Лабораторная работа №3
Стандарты сжатия изображений с потерей качества. Стандарт JPEG.

Широко используемым на практике подклассом систем сжатия изображений с потерей качества являются системы, основанные на преобразовании блоков из точек исходного изображения. Типичным преобразованием является разложение по некоторой системе базисных функций. Кодированию подвергается не исходное изображение а коэффициенты его разложения (коэффициенты преобразования). Часто в качестве базисных используют гармонические функции и коэффициенты разложения интерпретируют как интенсивности соответствующих гармоник. Такого рода преобразования называют переходом в частотную область.

Введем две $N \times N$ матрицы преобразования

$$T_c = \{t_c(u,i)\}, u,i = 0,1,\ldots,N-1$$

и

$$T_r = \{t_r(v,j)\}, v,j = 0,1,\ldots N-1.$$

Если двумерное изображение есть $N \times N$ матрица X, то $N \times N$ линейное преобразование для X можно записать как $Y = T_c X T_r^T$, где A^T означает транспонирование матрицы A. Матрицы T_c, T_r также называют ядрами преобразования. Строки этих матриц являются базисными функциями. Умножение X слева на матрицу T_c можно рассматривать как преобразование строк. Последующее умножение X справа на T_r можно интерпретировать как преобразование столбцов. Если преобразование симметрично, т.е. если $T_c = T_r = T$, то преобразование выполняется по формуле

$$Y = TXT^T .$$

Преобразование X в Y вводится с целью получения в Y более компактного представления данных из X. Желательно было бы найти такое преобразование, для которого выполнялись бы следующие свойства:

- Локализация большой части энергии в небольшом числе коэффициентов преобразования. Это позволило бы при кодировании с потерей качества после вычисления преобразования исключить из рассмотрения наименее информативные коэффициенты преобразования.
- Некоррелированность коэффициентов преобразования. Это свойство обеспечивает возможность кодировать каждый коэффициент независимо от других без потери потенциально достижимого уровня сжатия.
- Ортонормированность преобразования. Если это свойство выполняется, сумма ошибок аппроксимации коэффициентов преобразования в точности равна ошибке аппроксимации изображения в целом. При неортонормированном преобразовании ничтожно малая ошибка преобразования

какого-либо коэффициента, вообще говоря, могла бы привести к значительному искажению изображения.

- Низкая сложность вычисления коэффициентов преобразования. В частности, желательно использовать так называемые *сепарабельные* преобразования, для которых вычисление коэффициентов преобразования матрицы X размера $N \times N$ могут быть выполнено поэтапно. Сначала вычисляется преобразование каждой из N строк матрицы X, затем поочередно вычисляется преобразование каждого из N полученных на первом этапе векторов коэффициентов преобразования.

Наиболее эффективным в смысле некоторых из перечисленных требований является преобразования Карунена-Лоэва. Это преобразование является оптимальным в том смысле, что оно ортонормированно и гарантирует некоррелированность коэффициентов преобразования (элементов Y). Однако базисные функции этого преобразования зависят от преобразуемого изображения. Это означает, что битовый поток, описывающий сжатое изображение должен содержать не только коэффициенты преобразования, но и описание системы базисных функций. Кроме того, использование заранее известных базисных функций позволяет использовать их алгебраические свойства для уменьшения вычислительной сложности преобразования.

На практике используют дискретное преобразование Фурье (ДПФ) и дискретное косинусное преобразование (ДКП). Они привлекательны с точки зрения сложности, поскольку для них известны быстрые вычислительные алгоритмы, базисные функции этих преобразований не зависят от изображения, и, вместе с тем, например, в смысле компактности представления изображения, ДКП оказывается близким к преобразованию Карунена-Лоэва.

Кодирование, основанное на ДКП

Кодирование, основанное на ДКП, представляет собой основу всех стандартов сжатия изображений. Преобразование выполняется для блоков из $N \times N$ точек, причем в стандартах принято $N = 8$. Такой размер блоков выбран по следующим причинам. С точки зрения аппаратной и программной реализации размер блока 8×8 не накладывает существенных ограничений на размер требуемой памяти, вычислительная сложность ДКП для блока 8×8 также является приемлемой для большинства вычислительных платформ.

Выбор ДКП в качестве стандартного решения диктуется следующими причинами:

- для изображений с сильно коррелированными отсчетами (коэффициент корреляции >0.7) эффективность ДКП в смысле компактности представления данных близка к преобразованию Карунена-Лоэва,

- ДКП представляет собой ортогональное сепарабельное преобразование, независящее от изображения, поэтому его вычислительная сложность относительно невелика.

ДКП для блоков 8×8 может быть записано в виде

$$y_{kl}=\frac{c(k)c(l)}{4}\sum_{i=0}^{7}\sum_{j=0}^{7}x_{ij}\cos\left(\frac{(2i+1)k\pi}{16}\right)\cos\left(\frac{(2i+1)l\pi}{16}\right),\quad k,l=0,1,\ldots 7, \qquad (1)$$

где

$$c(k)=\begin{cases}1/\sqrt{2}, k=0\\ 1, k\neq 0.\end{cases}.$$

Обратное ДКП может быть записано как

$$x_{kl}=\sum_{i=0}^{7}\sum_{j=0}^{7}y_{ij}\frac{c(k)c(l)}{4}\cos\left(\frac{(2i+1)k\pi}{16}\right)\cos\left(\frac{(2i+1)l\pi}{16}\right),\quad i,j=0,1,\ldots 7. \qquad (2)$$

Отметим, что с вычислительной точки зрения прямое и обратное ДКП почти одинаковы, т.е. при аппаратной реализации одни и те же модули могут быть использованы при выполнении прямого и обратного преобразования.

Важным свойством двумерного прямого и обратного ДК преобразований является их сепарабельность. Одномерное ДКП может быть вычислено по формуле

$$z_k=\frac{c(k)}{2}\sum_{i=0}^{7}x_i\cos\left(\frac{(2i+1)k\pi}{16}\right),\quad k=0,...,7. \qquad (3)$$

Из (1) и (3) следует, что двумерное ДКП может быть представлено как

$$y_{kl}=\frac{c(k)}{2}\sum_{i=0}^{7}\left[\frac{c(l)}{2}\sum_{j=0}^{7}x_{ij}\cos\left(\frac{(2j+1)l\pi}{16}\right)\right]\cos\left(\frac{(2j+1)k\pi}{16}\right).$$

Пусть

$$z_{il}=\frac{c(l)}{2}\sum_{j=0}^{7}x_{ij}\cos\left(\frac{(2j+1)l\pi}{16}\right),\quad i,l=0,...7,$$

обозначает выход одномерного ДКП строк X. Таким образом, двумерное ДКП может быть получено вначале выполнением одномерного ДКП строк X со следующим за ним одномерным ДКП столбцов Z. В матричной форме преобразование

$$Y=TXT^T$$

можно теперь записать в виде последовательности преобразований:

$$Z=TX^T,\ Y=TZ^T=TXT^T.$$

Система кодирования, использующая ДКП, выполняет следующие вычисления:

1. Вычисляется ДКП для каждого блока изображения, состоящего из 8×8 точек. Результатом преобразования является блок из 8×8 коэффициентов ДКП.
2. При кодировании с потерей качества далее следует операция квантования, результатом которой является устранение наименее информативных коэффициентов ДКП.

3. Ненулевые квантованные значения y_{kl} сжимаются без потерь с использованием неравномерного кодирования.

В большинстве приложений используется кодирование длин серий со следующим за ним кодированием кодом Хаффмена.

Пример.

Рассмотрим один блок изображения

$$X = \begin{vmatrix} 168 & 161 & 161 & 150 & 154 & 168 & 164 & 154 \\ 171 & 154 & 161 & 150 & 157 & 171 & 150 & 164 \\ 171 & 168 & 147 & 164 & 164 & 161 & 143 & 154 \\ 164 & 171 & 154 & 161 & 157 & 157 & 147 & 132 \\ 161 & 161 & 157 & 154 & 143 & 161 & 154 & 132 \\ 164 & 161 & 161 & 154 & 150 & 157 & 154 & 140 \\ 161 & 168 & 157 & 154 & 161 & 140 & 140 & 132 \\ 154 & 161 & 157 & 150 & 140 & 132 & 136 & 128 \end{vmatrix}$$

После удаления среднего значения 128 из каждого элемента X и выполнения ДКП матрица Y имеет вид

$$Y = \begin{vmatrix} 214 & 49 & -3 & 20 & -10 & -1 & 1 & -6 \\ 34 & -25 & 11 & 13 & 5 & -3 & 15 & -6 \\ -6 & -4 & 8 & -9 & 3 & -3 & 5 & 10 \\ 8 & -10 & 4 & 4 & -15 & 10 & 6 & 6 \\ -12 & 5 & -1 & -2 & -15 & 9 & -5 & -1 \\ 5 & 9 & -8 & 3 & 4 & -7 & -14 & 2 \\ 2 & -2 & 3 & -1 & 1 & 3 & -3 & -4 \\ -1 & 1 & 0 & 2 & 3 & -2 & -4 & -2 \end{vmatrix}$$

Пока никакого сжатия не произошло. Отметим, что по сравнению с X в матрице Y коэффициенты с большими амплитудами сосредоточены около коэффициента y_{00}, называемого коэффициентом постоянного тока. Такая ситуация типична и именно это явление используется для сжатия изображений. Процесс квантования, который ведет к сжатию коэффициентов ДКП выражается следующим образом

$$z_{kl} = round(y_{kl} / q_{kl}) = \lfloor (y_{kl} \pm \lfloor q_{kl} / 2 \rfloor) / q_{kl} \rfloor, \quad k,l = 0,1,\ldots,7,$$

где q_{kl} обозначает kl-й элемент матрицы квантования Q размера 8×8, ($\lfloor x \rfloor$ обозначает наибольшее целое меньшее или равное x). Для нашего примера выберем матрицу Q вида

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

Тогда квантованный выход ДКП имеет вид

$$Z=\begin{vmatrix}13&4&0&1&0&0&0&0\\3&-2&1&1&0&0&0&0\\0&0&1&0&0&0&0&0\\1&-1&0&0&0&0&0&0\\-1&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0\\0&0&0&0&0&0&0&0\end{vmatrix}$$

В результате квантования произошло обнуление многих коэффициентов y_{kl}. Выбор матрицы Q определяется требуемым коэффициентом сжатия. Очевидно, что процесс квантования существенно экономит биты, необходимые для представления исходного изображения, т.к. только 11 ненулевых значений необходимы для представления Z вместо исходных 64 значений, необходимых для представления X. Таким образом, можно говорить о том, что уже получено сжатие с коэффициентом 5.8. Матрица Z может быть эффективно представлена с использованием кодирования длин серий и кодирования кодом Хаффмена.

Декодирование начинается с восстановления из полученного битового потока закодированных неравномерным кодом длин серий нулей и значащих элементов матрицы Z. Так как кодирование коэффициентов матрицы Z выполняется без потерь, то декодер восстанавливает точное значение Z. Восстановление оценок коэффициентов разложения $\widehat{Y}$ по квантованным значениям Z выполняется по формуле

$$\widehat{y}_{kl}=z_{kl}q_{kl}\,.$$

В рассматриваемом примере получим

$$\widehat{Y} = \begin{vmatrix} 208 & 44 & 0 & 16 & 0 & 0 & 0 & 0 \\ 36 & -24 & 14 & 19 & 0 & 0 & 0 & 0 \\ 0 & 0 & 16 & 0 & 0 & 0 & 0 & 0 \\ 14 & -17 & 0 & 0 & 0 & 0 & 0 & 0 \\ -18 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{vmatrix}.$$

Теперь декодер выполняет обратное 8×8 ДКП для $\widehat{Y}$. Выход преобразователя имеет вид

$$\hat{X} = \begin{vmatrix} 171 & 160 & 149 & 149 & 158 & 166 & 166 & 162 \\ 174 & 164 & 155 & 154 & 160 & 164 & 161 & 156 \\ 171 & 164 & 157 & 156 & 158 & 158 & 151 & 145 \\ 161 & 157 & 154 & 154 & 155 & 151 & 144 & 137 \\ 156 & 155 & 155 & 156 & 156 & 152 & 145 & 140 \\ 159 & 160 & 160 & 160 & 157 & 153 & 148 & 145 \\ 161 & 161 & 160 & 156 & 150 & 144 & 141 & 139 \\ 159 & 158 & 155 & 148 & 139 & 132 & 129 & 128 \end{vmatrix}$$

Очевидно, что $\hat{X} \neq X$ и причина этого несовпадения (ошибки кодирования) лежит в квантовании коэффициентов ДКП.

Стандарт JPEG (сжатие с потерей качества)

Ниже приведено краткое описание стандарта JPEG для черно-белых изображений. Для цветных изображений в формате YUV (4:1:1) один тот же процесс кодирования используется для яркостной и двух цветоразностных компонент.

Изображение делится на непересекающиеся блоки, каждый размером 8×8 точек. Если размеры изображения не кратны 8, то точки последней строки или соответственно последнего столбца повторяются. Каждый блок подвергается двумерному ДКП. Коэффициенты ДКП квантуются и кодируются неравномерным кодом. Кодирование неравномерным кодом выполняется в две ступени. На первой ступени выполняется дифференциальное кодирование для коэффициентов постоянного тока и кодирование длин серий для остальных коэффициентов. На второй ступени производится кодирование кодом Хаффмена.

Пусть DC_i и DC_{i-1} обозначают коэффициенты постоянного тока в блоках i и $i-1$. Из-за высокой корреляции значений коэффициентов постоянного тока в соседних блоках, JPEG использует для них дифференциальное кодирование. Для черно-белых изображений(или компонент Y,U или V цветного изображения) точка изображения соответствует 8 битам и разность $DC_i - DC_{i-1}$ принимает значение в диапазоне -2047 до 2047. Этот диапазон подразделяется на 12 категорий, где $i-$ я категория включает все разности, которые могут быть представлены в двоичной форме последовательностью длины i бит. Это первые 12 категорий из таблицы 1. После просмотра таблицы каждый коэффициент постоянного тока может быть описан парой *(категория, амплитуда)*. Если значение $DC_i - DC_{i-1} \geq 0$, то амплитуда это просто двоичное представление этой разности с числом бит, задаваемым категорией. Если же $DC_i - DC_{i-1} < 0$, то амплитуда кодируется обратным кодом абсолютного значения разности. Из указанной пары : *(категория, амплитуда)* только категория кодируется кодом Хаффмена.

Для 8-битовых изображений остальные коэффициенты, называемые коэффициентами переменного тока, могут принимать значения в диапазоне -1023 до 1023. Как и для коэффициентов постоянного тока этот диапазон делится на 10 категорий и каждый коэффициент может быть описан парой: *(категория, амплитуда)*. После квантования большинство коэффициентов переменного тока оказывается равными 0, так что кодировать необходимо только небольшое число ненулевых коэффициентов. Коэффициенты переменного тока обрабатываются в зигзагообразном порядке. На рис. 4 показано, каким образом производится считывание элементов матрицы размера 8×8. Этот порядок позволяет обеспечить более эффективное кодирование длин серий.

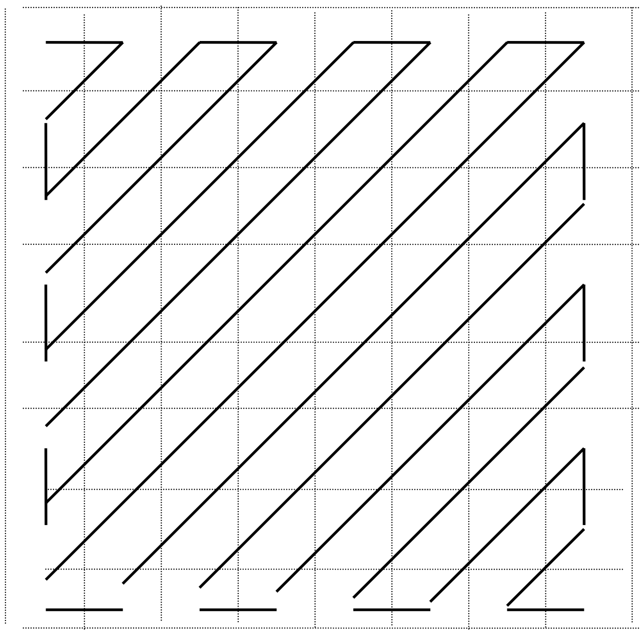

Рис.4. Зигзагообразный порядок считывания коэффициентов преобразования

Кодер длин серий выдает ненулевое значение следующего за данным ненулевого коэффициента и длину серии нулей, т.е. число нулевых коэффициентов, предшествующих этому ненулевому коэффициенту. Таким образом, каждый ненулевой коэффициент может быть записан парой: *(длина серии/категория, амплитуда)*. Значение длина серии/категория кодируется кодом Хаффмена, а значение амплитуды вычисляется как и в случае кодирования разностей коэффициентов постоянного тока и добавляется к коду. Например, пусть коэффициенту предшествует 6 нулей и он принимает значение -18. Из таблицы 1 следует, что -18 принадлежит к категории 5. Обратный код 18 равен 01101. Таким образом, коэффициент представляется парой (6/5,01101). Пара (6/5) кодируется кодом Хаффмена и 5-битовое значение для -18 добавляется к коду. Если слово кода Хаффмена для (6/5) есть 1101, то кодовое слово для -18 - это 110101101.

В кодировании коэффициентов переменного тока имеются два специальных случая.

1. После некоторого ненулевого коэффициента все остальные символы равны 0. В этом случае передается специальный символ (0/0), который означает конец блока (EOB).
2. Встретилась комбинация длина серии/категория, отсутствующая в таблице кода Хаффмена. В этом случае передается специальное кодовое слово, называемое escape-код, за которым равномерным кодом передается значение длины серии, далее следует равномерный код ненулевого значения.

Для завершения описания стандарта необходимы таблицы кодов и матрицы квантования. Эти данные мы не приводим, поскольку они заняли бы слишком много места и не существенны для понимания сути алгоритмов. Кроме того, стандарт допускает использование произвольных матриц квантования и неравномерных кодов для кодирования битовых категорий коэффициентов постоянного и переменного тока. В этом случае матрицы и кодовые таблицы должны быть переданы в составе битового потока в оговоренном стандартом формате.

ЗАДАНИЕ

1. Выполнить ДКП для компоненты Y цветного изображения. Проквантовать полученные коэффициенты. Выполнить преобразование коэффициентов постоянного тока в последовательность категорий и амплитуд.
2. Исследовать статистические характеристики полученного потока данных и оценить возможный коэффициент сжатия изображения при использовании согласованных с полученными статистическими данными кодов Хаффмена.
3. Оценить величину отношения сигнал-шум. Получить восстановленное изображение и вывести его на экран.
4. Сопоставить полученные результаты с характеристиками стандартного JPEG-кодера.
5. **Задание для самостоятельного исследования.** Построить функцию скорость-искажение для кодирования по стандарту JPEG. Выявить блоки, для которых кодирование по такому алгоритму неэффективно. Предложить для них альтернативный способ кодирования.