Вісник Харківського національного університету
Серія «Математичне моделювання. Інформаційні технології. Автоматизовані системи управління»
УДК 681.325 № 605, 2003, с. 43-47

# Метод вычисления ошибки сжатия данных в системе Хаара

В. Г. Иванов

*Национальная юридическая академия Украины им. Я. Мудрого*

Outcomes of an evaluation of an error of data compression with usage of orthogonal conversions of Haar are reduced. It is shown, that the rejection of a conversion efficiency with the last serial number influences at restoring input datas only the last pair of references. Analytical expressions of computing expenditures of deriving of an error of compression are obtained and graphs of associations of these expenditures, as functions of an aspect ratio are reduced.

В задачах сжатия данных с использованием ортогональных преобразований необходимо оценить ошибку восстановления исходной информации при усечении обобщенного ряда, то есть отбрасывания коэффициентов с большими порядковыми номерами в сторону их уменьшения.

Известен метод оценки сходимости усеченного ряда Хаара [1, 2], который заключается в том, что вычисляют коэффициенты Хаара по формуле $C_k = \int_0^T f(x)\aleph_k(x)dx$, где $\aleph_k(x)$ - функции Хаара, восстанавливают исходную информацию $S_n(x)$ по этим коэффициентам, приравнивая некоторые из них нулю $S_n(x) = \sum_{k=1}^{n} C_k \aleph_k(x)$ и находят сумму квадратов разности между исходными и восстановленными данными

$$\varepsilon = \frac{1}{T}\int_T \left[ f(x) - \sum_{k=1}^{n} C_K \aleph_K(x) \right]^2 dx. \qquad (1)$$

Однако, этот метод требует больших вычислительных затрат, т.к. необходимо произвести само ортогональное преобразование, т.е. анализ в дискретном базисе Хаара, восстановить исходную информацию, т.е. провести синтез и наконец, вычислить ошибку ε по формуле (1).

Исследуя детально природу дискретных функций Хаара, можно установить, что отбрасывание коэффициента преобразования с последним порядковым номером влияет при восстановлении исходных данных только на последнюю пару отсчетов. Это вытекает из свойства локализации рядов Фурье-Хаара [1], заключающегося в том, что если изменить значение функции $f(x)$ на каком либо отрезке $[x_1,x_2]$, то это не отразится на сходимости ряда $S_n(x)$ в точке $x_0$, не принадлежащей $[x_1,x_2]$.

Это свойство используется в методе [3], заключающегося в том, что если задан вектор исходных отсчетов размерности N, то отбрасывают условный последний коэффициент ряда $S_n(x)$ с номером $N$, не вычисляя его, и последнюю пару исходных отсчетов заменяют их среднеарифметическим значением "$m$". Вычисляют ошибку $\varepsilon$, как сумму квадратов разности между этой парой отсчетов и полученным средним значением. Отбрасывают следующий ($N$-1) условный коэффициент ряда и предпоследнюю пару отсчетов заменяют их среднеарифметическим, вычисляют ошибку по определенному выше правилу и суммируют ее с ошибкой, полученной на первом шаге. Процедуру обнуления коэффициентов продолжают до получения заданной ошибки $\varepsilon$, причем порядковые номера заменяемых отсчетов их среднеарифметическим значением "$m$" соответствуют правилу образования коэффициентов Хаара [4].

Если размерность вектора $N$ исходных данных равна 16, то ошибку сжатия можно записать в виде:

$$\varepsilon = (x_{16} - m)^2 + (x_{15} - m)^2, \tag{2}$$

где $x_{16}$ и $x_{15}$ значения исходных отсчетов, а $m$ – их среднее арифметическое.

После несложных преобразований (2) можно привести к виду:

$$\varepsilon = x_{16}^2 + x_{15}^2 - \frac{1}{2}(x_{16} + x_{15})^2, \tag{3}$$

Объем вычислений при этом составит 3 операции сложения и 3 операции умножения. Считаем, что по времени выполнения операции суммирования и вычитания одинаковые, а временем выполнения операции деления на число, кратное степени двойки можно пренебречь.

В общем случае среднеквадратическая ошибка преобразования при отбрасывании коэффициентов с большими порядковыми номерами в сторону их уменьшения может быть записана в виде:

$$\varepsilon_k = \sum_K \left\{ x_{2k-N}^2 + x_{2K-(N+1)}^2 - \frac{1}{2}\left[x_{2k-N} + x_{2k-(N+1)}\right]^2 \right\}, \tag{4}$$

где $k=N, N\text{-}1, \ldots \frac{N}{2}+1$, или

$$\varepsilon_k = \sum_K \left\{ x_{4k-N}^2 + x_{4K-(N+1)}^2 + x_{4k-(N+2)}^2 + x_{4k-(N+3)}^2 - \frac{1}{4}\left[x_{4k-N} + x_{4k-(N+1)} + x_{4k-(N+3)}\right]^2 \right\}, \tag{5}$$

где $k=\frac{N}{2}, \frac{N}{2}-1, \ldots, \frac{N}{4}+1$.

Выражения (4) и (5) позволяют получить значение ошибки при сжатии информации в два и четыре раза, а число необходимых при этом вычислительных операций типа сложения-вычитания можно оценить выражением вида:

$$M = \sum_{k=2}^{N} \frac{N}{K}(2^{2+\log_2 K} + 1), \tag{6}$$

где $k$ - коэфициент сжатия, характеризующий число ненулевых коэффициентов ряда Хаара, причем $k$=2,4,…$N$.

В дискретном виде значение ошибки по выражению (1) будет выглядеть

$$\varepsilon^* = \sum_{i=1}^{N} \left[ x_i - \sum_{k=1}^{N} C_K \aleph_K \right]^2, \tag{7}$$

где $x_i$ – значение исходных данных, а $C_k$ и $\aleph_k$ соответственно коэффициенты и функции Хаара.

Выполняя анализ и синтез в базисе Хаара с использованием быстрых алгоритмов [4], число необходимых вычислительных операций для выражения (7) можно записать как:

$$M^* = 4N сл.\text{-}4сл.\text{+}2умн.\text{-}4умн.\text{+}Nсл.\text{+}Nумн. \qquad (8)$$

Заменяя также, как и в (6) одну операцию умножения, как две операции сложения и приравнивая операции суммирования и вычитания, выражение (8) приводится к виду:

$$M^* = 11N\,сл. \qquad (9)$$

Используя при вычислении (7) прямые методы, т.е. не быстрые алгоритмы, получим оценку числа операций типа сложения-вычитания:

$$Q = 3Nсл.^2\text{-}2Nсл. \qquad (10)$$

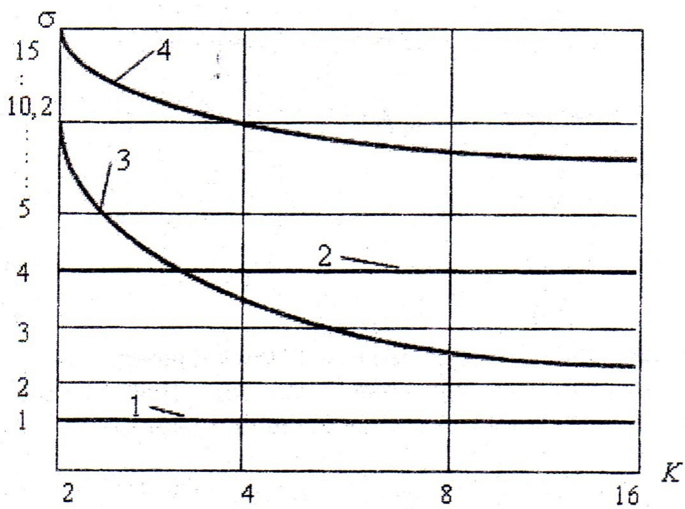


*Рис. 1. Графики зависимостей вычислительных затрат при получении ошибки сжатия*

На рис. 1. приведены графики зависимости вычислительных затрат при получении ошибки сжатия, как функции коэффициента сжатия (выражения 6, 9 и 10). Причем график 1 построен по (10), а графики 2 и 3 получены в относительных единицах как $\sigma = \dfrac{Q}{M^*}$ и $\sigma = \dfrac{Q}{M}$ при $N$=16.

Из анализа приведенных кривых следует, что при коэффициенте сжатия равным двум, вычислительная эффективность σ метода по формуле (6) кривая 3, достаточно высока. Она в десять раз превосходит методы прямых вычислений (формула (10), кривая 1) и в два с половиной раза выше метода оценки ошибки с использованием быстрых вычислений (формула (9), кривая 2). С увеличением коэффициента сжатия эта эффективность падает, оставаясь однако, всегда выше метода прямых вычислений. Следует также заметить, что на практике системы сжатия данных на базе функций Хаара не используют коэффициеты сжатия больше, чем 4-5 раз, так как в противном случае ошибка преобразования становится слишком большой (20 и более процентов).

Покажем далее в нашей работе, что объем вычислений в методах по формулам (4) и (5) может быть уменьшен.

Как и ранее, для простоты рассуждений зададимся размерностью вектора исходных данных $N$ равным шестнадцати и образуем промежуточные суммы следующего вида :

$$C_{16}=X_{16}+X_{15};\ C_{15}=X_{14}+X_{13};\ C_{14}=X_{12}+X_{11};\ C_{13}=X_{10}+X_9;$$

$$C_{12}=X_8+X_7;\ C_{11}=X_6+X_5;\ C_{10}=X_4+X_3;\ C_9=X_2+X_1; \text{ и}$$

$$A_1=C_{16}+C_{15};\ A_2=C_{14}+C_{13};\ A_3=C_{12}+C_{11};\ A_4=C_{10}+C_9;$$

$$A_5=A_1+\mathrm{A}_2;\ A_6=A_3+A_4; \tag{11}$$

Для общего случая формулы будут выглядеть:

$$C_K=X_{2K-N}+X_{2K}, \text{ где } k=N,\ N\text{-}1,\ \ldots \frac{N}{2}-1. \tag{12}$$

$$A_i=C_{N-i}+C_{N-i+1}, \text{ где } i=1,\ 2,\ \ldots,\frac{N}{4}\ . \tag{13}$$

$$A_m=A_i+A_{i+1}, \text{ где } i=1,\ 3,\ \ldots,\frac{N}{4}-1\ , \tag{14}$$

а $m=\ \dfrac{N}{4}+1,\ \ldots,\ \left[\dfrac{N}{4}+\log(\dfrac{N}{4})\right].$

Причем индексы "$m$" и "$i$" в (14) изменяются одновременно.

Тогда формулы определения ошибки при сжатии информации соответственно в два, четыре и восемь раз принимают вид:

$$\varepsilon_K=\frac{1}{2}C_k^2-2X_{2K-N}\bullet X_{2K-(N+1)}, \text{ где } k=N,\ N\text{-}1,\ \ldots,\frac{N}{2}+1. \tag{15}$$

$$\varepsilon_K=\varepsilon_{2K}+\varepsilon_{2K-1}+2A_i, \text{ где } k=\frac{N}{2},\ \ldots\frac{N}{4}+1. \tag{16}$$

$$\varepsilon_K=\varepsilon_{2K}+\varepsilon_{2K-1}+6A_i, \text{ где } k=\frac{N}{4},\ \ldots\frac{N}{4}-1, \text{ а } i=1,\ 2,\ \ldots,\frac{N}{4}. \tag{17}$$

Общий объем вычислений по выражениям (15-17) составит:

$$M_б^*=\sum_{K=2}\frac{N}{K}(\log_2 N+K), \text{ где } K=2,\ 4,\ 8,\ \ldots,\ N. \tag{18}$$

Кривая 4 на рис. 1. соответствует полученному выражению (18) и построена в относительных единицах по отношению к методу с использованием не быстрых алгоритмов. Как видно из графиков, эффективность $\sigma$ предложенного метода (формула 18) с увеличением коэффициента сжатия уменьшается, оставаясь однако, всегда лучше (как минимум в 2 раза) известных и рассмотренных выше способов (кривые 1, 2 и 3).

Таким образом, предложенный быстрый метод оценки сходимости коэффициентов Хаара в задачах сжатия данных позволяет существенно сократить число необходимых вычислительных операций для нахождении ошибки при условном отбрасывании коэффициентов ряда, минуя этапы анализа и синтеза в этом базисе. Высвободившееся процессорное время устройств сжатия данных можно использовать для решения других задач информационного содержания, что в целом будет способствовать повышению эффективности работы систем цифровой обработки сигналов и позволит им приблизиться к режиму реального времени.

## ЛИТЕРАТУРА


1. Соболь И.М. Многомерные квадратурные формулы и функции Хаара. – М.: Наука, 1970. – 288 с.
2. Трахтман А.М. Введение в обобщенную спектральную теорию сигналов. – М.: Сов. радио, 1972. – 352 с.
3. Соболев Ю.В., Поляков П.Ф., Иванов В.Г. Метод оценки сходимости коэффициентов Хаара в задачах сжатия сообщений. // Изв. вузов СССР. "Радиоэлектроника", 1984. Т.XXVII, № 6056, с. 103-104.
4. Соболев Ю.В., Поляков П.Ф., Иванов В.Г. Эффективность анализа и синтеза в дискретном базисе Хаара. // Изв. вузов СССР. "Радиоэлектроника", 1983. Т.XXVI, № 6059, с. 54-56.