МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение
высшего образования
«Национальный исследовательский
Нижегородский государственный университет им. Н.И. Лобачевского»
(ННГУ)

Институт информационных технологий, математики и механики
Кафедра алгебры, геометрии и дискретной математики

Направление подготовки:
«Фундаментальная информатика и информационные технологии»
Профиль подготовки: «Инженерия программного обеспечения»

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА
БАКАЛАВРА

Тема:
«Эволюция алгоритмов сжатия изображений: от RLE до JPEG»

Выполнил:
студент группы: 3821Б1ФИ2
Казанцев Евгений Александрович

_____

Научный руководитель:
к.ф.-м.н., доцент каф. АГДМ
Смирнова Татьяна Геннадьевна

_____

Нижний Новгород
2025

Аннотация

Сжатие изображений — применение алгоритмов сжатия данных к изображениям, хранящимся в цифровом виде. В результате сжатия уменьшается размер изображения, из-за чего уменьшается время передачи изображения по сети и экономится пространство для хранения.

Сжатие изображений подразделяют на сжатие с потерями качества и сжатие без потерь.

Преимущество методов сжатия с потерями над методами сжатия без потерь состоит в том, что первые дают возможность большей степени сжатия, при этом, продолжая удовлетворять поставленным требованиям, а именно — искажения должны быть в допустимых пределах чувствительности органов человеческого восприятия.

Методы сжатия с потерями используются для сжатия аналоговых данных — чаще всего звука или изображений.

В таких случаях распакованный файл может сильно отличаться от оригинала по размеру, но практически неотличим для человека «на слух» и «на глаз» в большинстве применений.

Множество методов фокусируются на физических особенностях органов чувств человека. К примеру психоакустические модели определяют то, как сильно звук может быть сжат без ухудшения воспринимаемого человеком качества звука.

При использовании сжатия с потерями необходимо учитывать, что повторное сжатие обычно приводит к деградации качества. Однако, если повторное сжатие выполняется без каких-либо изменений сжимаемых данных, качество не меняется. Так например, сжатие изображения методом JPEG, восстановление его и повторное сжатие с теми же самыми параметрами не приведет к снижению качества.

ОГЛАВЛЕНИЕ

Введение

Сжатие информации с потерями — это алгоритмическое преобразование данных, направленное на уменьшение их объёма за счёт допустимых потерь качества. Необходимость сокращения объёма данных актуальна с появлением цифровых технологий и остается актуальной по сей день. Несмотря на развитие технологий хранения и передачи данных, ограничения по объёму памяти и скорости передачи информации по-прежнему требуют эффективных методов сжатия. Объёмы данных продолжают расти, их структура усложняется, а качество — повышается. Это особенно актуально для мультимедийных данных — изображений, видео и звука.

В качестве основных литературных источников данной работы используются: книга Д. Сэломона «Сжатие данных, изображений и звука» [1], труды П. Пенна [2], а также две книги Т. Рафгардена из серии «Совершенный алгоритм» [3,4]. В книге [1] содержится формальное описание алгоритмов сжатия с потерями, в частности алгоритма JPEG, включая его математическую основу — дискретное косинусное преобразование (DCT). Труды П. Пенна [2] дают детализированное объяснение реализации алгоритма JPEG на программном уровне. Книги [3,4] содержат информацию о базовых алгоритмических принципах, используемых в методах сжатия. В ходе работы также использовалась официальная документация библиотеки OpenCV [5] и языка программирования Python 3 [6], а так же библиотеки для создания графического интерфейса PyQt [7].

Вернуться и доработать

Существует огромное множество методов сжатия данных, основанных на различных предположениях о структуре сжимаемой информации. Все алгоритмы сжатия можно разделить на две основные группы: Сжатие без потерь — предполагает возможность точного восстановления исходных данных после сжатия. Сжатие с потерями — допускает потерю части данных ради значительного сокращения объема. Этот метод часто применяется к изображениям, видео и аудиоданным, где некоторая потеря качества незаметна для восприятия.

В данной работе рассматривается алгоритм сжатия с потерями на примере широко используемого метода JPEG. Алгоритм JPEG основан на использовании дискретного косинусного преобразования (DCT) для преобразования изображения, последующем квантовании и энтропийном кодировании. В рамках данной реализации из рассмотрения исключаются этапы энтропийного кодирования, в частности кодирование Хаффмана, а также альтернативные подходы, такие как вейвлет-преобразования. Этот метод обеспечивает высокую степень сжатия при сохранении приемлемого качества изображения. JPEG стал важным этапом в развитии технологий сжатия данных, объединив достижения математических и инженерных исследований своего времени, и дал мощный толчок развитию в области сжатия изображений.

Основной задачей работы является исследование алгоритма JPEG в контексте его математической основы и алгоритмической реализации. Предполагается рассмотреть влияние различных параметров сжатия на качество изображения, а также провести сравнение производительности различных реализаций алгоритма. В качестве возможных модификаций рассматриваются методы адаптивного квантования и постобработки изображения после декодирования.

Целью работы является изучение развития алгоритмов сжатия с потерями, начиная с фундаментальных математических открытий (например, дискретного косинусного преобразования), простых алгоритмов сжатия без потерь, и заканчивая созданием аналогичного алгоритма JPEG. В этой работе реализуется собственный алгоритм, основанный на ключевых этапах стандарта JPEG. Его структура почти аналогична базовому варианту. В дальнейшем по тексту он будет обозначаться как “модификация алгоритма JPEG” или просто “JPEG”. Так же планируется разработать программу с интерфейсом, позволяющую проводить эксперименты по сжатию изображений с использованием модификации алгоритма JPEG. Программа должна обладать следующими функциями:

- кодирование изображения методом JPEG с возможностью настройки параметров квантования и качества;
- визуализация результатов кодирования и декодирования;
- сравнение качества изображения до и после сжатия с использованием метрик SSIM и PSNR;
- возможность включения адаптивного квантования для улучшения качества при низких битрейтах.

Для достижения поставленной цели сформулированы следующие задачи:

- изучить теоретическую основу алгоритма JPEG, включая дискретное косинусное преобразование и квантование;
- реализовать алгоритм JPEG на языке программирования Python с использованием библиотеки OpenCV;
- разработать интерфейс для управления параметрами алгоритма и отображения результатов;
- провести эксперименты на различных изображениях, оценить влияние параметров на степень сжатия и качество изображения;
- проанализировать результаты экспериментов и сделать выводы о целесообразности использования различных модификаций алгоритма.

1 Основные понятия и определения

Перед тем как углубиться в описание алгоритма JPEG, важно рассмотреть основные понятия, лежащие в основе цифровой обработки и сжатия изображений.

Цифровое изображение — двумерный массив, где каждая ячейка представляет собой пиксель.

Таблица 1. Упрощенная схема цифрового изображения как двумерного массива пикселей

P(0,0)	P(0,1)	P(0,2)	P(0,3)
P(1,0)	P(1,1)	P(1,2)	P(1,3)
P(2,0)	P(2,1)	P(2,2)	P(2,3)
P(3,0)	P(3,1)	P(3,2)	P(3,3)

Пиксель (pixel) — наименьшая единица изображения, содержащая информацию о цвете или яркости.

Таблица 2. Представление пикселей с помощью RGB-значений и их цветов

RGB значения	**Цвет**
RGB(255,0,0)	
RGB(0,255,0)	
RGB(0,0,255)	

Наиболее распространённой моделью представления цветных изображений является RGB.

RGB (Red, Green, Blue) — аддитивная цветовая модель, в которой каждый цвет формируется путём смешивания трёх базовых компонентов: красного, зелёного и синего. Эта модель близка к принципу работы матриц дисплеев и сенсоров цифровых камер, поэтому RGB часто используется как исходный формат цветовых изображений.

В зависимости от глубины цвета, каждый пиксель кодируется определённым числом бит. Например, в изображении с глубиной 24 бита (8 бит на каждый из каналов RGB) один пиксель может представлять более 16 миллионов оттенков. С увеличением разрешения и глубины цвета возрастает и общий объём изображения, что создаёт необходимость в эффективном сжатии данных.

Однако при сжатии изображений RGB-пространство не является наиболее эффективным. Это связано с тем, что RGB не разделяет яркость и цветовую информацию, а человеческий глаз по-разному чувствителен к этим компонентам. В связи с этим перед сжатием изображения, как правило, преобразуются в цветовое пространство **YCbCr**, где **Y** отражает яркость (luminance), а **Cb** и **Cr** — цветовые отклонения от серого (chrominance).

Также необходимо упомянуть несколько ключевых понятий, тесно связанных с задачами кодирования изображений:

Разрешение изображения — количество пикселей по горизонтали и вертикали.

Избыточность — повторяющаяся или избыточная информация, которую можно исключить без потери качества или с минимальными потерями.

Энтропия — мера неопределённости или информационной насыщенности данных, определяющая нижнюю границу возможной степени сжатия.

Формат **JPEG** представляет собой стандарт кодирования изображений с потерями, определяющий последовательность преобразований, квантования и кодирования данных. Разработка формата началась в 1986 году с формированием международной рабочей группы под названием *Joint Photographic Experts Group*. На тот момент существовало множество несовместимых форматов для хранения графики, что затрудняло обмен изображениями между системами. JPEG стал первым по-настоящему универсальным и широко распространённым стандартом сжатия фотоизображений. Официальная спецификация была опубликована в 1992 году и используется до сих пор.

2 Алгоритмы и структуры данных

2.1 Преобразование цветового пространства (RGB → YCbCr)

Перед сжатием изображений в формате JPEG зачастую используется преобразование из цветового пространства RGB в YCbCr. Это обусловлено как особенностями восприятия цвета человеком, так и требованиями алгоритмов сжатия.

Такое разделение позволяет применить хрома-субдискретизацию — уменьшение разрешения цветовых компонент — без заметного ухудшения качества изображения. Для RGB с диапазоном значений от 0 до 255, преобразование в YCbCr выполняется по следующим формулам:

$$\begin{aligned} Y &= 0.299R + 0.587G + 0.114B; \\ Cb &= -0.168736R - 0.331264G + 0.5B + 128; \\ Cr &= 0.5R - 0.418688G - 0.081312B + 128. \end{aligned} \tag{1}$$

Здесь значения R, G, B принимаются в диапазоне $[0, 255]$. Смещение на 128 единиц в формулах для Cb и Cr необходимо для корректного представления как положительных, так и отрицательных значений в беззнаковом формате (unsigned int).

Поскольку компоненты Cb и Cr менее значимы для восприятия, их можно хранить с пониженным разрешением. Наиболее распространённый формат — 4:2:0, при котором на каждые 4 пикселя хранится 4 значения яркости Y и по 1 значению Cb и Cr. Это даёт значительное уменьшение объёма данных без существенного ущерба для качества изображения.

2.2 Преимущества YCbCr

- Обеспечивает эффективное сжатие без видимых потерь качества;
- Разделение яркости и цвета позволяет применять различные методы компрессии к компонентам Y и CbCr;
- Поддерживается большинством стандартов обработки изображений и видео.

Таким образом, преобразование RGB $\rightarrow$ YCbCr является ключевым шагом в JPEG и других алгоритмах сжатия, позволяющим использовать особенности восприятия цвета человеком для достижения более высокой степени сжатия.

2.3 Субдискретизация

На втором этапе сжатия изображения применяется методика, известная как субдискретизация цветности (англ. chroma subsampling). Её ключевая идея заключается в сокращении объёма данных, содержащих информацию о цвете, путём уменьшения пространственного разрешения цветовых компонентов. Это возможно благодаря физиологической особенности зрительной системы человека: она в гораздо большей степени чувствительна к деталям яркости (компонент Y — luminance), чем к изменениям цветовых оттенков (компоненты Cb и Cr — chrominance).

В практической реализации данный метод означает, что цветовая информация (Cr и Cb) кодируется с меньшей точностью по сравнению с яркостной. Существует несколько вариантов схем хрома-субдискретизации, которые различаются по степени уменьшения разрешения цветовых компонентов:

4:4:4 — нет сжатия, все компоненты Y, Cb и Cr сохраняют оригинальное разрешение.

4:2:2 — разрешение Cb и Cr уменьшается в два раза по сравнению с Y.

4:2:0 — стандарт для большинства сжатых форматов, где разрешение Cb и Cr уменьшается в два раза как по вертикали, так и по горизонтали.

Схема 4:2:0 является наиболее распространенной для видео и изображений, так как она дает хороший баланс между качеством и сжатием. Она предполагает, что на каждые четыре пикселя яркости (расположенных в виде блока 2×2) приходится всего один усреднённый пиксел Cr и один усреднённый пиксел Cb. То есть цветовые значения вычисляются как средние значения для группы пикселей, и это значение применяется ко всей группе. В результате разрешение цветовых компонентов уменьшается в четыре раза (на 75%) по сравнению с яркостной компонентой.

Для визуализации можно представить исходное изображение как состоящее из трёх равнозначных слоёв:

Y (яркость),

Cb (синий цветовой компонент),

Cr (красный цветовой компонент).

До субдискретизации каждый из этих компонентов имел одинаковое разрешение и соответственно одинаковый вклад в общий объём данных:

$$Y + Cb + Cr = 1 + 1 + 1 = 3 \text{ единицы информации.} \tag{2}$$

После применения схемы 4:2:0, каждая из цветовых компонент уменьшается до $\frac{1}{4}$ исходного размера, а яркостная остаётся без изменений:

$$Y + \frac{1}{4}Cb + \frac{1}{4}Cr = 1 + 0.25 + 0.25 = 1.5 \text{ единицы информации.} \tag{3}$$

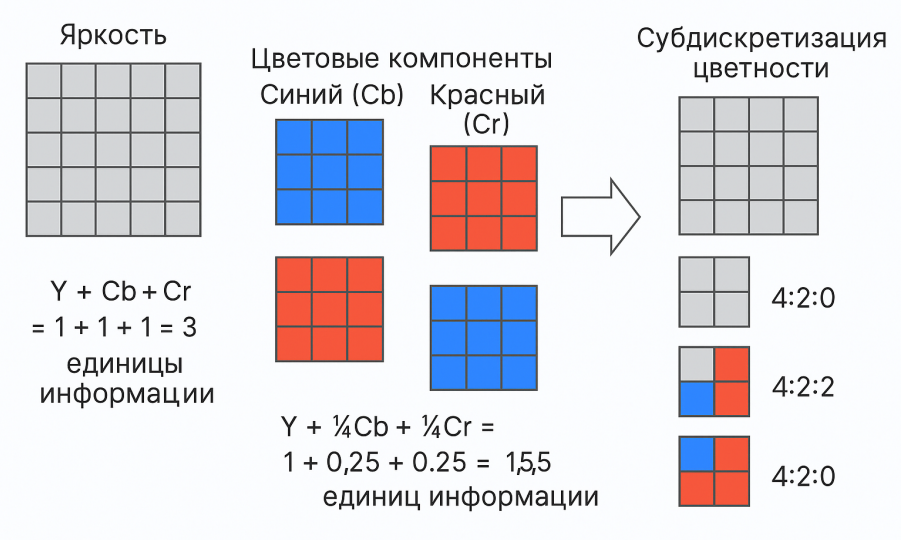


Рисунок 1. Как происходит преборазование

Таким образом, достигается двукратное уменьшение общего объёма данных, подлежащих хранению или передаче.

Что важно, такое преобразование происходит практически без визуальных потерь качества. Несмотря на то, что цветовые данные редуцируются, человеческий глаз практически не воспринимает различие между оригинальным и сжатым изображением. Именно поэтому субдискретизация цветности широко применяется в большинстве алгоритмов сжатия изображений и видео (JPEG, MPEG, H.264), где критически важно достичь компромисса между качеством и эффективностью хранения данных.

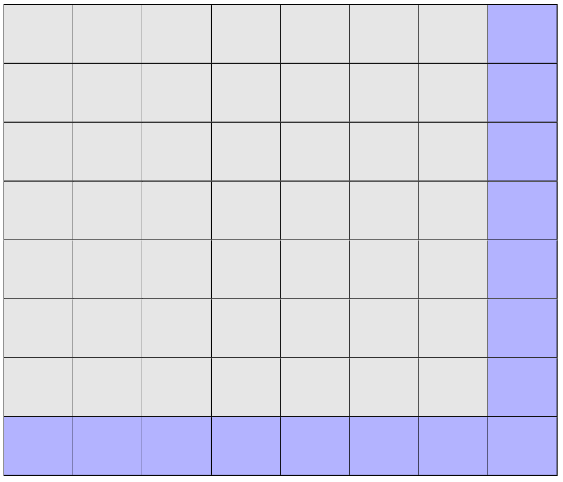
2.4 Дискретное косинус-преобразование

После преобразования изображения из цветового пространства RGB в YCbCr и выполнения хрома-субдискретизации (уменьшения разрешения цветовых компонент Cb и Cr), выполняется разбиение изображения на блоки фиксированного размера 8×8 пикселей. Однако размеры исходного изображения не всегда кратны восьми, что делает невозможным непо-

средственное и равномерное разбиение без остатка. В таких случаях производится предварительная корректировка размеров изображения: вычисляется необходимое количество дополнительных строк и столбцов, которые должны быть добавлены к правому и нижнему краю изображения для достижения кратности 8.

Добавление недостающих строк и столбцов осуществляется с помощью повторения граничных значений (пикселей) изображения, чтобы минимизировать искажения, возникающие на краях.

Таблица 3. Расширенный блок 8×8 с добавленными пикселами



Этот шаг реализуется при помощи библиотеки OpenCV, предоставляющей функцию для симметричного дополнения изображений. После этого скорректированное изображение разбивается на непересекающиеся блоки размером 8×8.

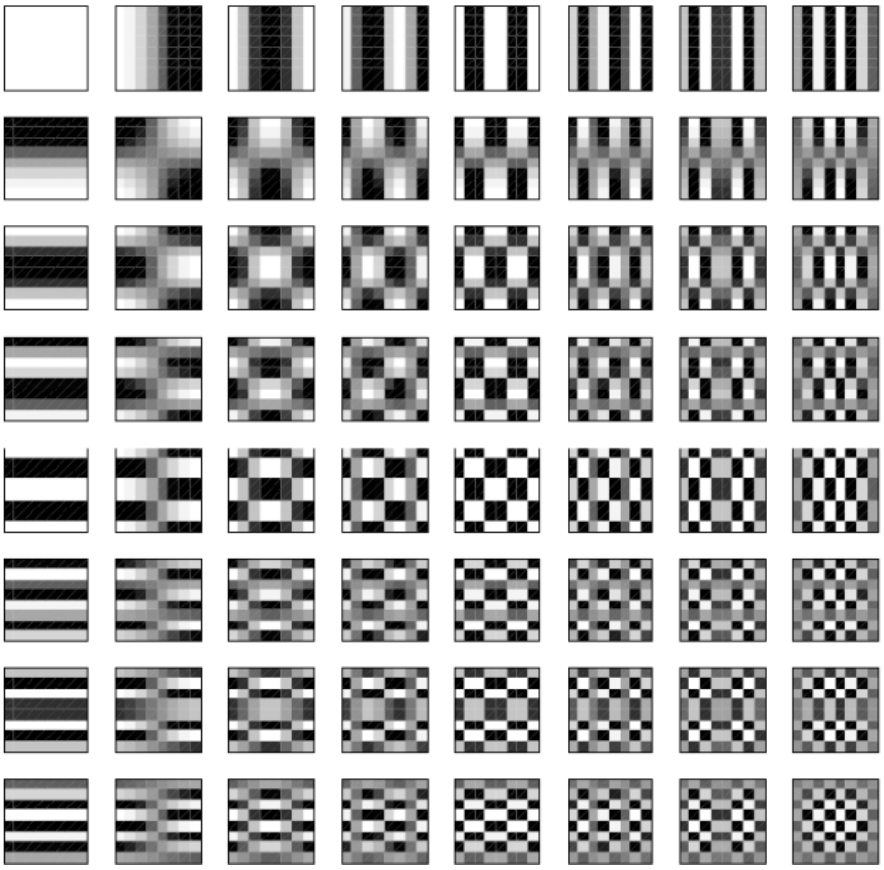


Рисунок 2. Блоки 8x8

Для удобства хранения и последующей обработки эти блоки сохраняются в виде многомерных массивов с использованием библиотеки NumPy. Это позволяет эффективно обращаться к отдельным блокам, выполнять над ними математические преобразования и организовать дальнейший процесс.

Это важный шаг, позволяющий значительно повысить эффективность сжатия. Разделение на небольшие блоки позволяет применять алгоритмы сжатия локально, что упрощает обработку и улучшает степень компрессии.

Дальнейшая задача заключается в оценке уровня детализации внутри каждого блока. Если блок содержит мало визуальных деталей, его можно закодировать с меньшим числом битов, минимизируя потери качества. Для анализа степени детализации и выделения значимых компонент используется дискретное косинусное преобразование (ДКП, англ. Discrete Cosine Transform, DCT). Преобразование позволяет перейти от пространственного представления блока к частотному, выявляя частотные составляющие, которые имеют наибольшее значение для визуального восприятия.

Для понимания принципов работы ДКП целесообразно сначала рассмотреть его одномерную (векторную) форму, поскольку она обладает большей наглядностью и сохраняет все ключевые идеи, лежащие в основе двумерного варианта, применяемого в обработке изображений.

На рисунке 3 показано восемь волн косинуса, $\omega(f) = \cos f\Theta$, при $0 \leq \Theta \leq \pi$ с частотами $f = 0, 1, \ldots, 7$.

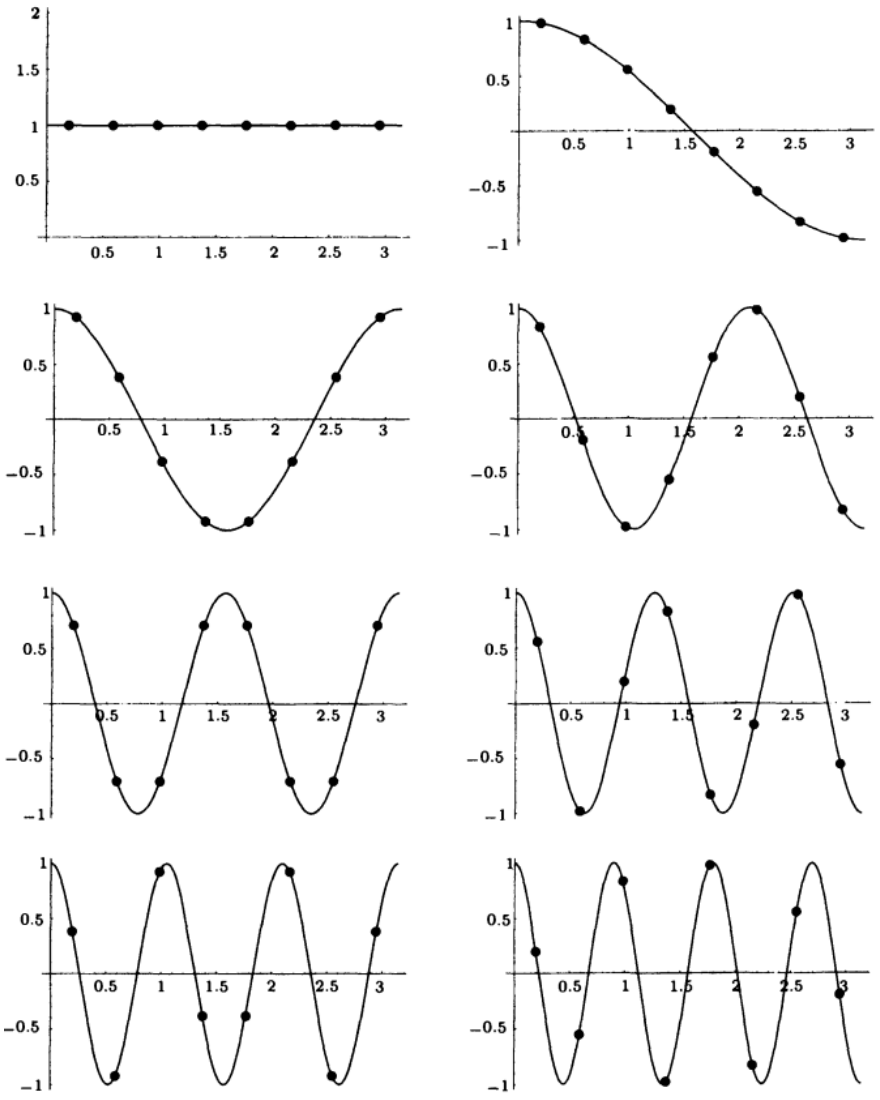


Рисунок 3. Графики функций $\omega(f) = \cos(f\Theta)$ для различных частот f.

На каждом графике отмечено восемь значений функции $\omega(f)$ с абсциссами

$$\Theta = \frac{\pi}{16}, \frac{3\pi}{16}, \frac{5\pi}{16}, \frac{7\pi}{16}, \frac{9\pi}{16}, \frac{11\pi}{16}, \frac{13\pi}{16}, \frac{15\pi}{16} \tag{4}$$

которые формируют базисный вектор v_f. В результате получится восемь векторов $v_f, f = 0, 1, \ldots, 7$ (всего 64 числа).

Таблица 4. Таблица значений $\cos(k\theta)$ для разных k и θ

θ	0.196	0.589	0.982	1.374	1.767	2.160	2.553	2.945
$\cos 0\theta$	1.	1.	1.	1.	1.	1.	1.	1.
$\cos 1\theta$	0.981	0.831	0.556	0.195	-0.195	-0.556	-0.831	-0.981
$\cos 2\theta$	0.924	0.383	-0.383	-0.924	-0.924	-0.383	0.383	0.924
$\cos 3\theta$	0.831	-0.195	-0.981	-0.556	0.556	0.981	0.195	-0.831
$\cos 4\theta$	0.707	-0.707	-0.707	0.707	0.707	-0.707	-0.707	0.707
$\cos 5\theta$	0.556	-0.981	0.195	0.831	-0.831	-0.195	0.981	-0.556
$\cos 6\theta$	0.383	-0.924	0.924	-0.383	-0.383	0.924	-0.924	0.383
$\cos 7\theta$	0.195	-0.556	0.831	-0.981	0.981	-0.831	0.556	-0.195

Они служат базисом одномерного косинусного преобразования, причём частота смены знаков увеличивается по строкам.

Можно показать, что все векторы v_i ортогональны между собой (из-за специального выбора восьми точек отсчета Θ). То же самое можно обнаружить прямым вычислением с помощью подходящей математической программы. Значит, эти восемь векторов можно поместить в матрицу размером 8 на 8 и рассмотреть соответствующее ей ортогональное преобразование - вращение в восьмимерном пространстве, которое называется одномерным дискретным косинус-преобразованием (DCT).

Одномерное преобразование косинусов Дискретное (DCT) можно интерпретировать как представление вектора в базисе, состоящем из векторов v_i В этом контексте любой вектор p из соответствующего векторного пространства может быть выражен через линейную комбинацию этих базисных векторов v_j.

Например, выберем 8 (коррелированных) чисел $p = (0.6, 0.5, 0.4, 0.5, 0.6, 0.5, 0.4, 0.55)$ в качестве тестовых данных. Выразим вектор p в виде суммы $p = \sum_i \omega_i v_i$ восьми веторов v_i. Решив эту систему из 8 линейных уравнений, находим восемь весов

$$\begin{aligned} &\omega_0 = 0.506, \omega_1 = 0.0143, \omega_2 = 0.0115, \omega_3 = 0.0439, \\ &\omega_4 = 0.0795, \omega_5 = -0.0432, \omega_6 = 0.00478, \omega_7 = -0.0077. \end{aligned} \tag{5}$$

Вес ω_0 не сильно отличается от элементов вектора p , но остальные семь весов гораздо меньше. Это показывает, как DCT (или любое другое ортогональное преобразование) производит сжатие. Теперь можно просто записать эти восемь весов в сжатый файл, где онибудут занимать меньше места, чем восемь компонентов исходного вектора p.

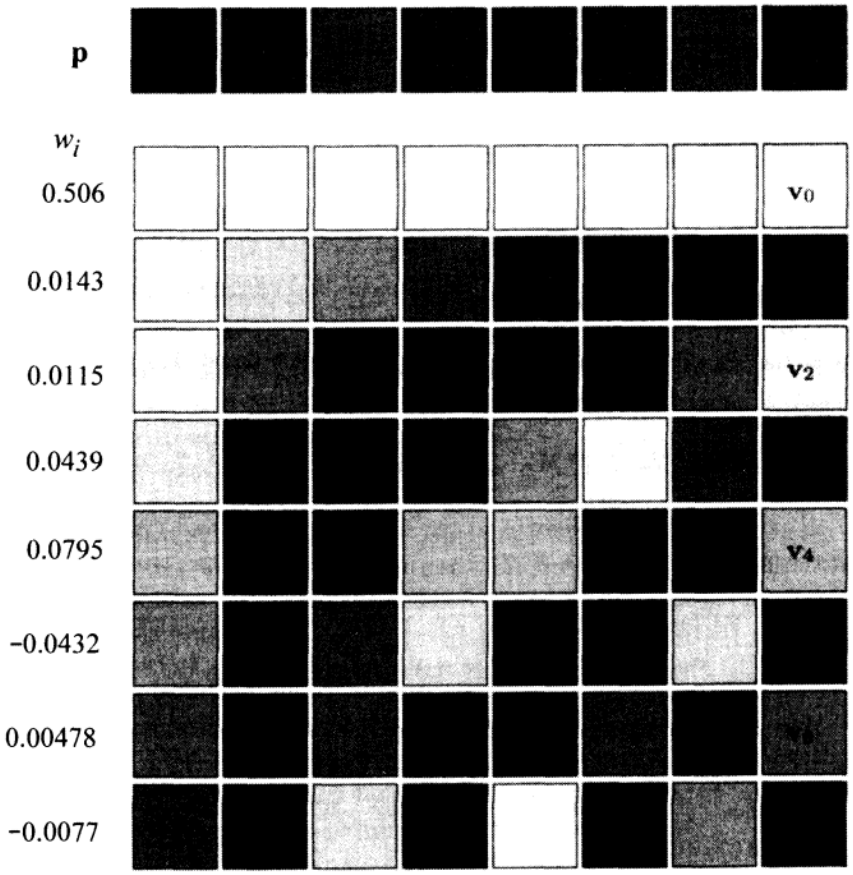


Рисунок 4. Графическое представление одномерного DCT.

Все восемь векторо v_i показаны в виде ряда из восьми маленьких серых квадратиков, причем значение $+1$ представлено белым цветом, а значение -1 окрашено в ченый цвет. Каждый из восьми компонентов вектора p выражен в виде взвешенной суммы восьми серых оттенков.

На практике одномерное DCT проще всего вычислять по формуле

$$G_f = \frac{1}{2}C_f \sum_{t=0}^{7} p_t \cos\frac{(2t+1)f\pi}{16} \tag{6}$$

$$\text{где} \quad C_f = \begin{cases} \frac{1}{\sqrt{2}}, & f = 0, \\ 1, & f > 0. \end{cases} \quad \text{При} \quad f = 0, 1, \ldots, 7.$$

Здесь исходными данными (пикселами, фрагментами звука или другими элементами) являются величины p_t, а им соответствующими коэффициентами DCT служат числа G_f. Формула (6) очень проста, но процесс вычисления по ней медленный. Декодер получает на входе коэффициенты DCT, делит их на восьмерки и применяет к ним обратное преобразование DCT (inverse DCT, IDCT) для восстановления исходных данных (тоже в виде групп по 8 элементов). Простейшая формула для вычисления IDCT имеет вид

$$p_t = \frac{1}{2}\sum_{j=0}^{7} C_j G_j \cos\frac{(2t+1)j\pi}{16}, \text{при } t = 0, 1, \ldots, 7 \tag{7}$$

2.5 Двумерное (матричное) DCT

Из опыта хорошо известно, что пикселы изображения имеют корреляцию не только по горизонтали, но и по вертикали. То есть пиксели взаимосвязаны как с соседними пикселами слева и справа, так и с пикселами сверху и снизу. Поэтому методы сжатия изображений используют двумерное DCT, которое задается формулой

$$G_{ij} = \frac{1}{\sqrt{2n}}C_i C_j \sum_{x=0}^{n-1}\sum_{y=0}^{n-1} p_{xy} \cos\frac{(2y+1)j\pi}{2n} \cos\frac{(2x+1)i\pi}{2n} \tag{8}$$

при $0 \le i, j \le n-1$. Изображение разбивается на блоки пикселов p_{xy} размера $n \times n$ (в данном примере $n = 8$), и уравнение (8) используется для нахождения коэффициентов G_{ij} для каждого блока пикселов. Если частичная потеря информации допустима, то коэффициенты подвергаются квантованию, что будет подробно рассмотрено в следующем разделе. Декодер восстанавливает сжатый блок данных (точно или приближенно), вычисляя обратное DCT (IDCT) по формуле

$$p_{xy} = \frac{1}{4}\sum_{i=0}^{7}\sum_{j=0}^{7} C_i C_j G_{ij} \cos\frac{(2y+1)j\pi}{16} \cos\frac{(2x+1)i\pi}{16} \tag{9}$$

$$\text{где} \quad C_f = \begin{cases} \frac{1}{\sqrt{2}}, & f = 0, \\ 1, & f > 0. \end{cases}$$

Двумерное дискретное косинусное преобразование (DCT) можно интерпретировать двумя способами: как композицию двух вращений, а также через представление вектора в базисе n-мерного векторного пространства. В первой интерпретации используется блок $n \times n$ пикселов (рис. (5)a, где элементы обозначены буквой "L")

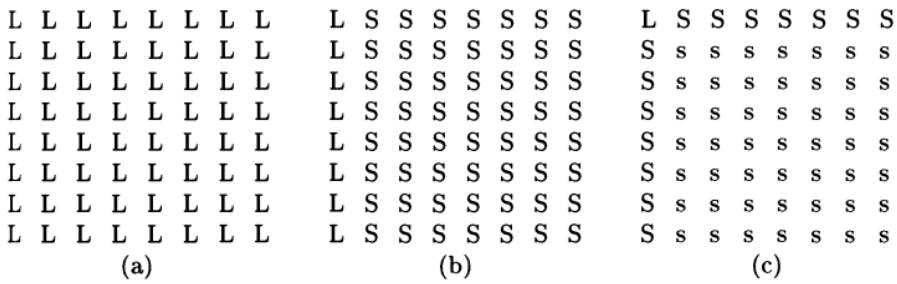


Рисунок 5. Двумерное DCT и двойное вращение.

Сначала рассматриваются строки этого блока как точки $(p_{x,0}, p_{x,1}, \ldots, p_{x,n-1})$ в n- мерном пространстве, которые поворачиваются в этом пространстве с помощью преобразования, задаваемого внутренней суммой

$$G1_{x,j} = C_j \sum_{y=0}^{n-1} p_{xy} \cos \frac{(2y+1)j\pi}{2n}$$

из уравнения (8). Результатом этого вращения служит блок $G1_{x,j}$ из $n \times n$ коэффициентов, в котором в строках доминируют первые элементы (обозначенные как «L» на рис. (5)b), в то время как все остальные элементы малы (они обозначены как «S»). Внешняя сумма, приведенная в уравнении (8), равна

$$G_{ij} = \frac{1}{\sqrt{2n}} C_i \sum_{x=0}^{n-1} p_{xy} G1_{x,j} \cos \frac{(2x+1)i\pi}{2n}$$

Здесь уже столбцы матрицы $1_{x,j}$ рассматриваются в качестве точек n - мерного векторного пространства, над которыми совершается преобразование вращения. В результате получается один большой коэффициент в верхнем левом углу блока (на рис.(5)c - это «L») и $N^2 - 1$ маленьких коэффициентов в остальных местах («S» и «s» на рисунке). Эта интерпретация рассматривает двумерное DTC в виде двух разных вращений размерности n.

Вторая интерпретация (при $n = 8$) использует уравнение (8) для создания 64 блоков по 8×8 величин в каждом. Все 64 блока рассматриваются в качестве базиса 64-мерного векторного пространства (это базисные изображения). Любой блок B из 8×8 пикселов можно выразить как линейную комбинацию этих базисных изображений, и все 64 веса этой линейной комбинации образуют коэффициенты DCT блока B.

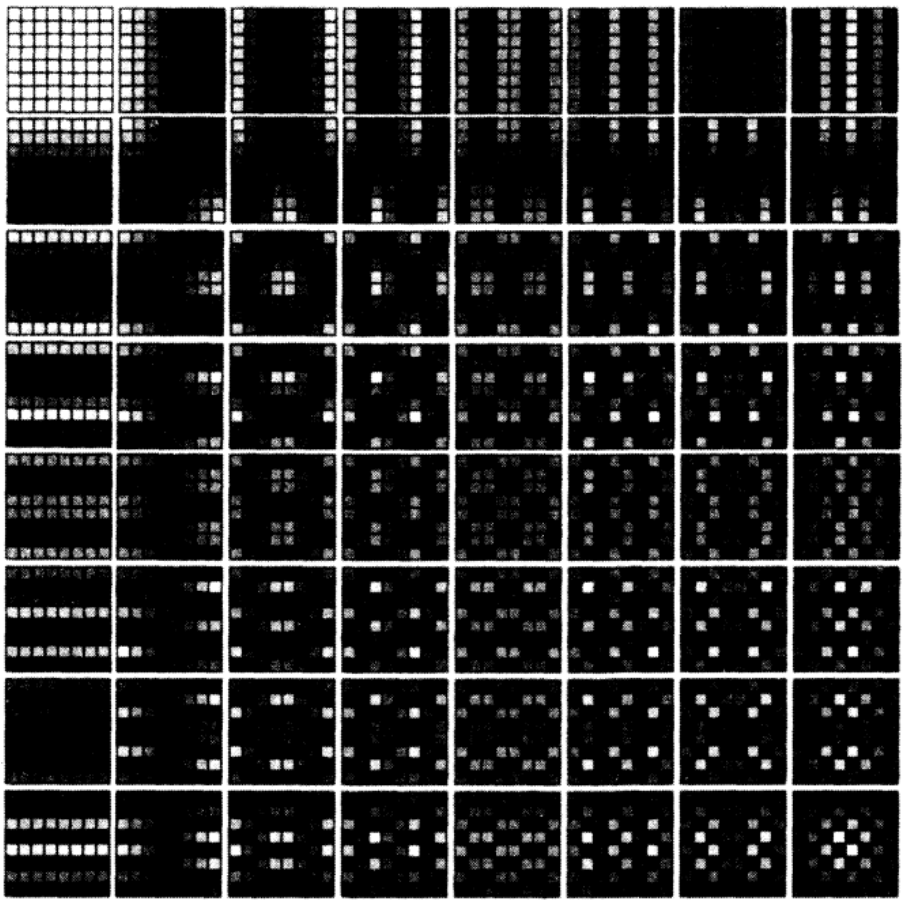


Рисунок 6. 64 базисных изображения двумерного DCT.

На рисунке выше представлено графическое отображение 64 базисных функций двумерного дискретного косинусного преобразования (DCT) при размере блока $n = 8$. Каждый элемент с координатами (i, j) на данном рисунке соответствует блоку размером 8×8, полученному в результате вычисления выражения $\cos(i \cdot s) \cdot \cos(j \cdot t)$, где переменные s и t изменяются независимо в пределах, определённых уравнением (4).

Подведем итог этого, довольно сложного шага. Сжатие любого изображения с помощью DCT можно теперь сделать следующим образом.

1) Разделить изображение на k блоков пикселов размером $n \times n$ (чаще всего используется размер 8×8).
2) Применить дискретное косинусное преобразование (DCT) к каждому блоку $B^{(i)}$, представив его в виде линейной комбинации 64 базисных функций, показанных на рисунке выше. В результате получится набор блоков (далее будем называть их векторами) $W^{(i)}$, каждый из которых содержит 64 коэффициента $\omega_j^{(i)}$, где $j = 0, 1, \ldots, 63$.
3) Векторы $W^{(i)}$, где $i = 1, 2, \ldots, k$, группируются по компонентам, формируя 64 отдель-

ных вектора коэффициентов

$$\mathbf{C_j} = \{\omega_j^{(1)}, \omega_j^{(2)}, \ldots, \omega_j^{(k)}\}.$$

В частности, вектор $\mathbf{C_0}$ состоит из k коэффициентов DC.

4) Каждый вектор коэффициентов $C^{(j)}$ подвергается квантованию независимо от остальных. Полученные квантованные векторы $Q^{(j)}$ далее могут быть дополнительно сжаты с использованием методов, таких как RLE, код Хаффмана или других алгоритмов, и записаны в сжатый файл.

2.6 Квантование

В процессе сжатия изображений важным этапом является **квантование**, под которым понимается преобразование чисел с высокой точностью (чаще всего вещественных) в значения с меньшей точностью, например — округление до ближайшего целого или замена на приближённое значение из ограниченного набора.

Это позволяет существенно уменьшить объём передаваемой или сохраняемой информации за счёт исключения наименее значимых деталей, воспринимаемых человеком слабо либо вовсе незаметных.

Существует два основных подхода к квантованию — **скалярное** и **векторное**.

Скалярное квантование представляет собой поэлементную обработку данных, когда каждое значение преобразуется независимо от других. Этот метод прост в реализации и интуитивно понятен, но не всегда обеспечивает оптимальный результат: полезная информация может быть утеряна даже в значимых фрагментах, если не учитывать взаимосвязь между соседними значениями.

Для повышения эффективности используется **векторное** квантование, где обрабатываются не отдельные числа, а целые блоки данных — группы пикселей, которые рассматриваются как многомерные векторы. Перед началом сжатия формируется специальный набор типовых блоков — кодовая книга. Каждый блок изображения сравнивается со всеми элементами этой книги, и выбирается наиболее близкий (в смысле минимальной разности значений). Вместо самого блока в выходной поток данных записывается лишь индекс (или указатель) соответствующего эталонного блока из кодовой книги.

Поскольку индекс, как правило, занимает меньше места, чем полный блок, такой подход позволяет добиться существенного уменьшения размера изображения. При этом достигается разумный баланс между степенью сжатия и качеством восстановления, особенно если кодовая книга заранее адаптирована под характеристики изображений, с которыми предполагается работать.

В стандарте JPEG реализовано канально-зависимое скалярное квантование, при котором каждый коэффициент, полученный в результате дискретного косинусного преобразова-

ния (DCT), квантуется независимо от остальных. Для каждого из цветовых компонентов — яркостного (Y) и двух хроминансных (Cb и Cr) — применяются отдельные квантовочные матрицы, отражающие различную чувствительность человеческого зрения к яркости и цвету. Эти матрицы сформированы на основе психовизуальных моделей и оптимизированы так, чтобы минимизировать визуально заметные искажения при сжатии.

Матрица квантования яркости (Y):

$$\begin{bmatrix} 16 & 11 & 10 & 16 & 24 & 40 & 51 & 61 \\ 12 & 12 & 14 & 19 & 26 & 58 & 60 & 55 \\ 14 & 13 & 16 & 24 & 40 & 57 & 69 & 56 \\ 14 & 17 & 22 & 29 & 51 & 87 & 80 & 62 \\ 18 & 22 & 37 & 56 & 68 & 109 & 103 & 77 \\ 24 & 35 & 55 & 64 & 81 & 104 & 113 & 92 \\ 49 & 64 & 78 & 87 & 103 & 121 & 120 & 101 \\ 72 & 92 & 95 & 98 & 112 & 100 & 103 & 99 \end{bmatrix}$$

Матрица квантования хроматических компонент (Cb и Cr):

$$\begin{bmatrix} 17 & 18 & 24 & 47 & 99 & 99 & 99 & 99 \\ 18 & 21 & 26 & 66 & 99 & 99 & 99 & 99 \\ 24 & 26 & 56 & 99 & 99 & 99 & 99 & 99 \\ 47 & 66 & 99 & 99 & 99 & 99 & 99 & 99 \\ 99 & 99 & 99 & 99 & 99 & 99 & 99 & 99 \\ 99 & 99 & 99 & 99 & 99 & 99 & 99 & 99 \\ 99 & 99 & 99 & 99 & 99 & 99 & 99 & 99 \\ 99 & 99 & 99 & 99 & 99 & 99 & 99 & 99 \end{bmatrix}$$

Избыточное квантование может привести к появлению блочных артефактов, особенно при низком битрейте. Эти артефакты возникают из-за чрезмерной потери информации в результатах квантования и часто становятся заметными на границах блоков.

Когда коэффициенты в матрице квантования слишком велики (например, при слишком сильном сжатии), значительная часть информации теряется, особенно в высокочастотных компонентах, которые отвечают за детали изображения. Этот процесс квантования приводит к значительным потерям точности в преобразованных данных. На изображении это выражается в виде артефактов, которые часто становятся видимыми в местах стыка блоков.

Однако, такого эффекта можно добитья и при слишком малом квантовании. На границах блоков, где алгоритм не может точно передать информацию, начинается видимый переход между различными уровнями яркости или цвета, что приводит к образованию четких линий или квадратных паттернов, которые визуально ощущаются как артефакты.

Так же, поскольку изображение разбивается на блоки размером 8×8 пикселей, и каждый блок сжимается независимо. Это может привести к потере контекста между блоками, особенно в случае сильного сжатия. Границы блоков не могут быть плавными, из-за чего возникают резкие переходы между блоками.

Эти артефакты могут выглядеть как "ступеньки"или резкие изменения яркости и цвета, которые не присутствуют в исходном изображении.

Как уменьшить блочные артефакты

Существует несколько подходов для минимизации блочных артефактов при сжатии JPEG:

1) Использование более низкого коэффициента сжатия: Уменьшение уровня сжатия помогает сохранить больше данных, что уменьшает квантование и, соответственно, количество потерь информации. Это снижает вероятность появления артефактов, но увеличивает размер файла.
2) Адаптивное квантование: В некоторых случаях используется адаптивное квантование, при котором для разных частей изображения выбираются различные матрицы квантования в зависимости от содержания. Например, для областей с высокой детализацией можно использовать более точное квантование, а для равномерных областей — более грубое.
3) Эксперименты с постобработкой: В некоторых случаях для устранения артефактов применяются методы постобработки, такие как сглаживание или фильтрация на уровне блоков.
4) Использование более сложных алгоритмов сжатия: Алгоритмы сжатия, такие как JPEG2000, используют более сложные методы, которые могут уменьшить или полностью устранить блочные артефакты. В частности, JPEG2000 использует вейвлет-преобразования вместо DCT, что позволяет избежать проблемы разделения изображения на блоки и улучшить качество сжатия.

JPEG не использует векторное квантование в силу его значительно большей вычислительной и алгоритмической сложности: построение и оптимизация кодовой книги требуют значительных ресурсов, а сам процесс кодирования с подбором ближайшего эталона становится существенно менее эффективным для реализации в массовых форматах хранения изображений. Кроме того, векторное квантование хуже адаптируется к переменной структуре изображений и менее совместимо с блочной природой DCT-представления.

2.7 Зигзаг-преобразование блоков

После применения DCT, и последующего квантования блоков, изображение представляется в виде матрицы коэффициентов 8×8, где каждый коэффициент представляет собой

степень различной частоты. Эти коэффициенты могут включать как высокочастотные (мелкие детали изображения), так и низкочастотные компоненты (общие очертания). Важным моментом является то, что после квантования высокочастотные коэффициенты, как правило, имеют значения, близкие к нулю, и не играют значительной роли для восприятия изображения. Поэтому их можно исключить или закодировать с меньшей точностью.

Зигзаг-преобразование используется для того, чтобы упорядочить эти коэффициенты в порядке, который позволяет более эффективно их кодировать. Суть метода в том, что он позволяет собирать все нулевые значения (или значения, близкие к нулю) в конец последовательности, что снижает объем данных при их дальнейшем кодировании.

Зигзаг-преобразование упорядочивает эти коэффициенты таким образом, что сначала идут наиболее значимые коэффициенты (низкочастотные), а затем — менее значимые (высокочастотные, которые в основном равны нулю).

Зигзагообразный порядок включает следующие шаги:

1) Начинается с первого элемента в верхнем левом углу матрицы (нулевой индекс).
2) Затем происходит движение по диагоналям матрицы, начиная с верхней левой части, последовательно переходя к правому нижнему углу.
3) Каждая диагональ матрицы представляет собой группу коэффициентов, которые последовательно вытягиваются в одном ряду.

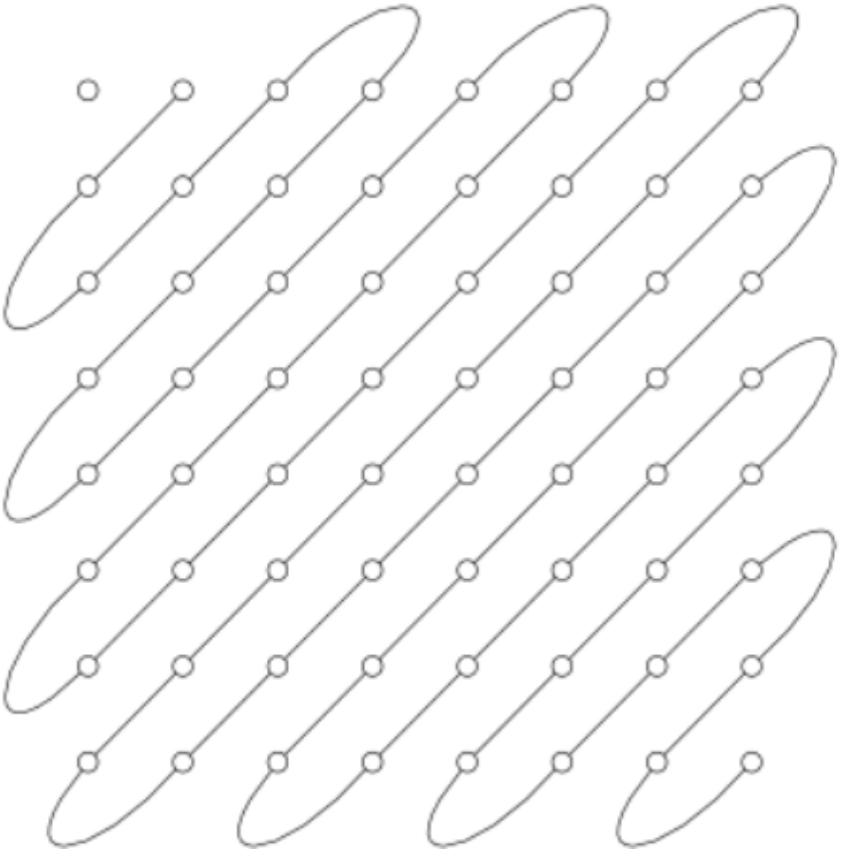


Рисунок 7. Обход блока зиг-заг.

В результате применения зигзагообразного обхода к блоку преобразованных коэффициентов дискретного косинусного преобразования формируется линейная последовательность. Один из примеров такой последовательности приведён ниже.

$$\text{Последовательность после зигзагообразного обхода: } \left\{ \begin{array}{cccccccc} 52 & -3 & 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{array} \right\} \quad (10)$$

Здесь первый элемент (52) соответствует DC-компоненте, а последующие значения — коэффициенты высоких частот (AC-компоненты), отсортированные по убыванию их вклада в общее изображение.

2.8 RLE-кодирование

После применения зигзаг-преобразования каждый блок коэффициентов преобразуется в одномерный массив, в котором значимая информация (низкочастотные коэффициенты) сосредоточена в начале, а большая часть оставшихся элементов принимает нулевое значение. Это делает полученную последовательность особенно удобной для дальнейшего сжатия с использованием методов, предназначенных для работы с повторяющимися элементами.

На этом этапе в классическом алгоритме JPEG традиционно применяется энтропийное кодирование, наиболее часто — с использованием кода Хаффмана, позволяющего достичь высокой степени сжатия за счёт построения оптимального префиксного кодового дерева, основанного на вероятностях появления различных значений. Однако реализация эффективного кодировщика Хаффмана требует значительных трудозатрат, включая построение статистической модели, генерацию кодовых деревьев и реализацию соответствующего декодирования.

В рамках данной работы энтропийное кодирование было намеренно опущено. Вместо него используется более простой и наглядный метод кодирования с повторениями — Run-Length Encoding (RLE), который позволяет проиллюстрировать фундаментальные принципы постобработки данных после зигзаг-преобразования и обеспечивает компрессию, достаточную для демонстрации эффективности базового JPEG-подобного сжатия. Это решение продиктовано необходимостью сосредоточиться на ключевых этапах сжатия и ограниченностью временных ресурсов, отведённых на реализацию дипломного проекта.

Применение RLE особенно эффективно в контексте JPEG-подобного представления данных, так как вследствие квантования и зигзагообразного упорядочивания большинство коэффициентов в конце последовательности равны нулю. Алгоритм RLE преобразует такие повторы в компактную форму, заменяя длинные последовательности одинаковых значений (в первую очередь нулей) парой «значение–длина повтора», что существенно снижает объём итоговых данных.

Таким образом, хотя в полномасштабной реализации алгоритма JPEG предпочтитель-

но использовать энтропийное кодирование, в рамках данной работы RLE выступает как более простая альтернатива, позволяющая сохранить общий принцип компрессии без существенного увеличения сложности реализации.

Таблица 5. Последовательность коэффициентов после зигзаг-преобразования

№	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
Значение	52	-3	0	0	0	2	0	0	0	0	0	0	0	0	1	0

Таблица 6. Преобразование последовательности с помощью RLE

№	**Пара (значение, количество нулей перед)**	**Пояснение**
1	(52, 0)	Первое значение, сразу записывается
2	(-3, 0)	Следующее ненулевое значение без нулей перед ним
3	(0, 3)	Три последовательных нуля
4	(2, 0)	Следующий ненулевой коэффициент
5	(0, 7)	Семь нулей подряд
6	(1, 0)	Следующий ненулевой коэффициент
7	(0, 2)	Завершающие два нуля

Применение RLE позволяет существенно сократить объём информации, подлежащей сохранению или передаче, при этом сохраняя возможность точного восстановления исходной последовательности. Хотя степень сжатия, достигаемая с помощью RLE, уступает алгоритмам энтропийного кодирования (таким как Хаффман), его простота и наглядность делают его отличным выбором для учебных и демонстрационных целей.

Таким образом, использование RLE в данной работе позволило реализовать полный цикл JPEG-подобного сжатия — от разделения изображения на блоки и применения косинус-преобразования до финального этапа кодирования, — при этом сосредоточившись на ключевых принципах без чрезмерного усложнения алгоритма.

2.9 Декодирование

Процесс декодирования JPEG является обратной операцией по отношению к этапам кодирования и направлен на восстановление изображения из сжатого представления. На выходе должен быть получен массив пикселей, максимально приближенный к исходному изображению. Процесс декодирования включает следующие этапы:

Распаковка закодированных данных

На первом этапе декодирования выполняется извлечение и интерпретация сжатых данных, полученных после применения метода RLE (run-length encoding, кодирование длин серий). В процессе кодирования многие коэффициенты DCT, особенно высокочастотные, равны нулю. Это свойство используется для повышения степени сжатия — нули заменяются на пары вида (количество нулей, ненулевое значение).

При декодировании производится восстановление полного набора коэффициентов блока размером 8×8 из такой сжатой последовательности. Например, последовательность:

$$\begin{aligned} &\text{Сжатая последовательность пар (количество нулей, значение):} \\ &\quad (0,\ 52),\ (0,\ -3),\ (3,\ 2),\ (9,\ 1) \\ &\text{После распаковки:} \\ &\quad 52,\ -3,\ 0,\ 0,\ 0,\ 2,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 1,\ 0 \end{aligned} \tag{11}$$

Обратный зигзагообразный обход

Поскольку в процессе кодирования двумерный массив коэффициентов представлялся в виде одномерного массива с помощью зигзагообразного обхода, теперь необходимо восстановить исходное расположение коэффициентов в блоке 8×8. Каждому элементу последовательности возвращается его позиция в соответствии со стандартной зигзагообразной схемой.

Обратное квантование

На этапе обратного квантования каждый коэффициент умножается на соответствующее значение из таблицы квантования, использованной при кодировании. Если обозначить квантованный коэффициент как $Q_{i,j}$, а таблицу квантования как $T_{i,j}$, то восстановленный DCT-коэффициент можно получить по формуле:

$$D_{i,j} = Q_{i,j} \cdot T_{i,j} \tag{12}$$

Это приближает коэффициенты к их исходным значениям до потери точности при квантовании.

Обратное дискретное косинусное преобразование (IDCT)

На этом этапе выполняется обратное дискретное косинусное преобразование (IDCT) для каждого блока. Восстановление значений пикселей в пространственной области осуществляется на основе 64 DCT-коэффициентов, по формуле (9).

Восстановление блоков

После применения IDCT к каждому блоку получаются блоки из 64 значений интенсивности, которые собираются в итоговое изображение. При этом важно точно соблюсти позиционирование блоков и границы изображения.

После этого идет обратное преобразование цветного пространства из **YCbCr** в **RGB**, что позволяет получить восстановленное изображение в привычном для восприятия формате.

Литература

1. Сэломон, Д. Сжатие данных, изображений и звука / Д. Сэломон. – М.: Техносфера, 2006. – 368 с.

2. Томас, Кормен. Алгоритмы Построение и анализ / Т. Кормен, Ч. Лейзерсон, Р. Ривест, К, Штайн - 2-е изд., - Москва: Вильямс, 2011. – 1296 с.

3. Официальная документация OpenCV. – Режим доступа: -URL: https://docs.opencv.org/4.x/index.html (дата обращения: 19.02.2025). – Текст: электронный.

4. Земцов, А. Н. Сравнительный анализ эффективности методов сжатия изображений на основе дискретного косинусного преобразования и фрактального кодирования (начало) / И. В. Петрова, Н. Г. Мамаев. – Волгоград: Издательство Волгоградского государственного технического университета, 2015. – 8 с.

5. Гарсия, Г. Обработка изображений с помощью OpenCV / Г. Гарсия, О. Суарес, Х. Аранда, Х. Терсеро, И. Грасиа, Н. Энано. – Москва: ДМК, 2016. – 210 с.

6. Официальная документация PyQt5. – Режим доступа: -URL: https://www.riverbankcomputing.com/ static/Docs/PyQt5/. (дата обращения: 05.01.2025). – Текст: электронный.

7. Официальная документация языка Python. – В режиме доступа: -URL: https://docs.python.org/3/ (дата обращения: 05.01.2025). – Текст: электронный.