

ПЕРЕМЕННОЕ КВАНТОВАНИЕ С МИНИМАЛЬНЫМИ ИСКАЖЕНИЯМИ

ВВЕДЕНИЕ

При решении различных задач науки и техники человеку приходится передавать, хранить или обрабатывать огромное количество визуальной информации, являющейся для него по-видимому психологически наиболее привычной. Действительно, человек стремится визуализировать даже сигналы в области невидимого спектра (радиолокация, инфракрасное тепловидение, рентгенография, гидролокация, ультразвуковое зондирование, акустическая голография и т.д.).

Как только объемы визуальной информации начинают превышать возможности технических систем передачи, хранения или обработки, возникает необходимость в ее экономном представлении т.е. в *сжатии*.

Основной операцией, отвечающей за качество декодированного изображения является *квантование*, разбивающее пространство сигналов на ячейки. Все сигналы, попавшие в одну и ту же ячейку считаются неотличимыми друг от друга. Зрительная система человека также разбивает пространство сигналов на ячейки сложной формы и не видит различий в сигналах из одной и той же ячейки. Это подтверждается огромным количеством психофизических опытов [7-10]. Каким должно быть квантование, чтобы при максимально возможном сжатии качество декодированного изображения оставалось неотличимым от оригинала для получателя в строго определенных условиях наблюдения? Формально эта задача рассматривается в теории ϵ -энтропии однако, несмотря на обилие работ, здесь, к сожалению до сих пор не найдено конструктивное решение [1-6]. С точки зрения источника сигналов квантователь должен отслеживать изменения статистических свойств изображения, а с позиций получателя должен согласовывать форму своих ячеек с формой и размером ячеек зрительной системы.

На данный момент накоплено огромное количество знаний о методах квантования изображений, которые отражены в прекрасных обзорах и монографиях [11-15]. Анализируя многочисленную литературу по квантованию изображений, можно заметить, что еще в ранних работах стали использовать переменные квантователи [16 -20], несмотря на то, что теории для таких квантователей не было разработано. В этих работах было обнаружено, что при фиксированном квантовании очень быстро достигается предел сжатия, за которым начинается ухудшение качества декодированного изображения. Можно несколько улучшить результат применяя методы статистического кодирования [4], однако другой путь - переход к *переменному квантованию*. оказался более эффективным. Действительно, с помощью простых измерений локальных свойств сигналов изображения можно убедиться в их существенной изменчивости в пределах изображения. Легко убедиться также и в локальной изменчивости зрительных порогов получателя - известное свойство маскирования - увеличение зрительных порогов вблизи перепадов яркости в пространстве и во времени [13].

В конце семидесятых годов одновременно в Японии [21], США [22] и СССР [23,24] были предложены варианты переменного блочного квантования изображений, которые затем в восьмидесятых были детально изучены в отечественных [24-35] и зарубежных работах [36,37]. В настоящее время можно сказать, что переменное квантование доминирует практически во всех алгоритмах сжатия изображений.

Несмотря на столь блистательную победу переменного квантования, потенциальные возможности многих подходов остались неизученными. Из-за отсутствия подходящих моделей переменного квантователя, построение "хорошего" квантователя до сих пор осуществляется эвристически на основе опыта и искусства разработчика. Теоретически неизвестны оптимальные процедуры изменения параметров квантователя, приводящие к максимальному сжатию при заданных требованиях к качеству декодированного изображения - задача оптимизации скорости, и минимальному искажению при фиксированной скорости - задача оптимизации искажений. В данной работе делается попытка решения второй задачи. Формализуется задача *переменного*

квантования изображений и известные алгоритмы оптимизации квантователей обобщаются на случай переменного квантования на основе многослойной модели.

1. НЕКОТОРЫЕ ОПРЕДЕЛЕНИЯ

Пусть F_0 - размеры поля зрения, на котором задано исходное изображение. В дискретном виде это изображение может быть представлено в виде некоторого массива чисел $U = \{u(i,j), i = 1, I; j = 1, J\}$, где I - число строк, а J - число столбцов. Массив U можно трактовать как последовательность скалярных величин u - элементов изображения, если ввести определенное правило их выборки из массива (развертку).

В одномерном случае квантование представляет собой отображение числовой оси \mathfrak{R}_1 на множество представительных уровней $r = \{r_l, l = 1, L\}$, т.е. $Q: \mathfrak{R}_1 \Rightarrow r$ и называется скалярным квантованием, а в многомерном случае векторным квантованием.

Квантование осуществляется в два этапа. На первом элементы числовой оси \mathfrak{R}_1 отображаются на множество индексов $J = \{1, 2, \dots, L\}$, т.е. $C: \mathfrak{R}_1 \Rightarrow J$. На втором этапе множество индексов J отображается на множество представительных уровней r , т.е. $D: J \Rightarrow r$.

При отображении C числовая ось \mathfrak{R}_1 разбивается на отрезки $\tau = \{\tau_l, l = 1, L\}$, $\cup \tau_l = \mathfrak{R}_1$ и $\tau_l \cap \tau_k = \emptyset$ для $l \neq k$ и каждое число u из \mathfrak{R}_1 , попавшее в отрезок τ_l заменяется на соответствующий индекс l из множества J , причем границы отрезков τ определяются набором пороговых уровней $t = \{t_l, l = 0, L\}$.

Для осуществления отображения D необходимо иметь множество представительных уровней r и последовательность индексов l из множества J . Каждому индексу l ставится в соответствие представительный уровень r_l .

Таким образом, скалярный квантователь q полностью определяется множеством пороговых t и представительных уровней r , т.е. $q = \{t, r\}$, а операция квантования представляется в виде $u \in \tau_l : Q(u) = r_l$. Если используется равномерный квантователь, то достаточно только трех чисел для его полного определения. Здесь возможны следующие варианты: $\{r_1, r_L, L\}$, $\{D, H, L\}$, или $\{r_1, S, L\}$, где r_1 и r_L первый и последний представительный уровни, $D = r_L - r_1$ - динамический диапазон, $H = (r_1 + r_L)/2$ или $H = r_0$ - средняя (опорная) точка и $S = D/(L-1)$ - шаг квантования, L - число уровней квантования. Для равномерного квантователя, таким образом, мы будем иметь $q = \{r_1, r_L\}$, $q = \{r_1, S\}$ или $q = \{D, H\}$, $i = 1, I, j = 1, J$.

Введем теперь формальное определение *переменного квантователя*. Переменным квантователем будем называть квантователь, параметры которого зависят от координаты элемента изображения в поле зрения получателя F_0 . Если задано правило выборки элементов из массива U , то можно записать для скалярного случая $q(i) = \{t(i), r(i)\}$, где $i = 1, 2, \dots$ - номера элементов изображения. Для двумерных координат в поле зрения F_0 будем иметь $q(i,j) = \{t(i,j), r(i,j)\}$, где $i,j \in F_0$ - координаты элемента в поле зрения получателя.

Для того, чтобы определить параметры квантователя необходимо произвести некоторые операции с изображениями. Множество элементов или векторов, на основе которых определяются параметры квантователя, будем называть *полем определения* квантователя и обозначать через $F_d(i,j)$. Множество элементов или векторов, подвергающихся одинаковому квантованию будем называть *полем зрения* квантователя и обозначать через $F_v(i,j)$. Очевидно, что $F_v(k,l) \cap F_v(m,n) = \emptyset$, $k \neq m$ или $l \neq n$. На практике эти параметры фиксируются т.е. $F_v(i,j) = F_v$ и $F_d(i,j) = F_d$.

2. ПРОЦЕДУРЫ ИЗМЕНЕНИЯ ПАРАМЕТРОВ КВАНТОВАТЕЛЯ.

Как было отмечено выше, для успешного декодирования (операция D) на приемную сторону необходимо передать представительные уровни r и последовательность индексов l . При фиксированном квантовании ($F_v = \infty$) представительные уровни квантователя передаются один раз в начале связи, а затем достаточно передавать только индексы.

Для случая переменного квантования при каждом изменении квантователя приходится об этом информировать приемник. Здесь становится актуальной проблема выбора процедуры изменения параметров квантователя. Формально параметры

квантователя могут переопределяться в каждой точке изображения ($F_v = 1$), однако, очевидно, тогда система сжатия будет неэффективной, поскольку придется передавать большой объем дополнительной информации. Существует два классических приема, позволяющих уменьшить объем дополнительной информации. Первый - *предсказание* [35], при котором значения параметров квантователя в текущей точке пытаются угадать по уже переданным на приемную сторону параметрам квантователей из некоторой окрестности квантуемой точки

$$q(i) = P(q(i-m), q(i-m-1), \dots, q(i-1)), \quad (1)$$

где m - определяет размер окрестности, а P - оператор предсказания.

Второй - *интерполяция*, при которой в редких точках происходит переопределение параметров квантователя на основе пикселей из его поля определения, а в остальных их детерминированное изменение, с помощью интерполяционных функций [30,41].

$$q(j) = A(q(i-m), q(i+m)), \quad j = i-m+1, \dots, i+m-1, \quad (2)$$

где A - оператор интерполяции.

И в том и в другом случае делаются предположения о медленном изменении параметров квантователя от точки к точке. В ряде случаев, например, если среднее сигнала практически не меняется, можно зафиксировать какую-нибудь опорную точку квантователя и относительно нее изменять другие параметры. В другом случае можно перестраивать только опорную точку, а остальные параметры зафиксировать и т.д. В общем случае, например для равномерного квантователя можно варьировать такими параметрами, как D диапазоном, S - шагом квантования, H - опорной точкой, L - числом уровней квантования и шагом интерполяции Z . Поскольку параметры D и S зависимы, реально используется один из этих двух параметров, например шаг S . Если изменяются все параметры, мы имеем один 4 - параметрический переменный квантователь $q(S, H, L, Z)$, возможны также 4 трехпараметрических $q(S, H, L)$, $q(Z, H, L)$, $q(S, Z, L)$, $q(S, H, Z)$, 6 двухпараметрических $q(S, H)$, $q(S, L)$, $q(S, Z)$, $q(H, L)$, $q(H, Z)$, $q(L, Z)$ и 4 однопараметрических $q(S)$, $q(H)$, $q(L)$ и $q(Z)$ квантователей. Мы будем рассматривать только 2-х параметрические переменные интерполяционные квантователи типа $q(S, H)$ с фиксированными параметрами (L, Z) .

3.МНОГОСЛОЙНАЯ МОДЕЛЬ ПЕРЕМЕННОГО КВАНТОВАТЕЛЯ.

Реальные изображения имеют ярко выраженную структуру. Как правило это композиция участков с приблизительно одинаковыми свойствами, причем переход от одного участка к другому может быть как резким, так и плавным. Если предположить, что нарушение гладкости в изображении редко вызваны разрывами физической поверхности изображаемого объекта, а скорее взаимным затенением и заслонением объектов, то разумно принять многослойную модель, которая трактует разрывы в функции как переключение между двумя гладкими функциями. Эти гладкие функции - слои описывают поведение представительных и пороговых уровней переменного квантования. Таким образом, основная идея здесь заключается в аппроксимации сложной функции набором гладких, переключаемых с помощью некоторой L - значной функции, где L - число слоев в модели или число уровней в квантователе.

Если принять вероятностную трактовку, то многослойная модель совпадает с иерархической вероятностной моделью ансамбля изображений [38-40], при которой предполагается, что источник изображений S является сложным и состоящим из множества подисточников $S = \{S_l, l = 1, L\}$, где L - число подисточников. Согласно [40] сначала выбирается реализация случайного L - значного поля индексов $L = \{L(i, j) : (i, j) \in F_0\}$. После формирования символического поля L сложное изображение U разобьется на участки $U_l, l = 1, L$, т.е. $U = \cup U_l$. Сложное изображение формируется согласно правилу $u(i, j) = u_l(i, j)$, если $L(i, j) = l$, где $u_l(i, j)$ - элемент реализации l - го подисточника S_l . Таким образом сложное изображение полностью описывается набором $\{L, S_1, \dots, S_L\}$. Эти

параметры, как правило, заранее неизвестны. Для их определения часто используют итеративную процедуру, а в качестве критерия оптимизации - максимум апостериорной вероятности (МАВ оценивание), предварительно выбрав значение L и свойства поисточников S_j . Например, в [40] предлагается использовать частные случаи источника с Гиббсовским распределением. Для индексов - полуторомарковское символнозначное случайное поле, а в качестве S_j - гауссмарковские поля.

На первом этапе предполагают равную вероятность пикселей символьного поля. Затем от итерации к итерации вероятности переоцениваются и приближение сложного изображения простыми случайными полями становится все лучше. Если итеративный процесс сойдется, то мы получим оптимальную МАВ оценку символьного поля L для выбранного значения L и заданного набора подисточников S . Значение L должно быть достаточным, для того, чтобы охватить все возможное разнообразие однородных участков изображения. Очевидно, что увеличивая число подисточников L и усложняя их свойства, мы будем приближать свойства сложных реализаций U к свойствам реальных изображений.

Рассмотрим детерминированную трактовку и пусть сложные изображения являются кусочно-непрерывными функциями и имеется L гладких функций $R = \{ R_l, l = 1, L \}$. Предположим что, мы определили символнозначное поле L , тогда сложное изображение U будет композицией кусков функций R_l , т.е. $U = \cup R_l$, причем $R_l \cap R_k = \emptyset, l \neq k$, а его элементы будут определяться согласно правилу $u(i,j) = R_l(i,j)$, если $L(i,j) = l$. При выбранном значении L и функциях R_l задача оптимизации здесь заключается в поиске наилучшего поля индексов L , при котором достигается минимум выбранного критерия ошибок аппроксимации. По аналогии с вероятностным подходом для этого можно воспользоваться алгоритмом, при котором итеративно происходит переопределение поля индексов.

3.1 Выбор функций для многослойной модели

Как выбирать гладкие функции R_l ? Если при вероятностной трактовке гладкость определялась структурой используемого потенциала в Гиббсовском распределении, то здесь можно воспользоваться другой классификацией функций, например, по количеству непрерывных производных, как это принято в теории аппроксимации [41]. Не теряя общности для простоты рассмотрим одномерный случай. Обычно пространство функций, заданных на конечном отрезке числовой оси $[a,b]$ и имеющих n непрерывных производных обозначается через $C_n[a,b]$. Внутри каждого класса функции, имеющие n производных различаются по модулю непрерывности n - той производной. Модуль непрерывности некоторой функции R при определенном значении g обозначается через $w(R;g)$ и определяется выражением

$$w(R;g) = \max\{|R(x) - R(y)|: x, y \in [a, b], |x - y| < g, \quad (3)$$

Для непрерывных функций $w(R, g)$ может меняться в широких пределах. Классическим средством аппроксимации являются многочлены. Пусть R есть функция, которую необходимо приблизить многочленом A , принадлежащим к классу многочленов порядка не выше n , т.е. $A \in A_n$. Предположим, что нам удалось выбрать наилучший из возможных многочленов $A^* \in A_n$ т.е.

$$\|R - A^*\| = \text{dist}(R, A_n) = \min_{A \in A_n} \|R - A\| \quad (4)$$

Известна оценка для (4), выраженная через (3) [39]. Если $R \in C[a,b]$, т.е. непрерывна на отрезке $[a,b]$, то

$$\text{dist}(R, A_n) \leq 6 \cdot w(R, (b-a)/(2 \cdot (n-1))), \quad (5)$$

Легко видеть из (5), что если функция на отрезке имеет какую-нибудь особенность, например, быстрое изменение на каком-нибудь интервале, которое увеличивает значение w (.,.), аппроксимация функции на всем отрезке может оказаться плохой. Зависимость от локальных свойств аппроксимируемой функции - главный недостаток многочленной аппроксимации. Таким образом, если потребуются более сложные R , то многочлены могут стать непригодными для этой цели. Единственным способом уменьшения погрешности аппроксимации является уменьшение отношения $(b - a)/(n - 1)$. Это можно сделать увеличением n и/или уменьшением $(b - a)$. Заметим, что как разбиение отрезка $[a, b]$ на k частей, так и использование многочленов порядка $k \cdot n$ одинаковым образом влияют на ограничение (5). Однако, вычисление значения многочлена порядка $k \cdot n$ требует выполнения операций с $k \cdot n$ коэффициентами и базисными функциями, сложность которых возрастает с увеличением $k \cdot n$. Кроме того, приходится решать систему линейных уравнений n - го порядка. В то же время вычисление значения в некоторой точке кусочно-многочленной функции порядка n с k частями требует операций с n коэффициентами и локальной базисной функцией ограниченной сложности. Нахождение коэффициентов кусочно-многочленной функции требует решения, как правило, ленточной системы уравнений, что представляет собой более простую задачу. По этой причине обычно эффективнее уменьшать $(b - a)$, чем увеличивать n . Это утверждение служит обоснованием выбора в дальнейшем кусочно-многочленной аппроксимации для гладких функций R . Полный класс кусочных многочленов называется сплайнами. Удобным представлением кусочно-многочленной функции является разложение по, так называемым, базисным сплайнам (B - spline) [41].

4. ПЕРЕМЕННОЕ КВАНТОВАНИЕ С ПРОИЗВОЛЬНЫМ ОПЕРАТОРОМ ИНТЕРПОЛЯЦИИ.

Обозначим через $V = I \cdot J$ - число пикселей в изображении U . Предположим далее, что известно множество параметров квантователей, заданных в редких точках $Q = \{ q(m, n), m = 1, M, n = 1, N \}$, где $q(m, n) = \{ t_0(m, n), \dots, t_L(m, n); r_1(m, n), \dots, r_L(m, n) \}$ - квантователь в точке (m, n) , L - число уровней квантования, а $K = M \cdot N$ - число редких точек на изображении.

Зададим множество интерполирующих функций $A = \{ A_{mn}, m = 1, M, n = 1, N \}$, с помощью которых по множеству квантователей, заданных в редких точках восстанавливают множество квантователей в каждой точке исходного изображения $Q = \{ q^I(i, j), i = 1, I, j = 1, J \}$ согласно выражению

$$q^I(i, j) = \sum_{m=1}^M \sum_{n=1}^N A_{mn}(i, j) \cdot q(m, n), \quad (6)$$

Выражение (6) по сути дела означает, что по массивам представительных и пороговых уровней, заданных в редких точках $R = \{ r_l(m, n), m = 1, M, n = 1, N, l = 1, L \}$, $T = \{ t_l(m, n), m = 1, M, n = 1, N, l = 0, L \}$ восстанавливают массивы представительных и пороговых уровней в каждой точке исходного изображения $R = \{ r_l(i, j), i = 1, I, j = 1, J, l = 1, L \}$ и $T = \{ t_l(m, n), m = 1, M, n = 1, N, l = 0, L \}$, согласно выражениям

$$r_l(i, j) = \sum_{m=1}^M \sum_{n=1}^N A_{mn}(i, j) \cdot r_l(m, n), \quad (7)$$

$$t_l(i, j) = \sum_{m=1}^M \sum_{n=1}^N A_{mn}(i, j) \cdot t_l(m, n), \quad (8)$$

где везде $A_{mn}(i,j)$ - значение интерполирующей функции A в точке (i,j) .

Обозначая множество координат пикселей изображения, удовлетворяющих условию $t_{l-1}^r(i,j) < u(i,j) \leq t_l^r(i,j)$ при фиксированном l через F_l , можно записать правило переменного квантования $u_r(i,j) = r_l^r(i,j)$, если $(i,j) \in F_l$ где $U_r = \{u_r(i,j), i = 1, I, j = 1, J\}$ - массив восстановленных пикселей изображения. Функционал ошибки квантования можно записать в виде

$$E_F(Q^r) = \sum_{i=1}^I \sum_{j=1}^J [(u(i,j) - u_r(i,j))^2] = \sum_{l=1}^L \sum_{(i,j) \in F_l} [(u(i,j) - r_l^r(i,j))^2] = \sum_{l=1}^L E_{F_l}(R_l^r), \quad (9)$$

а задачу переменного квантования свести к минимизации функционала (9) по всем возможным массивам пороговых и представительных уровней т.е.

$$E = \min_{T_l^r, R_l^r} E_F(Q^r), \quad (10)$$

Иными словами мы будем решать задачу оптимизации (потерь) искажений при фиксированных параметрах модели (L, Z) переменного квантователя (т.е. фиксируется скорость передачи информации). Некоторые варианты решения задачи оптимизации скорости при фиксированных потерях можно найти в [24,27].

5. КОНСТРУКТИВНЫЕ АЛГОРИТМЫ ПЕРЕМЕННОГО КВАНТОВАНИЯ.

5.1. Оптимальное СКО переменное квантование

Решение задачи (10) требует полного перебора всевозможных массивов пороговых и представительных уровней. Эта процедура даже в одномерном случае является трудоемкой, а для указанного выше многомерного случая практически неосуществима. Необходим конструктивный алгоритм решения этой задачи. Воспользуемся идеями алгоритма Ллойда-Макса 1 [41] и применим его к нашей задаче. Для этого надо доказать, что условия оптимальности одного шага квантования в этом алгоритме соблюдаются и в случае переменного квантования. Для упрощения перейдем к векторному обозначению массивов U, U^r, R, T, R^r, T^r .

Введем произвольный оператор интерполирования A . Тогда согласно условиям (7) и (8) векторы R_l, R_l^r, T_l, T_l^r будут связаны уравнениями

$$R_l^r = A \cdot R_l, \quad l = 1, L \quad (11)$$

$$T_l^r = A \cdot T_l, \quad l = 0, L \quad (12)$$

Используя для множества номеров компонент вектора U , удовлетворяющих условию $t_{l-1}^r(i) < u(i) \leq t_l^r(i)$, такое же обозначение, как и для элементов соответствующих массивов - F_l , запишем выражение для функционала ошибки квантования в интервале $[T_{l-1}, T_l]$ в следующем виде

$$E_{F_1}(R_1) = \sum_{i \in F_1} [u(i) - r_1^r(i)]^2 = \|U_1 - S_1 \cdot A \cdot R_1\|^2, \quad (13)$$

где $U_1 = S_1 \cdot U$, $S_1 = \text{diag}(s_1^{11}, s_2^{22}, \dots, s_1^{VV})$,

$$s_1^{ii} = \begin{cases} 1, i \in F_1 \\ 0, \text{otherwise} \end{cases}$$

Основным требованием к значениям компонентов векторов представительных уровней является выполнение условия $t_0^r(i) < r_1^r(i) < t_1^r(i) < \dots < t_{L-1}^r(i) < r_L^r(i) < t_L^r(i)$. Это требование приводит к следующему условию оптимальности вектора представительных уровней R_1 при заданных пороговых векторах T_{1-1} и T_1 .

$$E_1^* = \min_{t_{1-1}^r(i) < r_1^r(i) \leq t_1^r(i)} E_{F_1}(R_1), \quad (14)$$

Задача (14) известна в теории, как задача на условный экстремум квадратичного функционала и решается методами квадратичного программирования. Решив L раз эту задачу для всех значений l , мы получим совокупность векторов представительных уровней R_l , $l = 1, L$, удовлетворяющих условию оптимальности (14), по которым согласно (11) определим представительные векторы R_l^r .

Теперь необходимо записать условие оптимальности порогового вектора T_l , при фиксированных векторах представительных уровней в следующем виде

$$E_{F_1}(R_1) + E_{F_{l+1}}(R_{l+1}) = \min_{t_{1-1}^r(i) < t_1^r(i) \leq t_{l+1}^r(i)} [E_{F_1}(R_1) + E_{F_{l+1}}(R_{l+1})], \quad (15)$$

Аналогично, согласно (12) получим искомый вектор T_1^r . Поскольку ошибка квантования (15) является суммой ошибок квантования каждой компоненты исходного вектора, выражение достигает минимума, если мы будем для каждой компоненты вектора принимать оптимальное решение. Если значение компоненты $u(i)$ расположено между двумя значениями компонентов $t_{1-1}^r(i)$ и $t_{l+1}^r(i)$ при заданных значениях $r_1^r(i)$ и $r_{l+1}^r(i)$, наилучшим решением в смысле (15) будет расположение значения компоненты $t_1^r(i)$ посередине между значениями компонент векторов представительных уровней, т.е. $t_1^r(i) = (r_1^r(i) + r_{l+1}^r(i))/2$. Если оператор A линеен, то учитывая представления (11) и (12), можно записать $A \cdot T = (A \cdot R_1 + A \cdot R_{l+1})/2$ откуда следует $A \cdot T = A \cdot (R_1 + R_{l+1})/2 \Rightarrow T_1^r = (R_1 + R_{l+1})/2$. Это означает, что это условие достаточно соблюдается в редких точках, а в промежуточных точках оно выполнится автоматически.

Окончательно условия оптимальности одного шага итерационной процедуры построения оптимального переменного квантователя можно записать

$$R_1^{*r} = O[T_{1-1}^r, T_1^r], \quad l = 1, L, \quad (16)$$

$$T_1^{*r} = (R_1^r + R_{l+1}^r)/2, \quad l = 1, L-1, \quad T_0^r < \dots < T_L^r. \quad (17)$$

где O - условное обозначение процедуры решения задачи (14).

Применяя последовательно эти процедуры, можно построить последовательность векторов квантователей, начиная с некоторого приближения $Q^0, Q^1, \dots, Q^k, \dots$, для которых будет удовлетворяться условие

$$E_F(P(Q^{k-1})) = E_F(Q^k) < E_F(Q^{k-1}), \quad (18)$$

где k - номер итерации, а P - условное обозначение процедур (16) и (17). Сходимость этой процедуры обусловлена (18) и ограниченностью значения ошибки квантования снизу нулем. Процедура сходится к стационарному вектору квантователей Q^* . Найденный оптимум является локальным. Нахождение глобального оптимума требует применения процедуры динамического программирования, обобщение которой на случай переменного квантования с произвольным оператором интерполяции не представляется возможным. Мы показали пример разработки алгоритма переменного квантования на основе СКО критерия. Сложность алгоритма СКО обуславливается непредсказуемостью положения уровней квантователя. Можно упростить алгоритм квантования введя структурные ограничения, например, древовидные.

5.2 Переменное квантование с древовидной структурой.

Согласно алгоритму на первом уровне дерева необходимо найти оптимальный вектор квантователей с одним уровнем Q . Пусть заданы квантователи в редких точках, по которым затем определяются квантователи в промежуточных точках $Q^r = A \cdot Q$. Иначе, полагая $Q = \{T_0, T_1, R_1\}$ и $Q^r = \{T_0^r, T_1^r, R_1^r\}$, можно записать $R_1^r = A \cdot R_1$. Вектор R_1 следует искать из условия минимума квадратичного функционала

$$E_V(R_1) = \|U - R_1^r\|^2 = \|U - A \cdot R_1\|^2, \quad (19)$$

при ограничениях типа неравенств т.е.

$$E_V^*(R_1) = \min_{T_0 \leq R_1 \leq T_1} E_V(R_1), \quad (20)$$

где $t_0(i) = u_{\min}$, а $t_1(i) = u_{\max}$, $u_{\min} = \text{MIN}(u(1), \dots, u(V))$, $u_{\max} = \text{MAX}(u(1), \dots, u(V))$. Воспользуемся более простой процедурой безусловной минимизации. Приравняв частные производные функционала по R_1 нулю, получим систему линейных уравнений $A^T \cdot U = A^T \cdot A \cdot R_1$, откуда оптимальное решение будет иметь вид

$$R_1^* = (A^T \cdot A)^{-1} \cdot A^T \cdot U, \quad (21)$$

Используя (21) получим для квантователей в промежуточных точках

$$R_1^{*r} = A \cdot (A^T \cdot A)^{-1} \cdot A^T \cdot U, \quad (22)$$

Удвоим число уровней квантования и образуем вектор разностей

$$\Delta = U - R_1^{*r} = U - A \cdot (A^T \cdot A)^{-1} \cdot A^T \cdot U, \quad (23)$$

Представим вектор Δ в виде $\Delta = \Delta^+ - \Delta^-$, где $\Delta^+ = S \cdot \Delta$, $\Delta^- = (E - S) \cdot \Delta$, а S - соответствующая матрица, получаемая из выражения

$$s_{ii} = \begin{cases} 1, \Delta(i) \geq 0 \\ 0, \Delta(i) < 0 \end{cases}, \quad i = 1, V \quad (24)$$

где s_{ii} - i -й диагональный элемент матрицы.

Далее необходимо задать векторы разностей D^+ и D^- в редких точках, по которым затем получим

$$\Delta^{r+} = A \cdot D^+, \quad \Delta^{r-} = A \cdot D^-, \quad (25)$$

Делая несложные преобразования, получим системы уравнений

$$A^T \cdot \Delta^+ = A^T \cdot S \cdot A \cdot D^+, \quad (26)$$

$$A^T \cdot \Delta^- = A^T \cdot (E - S) \cdot A \cdot D^-, \quad (28)$$

откуда, решая их, получим выражения для оптимальных векторов D^{*+} и D^{*-}

$$D^{*+} = (A^T \cdot S \cdot A)^{-1} \cdot A^T \cdot \Delta^+ = (A^T \cdot S \cdot A)^{-1} \cdot D, \quad (29)$$

$$D^{*-} = (A^T \cdot (E - S) \cdot A)^{-1} \cdot A^T \cdot \Delta^- = (A^T \cdot (E - S) \cdot A)^{-1} \cdot D, \quad (30)$$

где $D = A^T \cdot \Delta^+ = -A^T \cdot \Delta^-$.

Подставляя оптимальные векторы (29) и (30) в (25), а затем, умножая результат на соответствующие матрицы S и $(E - S)$ получим

$$\Delta_s^{*r+} = S \cdot A \cdot (A^T \cdot S \cdot A)^{-1} \cdot A^T \cdot \Delta^+, \quad (31)$$

$$\Delta_s^{*r-} = (E - S) \cdot A \cdot (A^T \cdot (E - S) \cdot A)^{-1} \cdot A^T \cdot \Delta^-, \quad (32)$$

Декодирование изображения на втором уровне дерева производится простым суммированием Δ_s^{*r+} и Δ_s^{*r-}

$$U^r = R_1^{*r} + \Delta_s^{*r+} + \Delta_s^{*r-}, \quad (33)$$

Если необходим вектор квантователей с большим числом уровней, процедура (33) не производится, число уровней удваивается, векторы вычитаются из исходных Δ^+ и Δ^- и полученные векторы разности, каждый в отдельности подвергается описанной выше процедуре.

5.3 Переменное равномерное квантование

Равномерный квантователь является простейшим устройством и требует минимального объема информации для его описания. Обобщим его на случай переменного квантования с произвольным оператором интерполяции.

Пусть заданы векторы T_0^r и T_L^r , определяющие множество минимальных и максимальных значений пороговых уровней квантования, заданных в редких точках. Тогда можно определить векторы $T_0^r + V^r$, $T_L^r - V^r$ и V^r , где $V^r = (T_L^r - T_0^r)/L$. Обозначив через F_1 и F_L множество индексов компонентов вектора U , для которых удовлетворяются условия $t_0^r(i) \leq u(i) < t_0^r(i) + b^r(i)$ и $t_L^r(i) \leq u(i) < t_L^r(i) + b^r(i)$, где $b^r(i)$ - i -я компонента вектора V^r , можно найти два вектора представительных уровней R_1^r и R_L^r из условия минимума квадратичных функционалов

$$E_{F_1}^* = \min_{t_0^r(i) < r_1^r(i) \leq t_0^r(i) + b^r(i)} E_{F_1}(R_1), \quad (34)$$

$$E_{F_L}^* = \min_{t_L^r(i) - b^r(i) < r_L^r(i) \leq t_L^r(i)} E_{F_L}(R_L), \quad (35)$$

где $E_{F_m}(R_m) = \|U_m - R_m^r\|^2 = \|U_m - S_m \cdot A \cdot R_m\|^2$,

$$U_m = S_m \cdot U, \quad S_m = \text{diag}(s_{m11}, s_{m22}, \dots, s_{mVV}), \quad m = 1, m = L,$$

$$s_{0ii} = \begin{cases} 1, & t_0^r(i) \leq u(i) < t_0^r(i) - b^r(i) \\ 0, & \text{otherwise} \end{cases} \quad s_{Lii} = \begin{cases} 1, & t_L^r(i) - b^r(i) \leq u(i) < t_L^r(i) \\ 0, & \text{otherwise} \end{cases}$$

Для простоты используем безусловную минимизацию. Решая системы уравнений

$$A^T \cdot U_m = A^T \cdot S_m \cdot A \cdot R_m, \quad m = 1, m = L, \quad (36)$$

получаем выражения для векторов представительных уровней

$$R_m = (A^T \cdot S_m \cdot A)^{-1} \cdot A^T \cdot U_m, \quad m = 1, m = L, \quad (37)$$

По полученным векторам можно определить остальные векторы уровней из выражения

$$R_k = R_1 + (k-1) \cdot \Delta, \quad k = 1, L \quad \Delta = (R_L - R_1) / (L-1), \quad (38)$$

Пороговые уровни можно расставить оптимальным образом

$$T_k = (R_k + R_{k+1}) / 2, \quad k = 1, L-1 \quad (39)$$

В промежуточных точках квантователи определяются с помощью интерполяции согласно (11) и (12).

Везде, где использовался метод безусловной минимизации, необходимо полученные решения проверять на выполнение условия $t_0^r(i) < r_1^r(i) < t_1^r(i) < \dots < t_{L-1}^r(i) < r_L^r(i) < t_L^r(i)$, $i = 1, K$ и если оно не удовлетворяется ограничивать их. Например, для произвольного вектора представительных уровней R_k , полученного в результате минимизации, необходимо произвести преобразование

$$R_k^* = \begin{cases} T_{k-1}, & R_k < T_{k-1}, \\ R_k, & T_{k-1} \leq R_k < T_k, \\ T_k, & R_k \geq T_k, \end{cases} \quad (40)$$

а затем использовать полученный вектор R_k^* вместо R_k .

6. БЫСТРЫЕ АЛГОРИТМЫ ПЕРЕМЕННОГО КВАНТОВАНИЯ.

Наложим ограничения на оператор интерполяции A . Пусть A есть кусочный многочлен нулевого порядка. Тогда квантователи, определенные в редких точках будут

Из (46) и (47) видно, что значения параметров квантователя в промежуточных точках (i,j) определяются только значениями параметров в точках (m,n) и не зависят от квантователей в других точках, что является следствием ортогональности оператора A . В более общем случае ортогональность не соблюдается и значение параметров квантователя в точке (i,j) может зависеть от того, какие значения примут параметры квантователей в остальных точках.

Выполнение условия (45) позволяет нам переписать условия оптимальности (16) и (17) в следующем виде

$$r_1(m, n) = O[t_{1-1}(m, n), t_1(m, n)], \quad m = 1, M, n = 1, N, \quad (48)$$

$$t_1(m, n) = (r_1(m, n) + r_{1+1}(m, n)) / 2, \quad (49)$$

где O - процедура оптимизации представительного уровня при заданных пороговых.

Легко видеть, что сложная задача квадратичного программирования свелась к K - кратному решению более простой скалярной задачи Ллойда-Макса 1.

Покажем, что при этом общий алгоритм древовидного переменного квантования с оператором интерполяции нулевого порядка сводится к известному быстрому алгоритму группового кодирования [23]. Действительно, из структуры матрицы A^T видно, что при условии ортогональности ее действие на вектор изображения сводится к простому суммированию значений элементов вектора, а действие матрицы $(A^T \cdot A)^{-1}$ к их нормированию. Выражение (21) будет иметь вид

$$t_1 = \left(\sum_i \sum_j U(i, j) / (K_m \cdot K_n) \right), \quad (50)$$

что означает вычисление арифметического среднего. Аналогично выражение (37) после очевидных преобразований будет иметь вид

$$r_1 = \left(\sum_{(i, j) \in F_1} U(i, j) \right) / N_1, \quad 1 = 1, 2. \quad (51)$$

где $F_1 = \{(i, j): U(i, j) \in [t_0, t_1]\}$,

$F_2 = \{(i, j): U(i, j) \in [t_1, t_2]\}$,

N_1 - число элементов множества F_1 .

Легко видеть, что в случае равномерного переменного квантования выражение, определяющее представительные уровни равномерного квантователя сводится к простому условному среднему элементов изображения.

В заключении введем формальное определение переменного квантователя, использующего кусочный многочлен нулевого порядка, который для краткости будем называть *блочной импульсно кодовой модуляцией* (БИКМ).

Пусть исходный массив U разбивается на непересекающиеся подмножества, которые в дальнейшем будем называть блоками $U = \{U_{mn}, m = 1, M, n = 1, N\}$.

Блок размером $F_d(m, n) = F_v(m, n) = Z = K_m \cdot K_n$ имеет вид

$$U_{mn} = \{u(m \cdot K_m + i, n \cdot K_n + j), \quad i = 1, K_m, j = 1, K_n\}, \quad (52)$$

Каждый блок рассматривается как отдельное изображение, для которого определяется свой квантователь $q(m, n)$ по одному из известных в литературе методов квантования. Таким образом для квантования всего массива пикселей U необходимо сформировать

множество $M \cdot N = K$ квантователей, или эквивалентно, K наборов представительных и пороговых уровней

$$Q = \{q(m, n) = (t_0(m, n), \dots, t_L(m, n); r_1(m, n), \dots, r_L(m, n)), \\ m = 1, M, \quad n = 1, N. \quad (53)$$

Заключение

На основе многослойной модели интерполяционного переменного квантования сформулирована задача оптимизации потерь в сжатии при фиксированной скорости передачи информации. Приводятся общие ее решения для СКО оптимального, древовидного и равномерного переменного квантователей при произвольном операторе интерполяции. Показано, что известные алгоритмы быстрого блочного квантования (групповое [23], блочное усеченное кодирование [22] и т.д.) являются частными случаями общих алгоритмов при операторе нулевого порядка. Приведенные алгоритмы могут быть использованы в любом методе сжатия не только в области значений элементов изображения, но и в области коэффициентов таких преобразований, как Фурье, дискретное косинусное, Лапласа и др.

В дальнейшем предполагается распространить полученные результаты на переменные квантователи с переменными L и Z и векторные квантователи.

ЛИТЕРАТУРА

1. Шеннон К. Математическая теория связи. Работы по теории информации и кибернетике.- М. - ИЛ. 1963 - С.243 - 332.
2. Berger T. Rate-Distortion Theory.-Englewood Cliffs. - N.J.: Prentice-Hall.- 1971.
3. Колмогоров А.Н., Тихомиров В.Н. ϵ - энтропия и ϵ - емкость множеств в метрических пространствах, Успехи математических наук. - 1959 14. 3.
4. Kieffer J.C. A Survey of the Theory of Source Coding with a Fidelity Criterion, IEEE Transactions on Inform. Theory, Vol.39, No.5. September 1993.
5. Tasto M. and Wintz P. A bound on the rate-distortion function and application to images, IEEE Trans. on Inf. Theory - 1972. v.IT - 18. - PP.150 - 159.
6. Anastassiou D. and Sakrison D. New Bounds to $R(D)$ for Additive Sources and Applications to Images Encoding, IEEE Trans. on Inf. Theory - 1979 - v.IT - 25. - no.2. - PP.145 - 154.
7. Sakrison D.J. On the role of the observer and a distortion measure in image transmission, IEEE Trans. Comm. - 1977. v.COM - 25. - PP. 1251 - 1267.
8. Granrath D.J. The Role of Human Visual Models in Image Processing, Proceedings of the IEEE. 1981 - v.69.-no.5.- PP. 552 - 561.
9. Hall C.F. and Hall E.L. A nonlinear model for spacial characteristics of the human visual system, IEEE Trans. Syst. Man Cybern. - 1977. - v.SMC - 7. PP. 161 - 170.
10. Mannos J.L. and Sakrison D.J. Effects of a Visual Fidelity Criterion on the Encoding of Images, IEEE Trans. on Inform. Theory - 1974. v.IT - 20. - no.4.- PP.525 - 535.
11. Kronander T. Perception based image coding. Proceedings, of SSAB, Gothenburg, Sweden, March. - 1989.
12. Джайн А.К. Сжатие видеинформации. Обзор ТИИЭР - 1981. т.69. вып.3. - С.71 - 117.
13. Нетравали А., Лимб Дж. Кодирование изображений Обзор. ТИИЭР. - 1980. т.68. вып.3. С.76 - 121.
14. Прэтт У. Цифровая обработка изображений. Пер с англ.-М Мир. 1982 . Т.1 312 с. и Т.2 480 с.
15. Jayant N. Signal Compression: Technology Targets and Research Directions, IEEE Journal on Selected Areas in Communications, Vol.10., No. 5, June, 1992.
16. Goodall W.M. Television by Pulse Code Modulation, Bell Syst. Techn. J. 30. 33 - 49. (January, 1951).
17. А.с. 120534 Способ анализа и синтеза телевизионного изображения/Ю.М.Брауде-Золотарев. Заявл. 22.11.1957, Оpubл. 1959. Бюл. 12.

18. Cutler C.C. Differential quantization on communication signals, U.S. Patent 2605361. 1952.
19. Kretzmer E.R. Reduced - alphabet representation of television signals, Conv. Rec. IRE - 1958. no.6.- P.58.
20. Schreiber W.F., Knapp C.F. Bandwidth reduction by digital coding, Conv. Rec. IRE - 1958. no.6 - P.58.
21. Kishimoto T., Mitsua E., Hoshida K.A. Method of still Picture Coding by using Statistics Properties (in Japanese), National Conference of the Inst. of Electronics and Communications of Japan. 1974.-no.974.
22. Delp E.J., Mitchel O.R. Image compression using Block Truncation Coding, IEEE Trans. Communs. - 1979. - v.COM - 27. - PP. 1335 - 1342.
23. Сардыко С.В., Цуккерман И.И. Групповое кодирование изображений. Техника кино и телевидения.-1977.- N9. С.52-54.
24. Тхор В.Б. Адаптивное квантование изображений, Обработка данных в информационных системах. Часть. 1. - М - 1986. - С.40-43.
25. Тхор В.Б. Сравнение методов трехуровневого блочного квантования изображений, Тезисы докладов. Всесоюзный симпозиум "Сокращение избыточности в цифровых телевизионных системах". -Тбилиси, 1983. С.86-87.
26. Тхор В.Б., Гогоберидзе Л.И. Трехуровневое групповое кодирование изображений, Цифровое кодирование изображений и коммутация изображений, Труды ГПИ им. В.И. Ленина. -1985.- N10(292). С. 55-58
27. Тхор В.Б. Кодирование гаусс-марковских полей методом блочной импульсно-кодовой модуляции. Обработка данных в информационных системах. Часть 11.-М - 1986. С.28-31.
28. Тхор В.Б., Гогоберидзе Л.И. Блочная импульсно-кодовая модуляция и свойства источника, Цифровое кодирование и коммутация сигналов изображений, Труды ГПИ им.В.И.Ленина.- 1987. N613 (8).- С.75-78.
29. Лебедев Д.С., Тхор В.Б. Интерполяционный подход к кодированию изображений методом блочной ИКМ, Тезисы докладов. 11 Всесоюзная конференция АСОИЗ - 86, 1986.- С.91-93.
30. Лебедев Д.С., Тхор В.Б. Интерполяционный подход к блочной импульсно-кодовой модуляции, Иконика. Кодирование и обработка изображений.-М. Наука.-1988.-С.5-21.
31. Тхор В.Б. Кодирование изображений с переменными параметрами квантования в системах обработки видеoinформации. Диссертация на соискание ученой степени кандидата технических наук.- Москва. 1988.
32. А.с. 1359914. Способ кодирования и декодирования телевизионного сигнала/Тхор В.Б. Заяв. 8.07.1985 Оpubл. 1987. Бюл. 46.
33. Тхор В.Б. Аппроксимационный подход к адаптивному квантованию изображений. Системы передачи и обработки информации. Часть 1.- Москва. - 1988.
34. Цифровое кодирование изображений, Под ред.Цуккермана И.И. - М. Радио и связь. - 1981.- 239 с.
35. Пилипчук Н.И., Яковлев В.П. Адаптивная импульсно-кодовая модуляция, Москва. Радио и связь.- 1986. - 296 с.
36. Yiyan W. and Coll D., Single Bit-Map Block Truncation Coding of Color Images, IEEE Journal on Selected Areas in Communications, Vol.10. no.5. 1992.
37. Monet P. and Dubois E., Block Adaptive Quantization of Images, IEEE Transactions on Communications, Vol.41., No.2., Feb. 1993.
38. Лебедев Д.С., Безрук А.А., Новиков В.М. Марковская модель изображения и рисунка, Препринт ИППИ АН СССР. М.- 1983. 40 с.
39. Лебедев Д.С. Статистическая модель изображения, Иконика. Пространственная фильтрация изображений. Фотографические системы М. Наука, 1970. С. 53-65.
40. Лебедев Д.С. Статистическая теория обработки видеoinформации. Учебное пособие. Московский Физикотехнический институт. Москва. 1988.
41. Де Бор К. Практическое руководство по сплайнам. - М. Радио и связь. 1985. - 303 с.
42. Max J. Quantizing for minimum distortion, IRE Trans. on Information Theory. - 1960.- V.IT-6. PP. 7-12.