блокове 8×8 . Због тога је овај стандард погодан за медицинске, сателитске и архивске слике, где су очување детаља и квалитет реконструкције посебно важни.

4.5 Робустност и локализација

Једна од најважнијих предности DWT трансформације јесте њена способност локализације сигнала и у простору и у фреквенцији. За разлику од FFT трансформације, која анализира цео сигнал као јединствену целину, вејвлет трансформације омогућавају да се одреди не само које су фреквенције присутне, већ и где се оне појављују.

Ова особина је веома значајна код природних слика, јер се структуре попут ивица, текстура и наглих промена осветљености јављају локално, само у одређеним деловима слике.

Вејвлети су посебно ефикасни код наглих прелаза осветљености, ивица објеката и структура различитих размера. Управо те појаве су честе у природним сликама, због чега је DWT погодан за анализу сложених визуелних садржаја.

Нагли прелази представљају места на којима се вредности пиксела брзо мењају. На пример, прелаз између тамног и светлог дела слике садржи велики број високофреквентних компоненти. DWT омогућава да се такве промене јасно издвоје и локализују.

Ивице су један од најважнијих елемената дигиталних слика јер описују облике и границе објеката. Код класичних трансформација, као што је DCT, ивице често доводе до појаве блок артефаката. Вејвлет трансформације значајно смањују овај проблем јер анализирају сигнал на више размера истовремено.

Промена скале односи се на појаву структура различитих величина. На пример, на сателитском снимку могу се истовремено појавити велики облаци, путеви и ситни објекти. DWT омогућава анализу таквих структура на различитим нивоима резолуције.

Захваљујући овим особинама, вејвлет трансформације су веома робусне при компресији и анализи сложених слика. Због тога имају широку примену у медицинској

визуелизацији, анализи видео сигнала, препознавању објеката и савременим системима компресије слика.

4.6 Конкретан пример DWT декомпозиције

Код дводимензионе DWT трансформације слика се разлаже на четири подопсега: LL , LH , HL и HH . Ови подопсези представљају различите делове информације садржане у слици.

Компонента LL представља апроксимацију слике. Она садржи нискофреквентне информације у оба правца и због тога подсећа на умањену и благо замућену верзију оригиналне слике. У њој се налази највећи део енергије сигнала.

Компоненте LH , HL и HH представљају детаље. Компонента LH наглашава хоризонталне детаље, компонента HL вертикалне детаље, док компонента HH садржи дијагоналне детаље и fine текстуре.



Слика 4.1. Пример једнонивоовске дводимензионе DWT декомпозиције слике на подопсезе LL , LH , HL и HH .

Слика 4.1 приказује како се оригинална слика разлаже на четири дела. Подопсег LL садржи главну структуру слике, док остали подопсези издвајају локалне промене осветљености и ивице.

Математички, први ниво DWT декомпозиције може се записати као примена нископропусног и високопропусног филтера у два правца. Ако са L означимо нископропусни филтер, а са H високопропусни филтер, добијају се подопсези

$$LL = L_x L_y(I), \quad LH = L_x H_y(I),$$

$$HL = H_x L_y(I), \quad HH = H_x H_y(I),$$

где је I оригинална слика, а индекси x и y означавају примену филтера по хоризонталном и вертикалном правцу.

Ова декомпозиција је важна јер омогућава да се највећи део информације сачува у LL компоненти, док се детаљи могу обрађивати одвојено. Управо због тога је DWT погодна за компресију слика, уклањање шума и анализу ивица у медицинским и сателитским снимцима.

Ако се жели даља анализа, DWT се може поново применити само на LL компоненту. Тако се добија вишескална декомпозиција, у којој сваки следећи ниво описује све грубљу структуру слике.

5 Упоредна анализа трансформација

У овом поглављу упоређују се три кључне нумеричке трансформације које се користе у мултимедији: FFT, DCT и DWT. Поређење се заснива на рачунској сложености, квалитету реконструкције, ефикасности компресије и нумеричкој стабилности.

5.1 Рачунска сложеност

Рачунска сложеност је важна јер одређује колико брзо алгоритам може да обради сигнал. У мултимедијалним системима, посебно код звука, видеа и преноса података у реалном времену, није довољно да трансформација буде математички исправна. Она мора бити и довољно брза за практичну примену.

Трансформација	Сложеност
DFT	$O(N^2)$
FFT	$O(N \log N)$
DCT-II	$O(N \log N)$ у ефикасним имплементацијама
DWT (Haar)	$O(N)$
DWT (Daubechies)	$O(N)$

Табела 5.1. Упоредна рачунска сложеност трансформација.

Из табеле се види да је директна DFT трансформација најскупља за рачунање. FFT значајно смањује број операција и зато је погодна за обраду великих сигнала. DWT има линеарну сложеност, што је чини веома ефикасном у задацима вишескалне анализе.

5.2 Квалитет реконструкције

Квалитет реконструкције представља меру колико успешно трансформација може да обнови оригинални сигнал након компресије или уклањања дела коефицијената. У практичним системима компресије није могуће сачувати све информације, па је важно да се и након одбацавања мање значајних компоненти очува што већи квалитет слике или звука.

Код FFT трансформације сигнал се представља глобалним фреквенцијским компонентама. Овај приступ веома добро описује периодичне структуре и спектралне особине сигнала, али има ограничења код локалних промена. Ако се задржи само мали број FFT коефицијената, реконструисана слика често губи оштре ивице и локалне детаље.

DCT показује веома добар квалитет реконструкције код природних слика. Захваљујући способности концентрације енергије у малом броју нискофреквентних коефицијената, могуће је одбацили велики број високофреквентних компоненти уз

релативно мали визуелни губитак. Због тога DCT даје добре резултате код глатких слика и блоковске компресије 8×8 , што је један од разлога њене примене у JPEG стандарду.

DWT трансформација омогућава вишескалну анализу сигнала и боље очување локалних структура. Посебно добре резултате даје код слика које садрже ивице, нагле промене и текстуре различитих размера. Захваљујући локализацији у времену и фреквенцији, DWT често обезбеђује природнију реконструкцију и мање артефаката у односу на класичне блоковске методе.

5.3 Ефикасност компресије

Ефикасност компресије зависи од тога колико добро трансформација концентрише енергију сигнала у мали број коефицијената. Ако се већина енергије налази у неколико коефицијената, онда је могуће одбацити велики број мање важних коефицијената без значајног губитка квалитета.

DCT је у овом погледу веома погодна за природне слике. Код таквих слика постоји много спорих промена осветљености, па се велики део информације пребацује у нискофреквентне коефицијенте. Због тога је DCT основа класичног JPEG стандарда.

DWT има другачију предност. Она не посматра слику само као скуп фреквенција, већ је разлаже на више нивоа детаља. Због тога је погодна када су важне ивице, локалне структуре и постепена реконструкција слике, што је основа JPEG-2000 стандарда.

FFT се ређе користи за директну компресију слика, јер су њене базисне функције глобалне. Ипак, она остаје веома значајна у анализи аудио сигнала, спектралном филтрирању и задацима у којима је потребно прецизно одредити фреквенцијски садржај сигнала.

5.4 Нумеричка стабилност

Нумеричка стабилност описује колико је поступак осетљив на грешке заокруживања, квантизацију и ограничену прецизност рачунара. У обради слика и звука ове грешке могу бити видљиве или чујне, па стабилност трансформације има практичан значај.

DCT је погодна за рад са реалним сигнаlima, јер користи реалне косинусне функције и добро концентрише енергију. FFT користи комплексне бројеве и захтева пажљивију обраду фаза и амплитуда. DWT има стабилност која зависи од избора вејвлета: Хаар је једноставан и стабилан, док Daubechies вејвлети дају глаткију репрезентацију, али су нумерички сложенији.



Слика 5.1. Визуелно поређење реконструкције у присуству Гаусовог шума, са нагласком на разлике у потискивању шума и очувању детаља.

Слика 5.1 приказује резултате реконструкције добијене након примене различитих трансформација на слику са шумом. Свака метода обрађује сигнал у различитом трансформационом домену, што утиче на њену способност да потисне шум и очува структурне детаље. Ово визуелно поређење помаже у процени колико ефикасно свака трансформација представља суштинске информације сигнала. Квантитативне разлике између метода могу се додатно анализирати коришћењем PSNR метрике.

5.5 Општи закључак упоредне анализе

Упоредна анализа показује да ниједна трансформација није најбоља у свим ситуацијама. FFT је природан избор када је потребна прецизна спектрална анализа, посебно у обради звука и телекомуникацијама. DCT је погодна за класичну компресију слика, јер добро концентрише енергију у малом броју коефицијената. DWT је најкориснија када је важно очување локалних структура, ивица и вишескалних детаља.

6 Примери из реалних мултимедијалних стандарда

У овом поглављу анализирају се практичне примене нумеричких трансформација у најважнијим мултимедијалним форматима. Циљ није само да се наведу стандарди, већ да се покаже зашто се одређена трансформација користи баш у датом контексту.

6.1 JPEG: DCT и квантизација

JPEG стандард заснива се на идеји да се слика најпре подели на мале блокове величине 8×8 пиксела. На сваки блок примењује се DCT-II трансформација, која вредности пиксела преводи у коефицијенте различитих просторних фреквенција.

Ради једноставнијег приказа, најпре се може посматрати једнодимензиона DCT-II трансформација:

$$X[k] = \alpha_k \sum_{n=0}^7 x[n] \cos \left(\frac{\pi(2n+1)k}{16} \right).$$

Код стварног JPEG блока величине 8×8 , ова идеја се примењује у два правца, по врстама и по колонама, чиме се добија дводимензиона DCT.

После трансформације примењује се квантизација:

$$C_q[i, j] = \text{round} \left(\frac{C[i, j]}{Q[i, j]} \right).$$

Матрица $Q[i, j]$ одређује колико ће поједини DCT коефицијенти $C[i, j]$ бити заокружени. Високофреквентни коефицијенти обично се јаче квантизују, јер људско око мање примећује губитак ситних детаља него промену основне структуре слике.

После квантизације следе зиг-заг скенирање, кодирање дужина серија и Huffman кодирање. Већина AC коефицијената постаје нула, што омогућава значајно смањење величине фајла.

6.2 MP3: MDCT и психоакустика

MP3 користи модификовану дискретну косинусну трансформацију (MDCT):

$$X[k] = \sum_{n=0}^{2N-1} x[n] \cos \left(\frac{\pi}{N} \left(n + \frac{1}{2} + \frac{N}{2} \right) \left(k + \frac{1}{2} \right) \right).$$

MDCT користи преклапајуће прозоре, па се прелази између суседних делова сигнала обрађују глаткије него код обичне блоковске анализе. То смањује појаву блок артефаката у звуку. Поред саме трансформације, MP3 користи и психоакустичке моделе слуха. Они описују чињеницу да човек не чује све фреквенције подједнако и да јачи звук може да прикрије слабији звук блиске фреквенције. Зато се део информација може уклонити без великог субјективног губитка квалитета [6].

6.3 MPEG видео компресија: компензација покрета и DCT

MPEG видео компресија комбинује анализу кретања и трансформационо кодирање. Најпре се покушава предвидети како се објекти померају из једног кадра у други. Уместо да се сваки кадар памти у потпуности, чува се разлика између стварног кадра и предвиђене слике [7].

Та разлика назива се резидуални сигнал. На њега се затим примењује DCT, јер је резидуални сигнал често погодан за компресију. После DCT трансформације следе квантизација и ентропијско кодирање. На тај начин MPEG смањује количину података неопходну за чување и пренос видеа.

6.4 JPEG-2000 (DWT): вејвлет компресија

JPEG-2000 користи 2Д вејвлет трансформацију, која слику разлаже на компоненте различитих размера. Математички, 2Д вејвлет трансформација слике $I(x, y)$ може се записати као

$$W_{LL}, W_{LH}, W_{HL}, W_{HH} = \text{DWT}_x(\text{DWT}_y(I(x, y))).$$

Компонента LL представља грубу апроксимацију слике и садржи највећи део енергије сигнала. Компоненте LH , HL и HH садрже хоризонталне, вертикалне и дијагоналне детаље. Овакав приступ омогућава да се слика прикаже постепено: најпре груба верзија, а затим све прецизнији детаљи.

JPEG-2000 је посебно погодан за примене у којима су важни очување детаља и флексибилна реконструкција, на пример у медицини, сателитској обради снимака и архивирању слика високог квалитета.

7 Закључак

У овом раду анализиране су три кључне нумеричке трансформације које чине основу савремених мултимедијалних технологија: брза Фуријеова трансформација (FFT), дискретна косинусна трансформација (DCT) и дискретне вејвлет трансформације (DWT). Свака од ових трансформација има специфичне математичке особине које је чине погодном за различите врсте дигиталних сигнала и примене.

FFT омогућава ефикасну фреквенцијску анализу и неопходна је у аудио обради, дигиталној филтрацији и анализи спектра. Сложеност од $O(N \log N)$ учинила је FFT стандардом у свим областима које захтевају брзу обраду сигнала [8].

DCT је постала кључна трансформација у компресији слика, нарочито у JPEG стандарду. Њена способност да концентрише енергију у мали број коефицијената чини је веома погодном за блоковску компресију слика. У компресији звука, као код MP3 формата, користи се MDCT, односно модификована дискретна косинусна трансформација. Она је блиско повезана са DCT идејом, али је прилагођена обради аудио сигнала помоћу преклапајућих прозора.

Вејвлет трансформације представљају савремени приступ обради сигнала који омогућава симултану локализацију и у времену и у фреквенцији. DWT је основни механизам JPEG-2000 стандарда и даје боље резултате при високим стопама компресије, нарочито за слике са наглим променама и ивицама.

Упоредна анализа показује да ниједна трансформација није оптимална за све проблеме: FFT доминира у анализи сигнала, DCT у компресији слика, док DWT омогућава најквалитетнију вишескалну репрезентацију. Заједно, ове трансформације чине математички темељ целокупне модерне мултимедијалне индустрије.

Разумевање нумеричких метода и њихових својстава омогућава развој ефикаснијих алгоритама и напреднијих стандарда у будућности. Даљи развој трансформација, укључујући адаптивне и хибридне приступе, отвара могућности за још квалитетнију и ефикаснију обраду мултимедијалног садржаја.

Надам се да ћу у наставку мог школовања и сам моћи да допринесем даљем развоју ове важне области чију примену видимо у свакодневном животу.

Захвалница

Захваљујем свим мојим професорима у основној школи и Математичкој гимназији који су током мог школовања, својим саветима, предлозима и подршком, посредно допринели изради овог рада. Велика је привилегија што ми је најважнији предмет у гимназији предавала разредна, професорка анализе са алгебром, Милица Мисојчић Огњановић, чији су рад, знање и посвећеност значајно утицали на мој однос према математици.

Пошто овај рад повезује знања из математике, физике, рачунарства и информатике, захвалност дугујем и професору физике Дејану Ђокићу, чији су часови и научни приступ допринели мом разумевању физичких појава и методологије истраживања. Такође, захваљујем професорки информатике и привременој разредној, Нини Алимпић, која ме је увела у свет дигиталног представљања сигнала и инжињерског начина размишљања.

Највећу захвалност ипак дугујем свом ментору, професору Вукашину Брковићу, на стрпљењу, разумевању и великој помоћи током читавог процеса израде рада.

Хвала професору на томе што је кроз предмет Нумеричка математика приближио сложене математичке моделе и омогућио нам да их не само теоријски разумемо, већ и практично имплементирамо у конкретном програмском језику. Такав приступ значајно је допринео мом интересовању за примену математике у информатици и физици и обради сигнала.

Посебно сам захвалан на томе што је ментор увек брзо, конкретно и са великом посвећеношћу одговарао на питања и помагао у разјашњавању недоумица током рада. Његови савети и смернице били су веома значајни за коначан изглед и квалитет овог рада.

Такође, захваљујем мојој сестри и браћи и другарима на подршци, охрабрењу и разумевању током израде матурског рада.

Литература

- [1] A. V. Oppenheim, R. W. Schaffer, *Discrete-Time Signal Processing*, Prentice Hall, 2010.
- [2] R. Bracewell, *The Fourier Transform and Its Applications*, McGraw–Hill, 2000.
- [3] G. Strang, T. Nguyen, *Wavelets and Filter Banks*, Wellesley–Cambridge Press, 1997.
- [4] G. Wallace, “The JPEG Still Picture Compression Standard”, *Communications of the ACM*, 1991.
- [5] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, 1992.
- [6] K. Brandenburg, “MP3 and AAC Explained”, *AES Journal*, 1999.
- [7] J. Cooley, J. Tukey, “An Algorithm for the Machine Calculation of Complex Fourier Series”, *Mathematics of Computation*, 1965.
- [8] D. Taubman, M. Marcellin, *JPEG2000: Image Compression Fundamentals, Standards and Practice*, Springer, 2002.