

Лабораторная работа №2

Эффективное кодирование

Цель работы: изучить методы оптимального кодирования Шеннона–Фано и Хаффмана

Задание на лабораторную работу:

1. Взять фотографию размером 128 x 128 с глубиной цвета 256 градаций серого.
2. Взять центральную строку пикселей $T: X [x_1, x_2, \dots, x_n]^T$ и выполнить ее квантование по формуле $X = \text{round}(X / 20) * 20$. Эта строка будет представлять собой сообщение.
3. Для каждого уникального символа вычислить его частоту появления в цифровой последовательности (строке).
4. Проанализировать первичный алфавит: количество символов алфавита, минимальную длину равномерного двоичного кода, значение энтропии.
5. Построить двоичный равномерный код.
6. Сформировать коды Шеннона–Фано для выделенных символов.
7. Сформировать коды Хаффмана для выделенных символов.
8. Оценить среднюю длину кодовой комбинации для кодов Хаффмана и Шеннона-Фано.
9. Оценить степень сжатия сообщения, закодированного равномерным кодом и кодом Шеннона-Фано, а также равномерным кодом и кодом Хаффмана.
10. Оценить относительную избыточность для сформированных кодов Хаффмана и Шеннона-Фано.

Методические указания:

Исходные данные для работы – упорядоченные по не возрастанию частоты встречаемости (вероятности появления) символы первичного алфавита.

Результат – кодовые таблицы и кодовое дерево.

Алгоритм формирования кодовых комбинаций по методу Шеннона-Фано:

1. Алфавит разбивается на два подмножества с примерно одинаковыми суммарными вероятностями. Верхнее подмножество получает префикс 1, нижнее – 0.
2. Полученные подмножества рекурсивно делятся и получают соответствующие двоичные цифры в префиксном коде до тех пор, пока в каждом подмножестве не остается по одному элементу.

Алгоритм формирования кодовых комбинаций по методу Хаффмана:

1. Символы входного алфавита образуют список свободных узлов. Каждый лист имеет вес, который может быть равен либо вероятности, либо количеству вхождений символа в сжимаемое сообщение.
2. Выбираются два свободных узла дерева с наименьшими весами.
3. Создается их родитель с весом, равным их суммарному весу.
4. Родитель добавляется в список свободных узлов, а двое его детей удаляются из этого списка.
5. Одной дуге, выходящей из родителя, ставится в соответствие бит 1, другой — бит 0.
6. Шаги, начиная со второго, повторяются до тех пор, пока в списке свободных узлов не останется только один свободный узел. Он и будет считаться корнем дерева.

Анализ алфавита:

Энтропия независимых случайных событий x с n возможными состояниями (от 1 до n):

$$H(x) = - \sum_{i=1}^n p_i \log_2 p_i$$

Число двоичных разрядов кода (длина двоичного кода) определяется по формуле Хартли:

$i = \log_2 N$, где N - количество символов в используемом алфавите (мощность алфавита), i - количество информации в одном символе в битах.

Относительная избыточность кода определяется формулой (Q):

$Q = 1 - \frac{H}{I}$, где H - энтропия сообщения, I - средняя информация на знак, определенная с учетом вероятностей их появления в сообщении.

Степень сжатия принято определять как коэффициент сжатия, равный:

$k = \frac{S_{\text{код}}}{S_{\text{исх}}}$, где $S_{\text{код}}$ - размер закодированного (сжатого) сообщения; $S_{\text{исх}}$ - размер исходного сообщения.

Содержание отчета:

1. Титульный лист с названием лабораторной работы, ФИО студента и группы.
2. Исходное изображение и фотография в заданном формате.
3. Полученная цифровая последовательность (сообщение) :
 - a. Если выделение цифровой последовательности выполнялось путем написания программы, то приводится листинг кода с подробными комментариями.
 - b. Если выделение цифровой последовательности выполнялось путем использования готового программного обеспечения, то приводятся скриншоты на каждый этап работы.
4. Упорядоченные символы первичного алфавита с указанием частоты встречаемости (вероятности появления). Количество символов алфавита. Значение энтропии. Расчетная длина двоичного кода.
5. Коды Шеннона–Фано.
6. Кодовое дерево Хаффмана.
7. Коды Хаффмана для символов.
8. Закодированное равномерным кодом сообщение, длина кодового слова, количество переданной информации при передаче сообщения.
9. Закодированная полученным кодом Шеннона–Фано последовательность, средняя длина кодового слова, количество переданной информации при передаче сообщения.
10. Закодированная полученным кодом Хаффмана последовательность, средняя длина кодового слова, количество переданной информации при передаче сообщения.
11. Расчеты по п. 9 – 10 задания на лабораторную работу.
12. Выводы по работе – оценка степени сжатия при разных методах кодирования.