

М. А. Иванов

ПРИМЕНЕНИЕ ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЙ В КОДИРОВАНИИ ИЗОБРАЖЕНИЙ*

ВВЕДЕНИЕ

Первое упоминание о вейвлетах появилось в литературе по цифровой обработке и анализу сейсмических сигналов в работах А. Гроссмана и Ж. Морлета. Однако набор базисных функций был избыточным, так как основной интерес авторов был направлен на анализ сигналов. Далее, математик И. Мейер показал существование вейвлетов, образующих ортонормальный базис в $L^2(R)$. Дискретизация вейвлет-преобразования была описана в статье И. Добеши, которая объединила разработки математиков и специалистов в области обработки сигналов. Добеши разработала семейство вейвлет-преобразований, имеющих максимальную гладкость для данной длины фильтра. Популярность вейвлетов увеличилась после введения С. Маллатом концепции кратномасштабного анализа. Он же, по-видимому, первым применил вейвлеты для кодирования изображений. И И. Добеши, и С. Маллат показали, что практическое выполнение вейвлет-преобразований осуществляется посредством двухполосного банка фильтров анализа-синтеза известного ранее в теории субполосного кодирования. Эта теория может быть описана в терминах вейвлетов. Главное различие между этими двумя направлениями заключается в критериях построения фильтров.

Некоторые идеи теории вейвлетов частично были разработаны уже очень давно. Например, А. Хаар опубликовал в 1910 году полную ортонормальную систему базисных функций с локальной областью определения. Эти функции называются теперь вейвлетами Хаара.

В настоящее время исследования в области вейвлетов ведутся по многим направлениям. В частности, разработана лифтинговая схема выполнения вейвлет-преобразований, имеющая ряд преимуществ по сравнению с традиционной. Активно исследуется целочисленное вейвлет-преобразование.

Цель настоящей статьи — изложить основы теории вейвлет-преобразований и дать обзор основных способов применения вейвлет-преобразований к задаче кодирования изображений.

* Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (грант № 01-01-794) и Министерства образования РФ.

Несмотря на то, что теория вейвлет-преобразований уже в основном разработана, точного определения, что же такое «вейвлет», какие функции можно назвать вейвлетами, насколько известно, не существует. Обычно под вейвлетами понимаются функции, сдвиги и растяжения которых образуют базис многих важных пространств, в том числе и $L^2(R)$. Эти функции являются компактными как во временной, так и в частотной области. Вейвлеты непосредственно связаны с кратномасштабным анализом сигналов. Вейвлеты могут быть ортогональными, полуортогональными, биортогональными. Эти функции могут быть симметричными, асимметричными и несимметричными. Различают вейвлеты: с компактной областью определения и не имеющие таковой. Некоторые функции имеют аналитическое выражение, другие — быстрый алгоритм вычисления связанного с ними вейвлет-преобразования. Вейвлеты различаются также степенью гладкости. Для практики желательно было бы иметь ортогональные симметричные или асимметричные вейвлеты. К сожалению, доказана теорема о том, что такими вейвлетами являются лишь вейвлеты Хаара. Функции Хаара не обладают достаточной гладкостью и не подходят для большинства приложений, поэтому для кодирования изображений обычно используют биортогональные вейвлеты.

Может показаться, что вейвлеты не являются чем-то фундаментально новым. В самом деле, сходные идеи появлялись на протяжении последних десятилетий: субполосное кодирование, успешно применяемое при кодировании речи, пирамидальные схемы кодирования изображений, преобразование и функции Габора (вейвлеты Габора). С развитием теории вейвлетов произошло как бы объединение и взаимообогащение этих идей, что привело к качественно новому результату.

ТЕОРИЯ ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЙ

Одним из основных средств обработки сигналов является линейное преобразование. Традиционно, кодеры, основанные на линейном преобразовании, делят на кодеры с преобразованием и субполосные.

Кодирование с линейным преобразованием основывается на применении ортогонального линейного преобразования. Классическим примером такого преобразования является дискретное преобразование Фурье (ДПФ), которое разлагает сигнал на синусоидальные компоненты. Другими примерами являются дискретное косинусное преобразование (ДКП) и преобразование Корунена—Лоэва (ПКЛ). Эти преобразования находятся путем вы-

числения свертки сигнала конечной длины с семейством базисных функций. В результате получается ряд коэффициентов, которые подвергаются дальнейшей обработке. Обычно эти преобразования применяются к непрерывным блокам исходного сигнала, что приводит к дополнительным искажениям, хотя есть модификации алгоритмов, ликвидирующие данный недостаток.

Субполосное кодирование реализуется посредством свертки исходного сигнала с несколькими полосовыми фильтрами и последующей децимации результата. Совокупность набора фильтров с дециматорами называется банком фильтров. Каждый получившийся в результате сигнал несет в себе информацию об определенной спектральной составляющей исходного сигнала при некотором пространственно-временном масштабе. Для обратного синтеза сигнала выполняются операции интерполяции субполосных сигналов, их фильтрация и сложение. Большинство методов синтеза фильтров направлено на устранение эффекта наложения спектров (aliasing), возникающего при децимации. В пространственной области этот эффект проявляется в виде дискретной структуры восстановленного изображения. Идеальный банк фильтров должен состоять из фильтров с прямоугольной характеристикой, предотвращающих вместе с тем aliasing. Такие фильтры, однако, приводят к так называемому эффекту Гиббса, вследствие которого искажения сильно заметны.

Сформулируем некоторые важные требования, предъявляемые к преобразованиям, используемым при кодировании изображений.

1. Масштаб и ориентация

В изображениях имеются объекты различных размеров. Поэтому, преобразование должно позволять анализировать изображение одновременно и независимо на различных масштабах, иначе говоря, допускать кратномасштабный анализ.

Для двумерного сигнала некоторая спектральная область соответствует определенному масштабу и ориентации. Ориентация базисных функций определяет способность преобразования корректно анализировать ориентированные структуры, типичные для изображений (линии, контуры). Таким образом, для решения задачи анализа желательно иметь преобразование, которое делило бы входной сигнал на локальные частотные области.

2. Пространственная локализация

Необходимость в пространственной локализации преобразования возникает тогда, когда информация о местоположении деталей изображения является важнейшей. Эта локальность, однако, не должна быть абсолютной,

блочной, как при ДКП, так как это ведет к потере свойства локальности в частотной области. В 1946 году Д. Габор предложил класс линейных преобразований, которые обеспечивают локальность и в частотной, и во временной областях. Вейвлеты являются еще одним примером функций, локализованных в пространственной и частотной областях.

3. Ортогональность

Вообще говоря, преобразование не обязательно должно быть ортогональным. Ортогональность функций лишь упрощает вычисления.

4. Быстрые алгоритмы вычисления

Это наиболее важное свойство, предъявляемое к линейным преобразованиям, применяемым для кодирования изображений, так как невозможность применения в реальном времени сводит на нет все их положительные свойства.

Непрерывное вейвлет-преобразование

Основным средством анализа стационарных непрерывных сигналов является преобразование Фурье непрерывного времени (CTFT):

$$F(\omega) = \int_{-\infty}^{\infty} f(x)e^{-i\omega x} dx ,$$

где $f(x)$ — исходный сигнал, а $F(\omega)$ — его преобразование Фурье.

С практической точки зрения, CTFT имеет ряд недостатков. Во-первых, для получения преобразования по одной частоте требуется информация о сигнале в любой момент времени. Во-вторых, на практике не все сигналы стационарны. Пик в сигнале во временной области распространится на все частоты. Для преодоления этих недостатков вводится оконное преобразование Фурье (STFT):

$$STFT_f(\omega, b) = \int_{-\infty}^{\infty} f(x)e^{-i\omega x} w(x-b) dx ,$$

где $w(x-b)$ — локальная функция окна, которая сдвигается вдоль временной оси. Преобразование становится зависимым от времени, и в результате получается пространственно-временное описание сигнала. Недостаток STFT в том, что в нем используется фиксированное окно, которое невозможно адаптировать к локальным особенностям сигнала.

Вейвлет-преобразование, рассмотренное ниже, решает эту проблему. Непрерывное вейвлет-преобразование (CTWT) есть скалярное произведение исходной функции и базисных функций:

$$\psi_{a,b}(x) = \frac{1}{\sqrt{a}} \psi\left(\frac{x-b}{a}\right), \quad a \in R^+, b \in R,$$

так что

$$CTWT_f(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} \psi\left(\frac{x-b}{a}\right) f(x) dx.$$

Базисные функции $\psi_{a,b} \in L^2(R)$ определены на некотором интервале. Они и называются вейвлетами. Их можно рассматривать как масштабирование и сдвиг некоторой функции-прототипа $\psi(x)$. Параметр b отвечает за расположение во времени, а a — за масштаб. Большие значения a соответствуют низким частотам, меньшие — высоким.

Для того чтобы было возможно обратное преобразование, функция $\psi(x)$ должна удовлетворять следующему условию:

$$C_\psi = \int_0^\infty \frac{|\Psi(\omega)|^2}{\omega} d\omega < \infty,$$

где $\Psi(\omega)$ — преобразование Фурье функции $\psi(x)$.

Если $\psi(x)$ — локальная функция, удовлетворяющая данному условию, то ее среднее значение равно нулю и обратное преобразование имеет вид:

$$f(x) = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_0^\infty CTWT_f(a,b) \frac{1}{\sqrt{a}} \psi\left(\frac{x-b}{a}\right) \frac{dad b}{a^2}.$$

Параметры a и b изменяются непрерывно, поэтому множество базисных функций избыточно. Необходима дискретизация этих значений при условии сохранения возможности обратного преобразования. Нетрудно показать, что при следующей дискретизации это условие соблюдается:

$$a = a_0^m; \quad b = nb_0 a_0^m, \quad m, n \in Z, a_0 > 1, b_0 \neq 0.$$

Возможен произвольный выбор b_0 . Выберем $b_0 = 1$. Видно, что параметр местоположения во времени зависит от параметра масштаба, с увели-

чением масштаба увеличивается шаг сдвига. Это интуитивно понятно, так как для анализа с большим масштабом детали не так важны. Таким образом, $\psi_{m,n}(x) = a_0^{-m/2} \psi(a_0^{-m}x - n)$ — базисные функции вейвлет-преобразования.

По аналогии с терминологией преобразования Фурье, введем ряды вейвлетов непрерывного времени (CTWS):

$$(CTWS_f)_{m,n} = d_{m,n} = \int_{-\infty}^{\infty} a_0^{-m/2} \psi(a_0^{-m}x - n) f(x) dx .$$

Восстановление исходного сигнала из последовательности возможно в том случае, если существуют числа $A > 0$ и $B < \infty$, такие что

$$A \|f(x)\|^2 \leq \sum_{m \in Z} \sum_{n \in Z} |d_{m,n}|^2 \leq B \|f(x)\|^2$$

для всех $f(x)$ в $L^2(R)$.

Это означает, что хотя и восстановленный сигнал будет отличаться от исходного, он будет близок к исходному в среднеквадратической норме. Если $A = B = 1$ и $a_0 = 2$, то возможно полное восстановление исходного сигнала и семейство базисных функций $\psi_{a,b}(x)$ образует ортогональный базис. Тогда

$$f(x) = C_\psi \sum_{m \in Z} \sum_{n \in Z} d_{m,n} 2^{-m/2} \psi(2^{-m}x - n) .$$

Кратномасштабный анализ

При анализе сигнала зачастую полезно представить сигнал в виде совокупности его последовательных приближений. Например, при передаче изображения по сети можно сначала передать его грубую версию, а затем, исходя из пропускной способности канала, последовательно ее уточнять.

Кратномасштабный анализ обладает целым рядом полезных свойств, главным из которых является возможность выделения из исходного сигнала его деталей различных масштабов. Это позволяет адаптировать кодер к локальным особенностям сигнала и, тем самым, повысить качество кодирования, не уменьшая степени сжатия.

Кратномасштабным анализом называется описание пространства $L^2(R)$ через иерархически вложенные подпространства V_m , которые не пересекаются, и объединение которых дает в пределе все $L^2(R)$, т. е.

$$\dots \subset V_2 \subset V_1 \subset V_0 \subset V_{-1} \subset V_{-2} \subset \dots, \bigcap_{m \in Z} V_m = \{0\}, \bigcup_{m \in Z} V_m = L^2(R).$$

Далее, эти пространства имеют следующее свойство: для любой функции $f(x) \in V_m$, ее «сжатая» версия будет принадлежать V_{m-1} ,

$$f(x) \in V_m \Leftrightarrow f(2x) \in V_{m-1}.$$

А также, существует такая функция $\phi(x) \in V_0$, что ее сдвиги $\phi_{0,n}(x) = \phi(x - n), n \in Z$ образуют ортонормированный базис пространства V_0 .

Из последнего свойства вытекает, что функции $\phi_{m,n}(x) = 2^{-m/2} \phi(2^{-m}x - n)$ образуют ортонормированный базис пространства V_m . Эти базисные функции называются масштабирующими.

Таким образом, любая функция $f(x) \in L^2(R)$ может быть представлена множеством ее последовательных приближений $f_m(x) \in V_m$. Другими словами, функция $f(x)$ есть предел аппроксимаций $f_m(x) \in V_m$:

$$f(x) = \lim_{m \rightarrow -\infty} f_m(x).$$

Из определения кратномасштабного анализа следует, что функция $f_m(x)$ есть ортогональная проекция $f(x)$ на подпространство V_m , то есть

$$f_m(x) = \sum_n \langle \phi_{m,n}(x), f(x) \rangle \phi_{m,n}(x) = \sum_n c_{m,n} \phi_{m,n}(x).$$

Так как $\phi(x) = \phi_{0,0}(x) \in V_0 \subset V_{-1}$, то

$$\phi_{0,0}(x) = \sqrt{2} \sum_n h_n \phi_{-1,n}(x) = 2 \sum_n h_n \phi(2x - n),$$

где h_n — некоторая последовательность.

Это равенство принято называть масштабирующим уравнением. Имеют место следующие соотношения:

$$\sum_n h_n = 1,$$

$$\delta_{0,k} = \langle \phi_{0,0}(x), \phi_{0,k}(x) \rangle = 2 \sum_n h_n h_{n+2k},$$

и в спектральной области $|H(\omega)|^2 + |H(\omega + \pi)|^2 = 1$, что приводит к $H(0) = 1, H(\pi) = 0$.

Введем в рассмотрение еще один объект — пространства W_m :

$$V_{m-1} = V_m \oplus W_m, \bigcap_{m \in \mathbb{Z}} W_m = \{0\}, \overline{\bigcup_{m \in \mathbb{Z}} W_m} = L^2(R).$$

Пусть $\psi(x) = \psi_{0,0}(x)$ — базисная функция W_0 , тогда

$$\psi_{0,0}(x) = \sqrt{2} \sum_n g_n \phi_{-1,n}(x)$$

для некоторой последовательности g_n .

Определим семейство вейвлет-функций $\psi_{m,n}(x) = 2^{-m/2} \psi(2^{-m}x - n)$.

Имеют место соотношения: $0 = \langle \phi_{0,0}, \psi_{0,0} \rangle = 2 \sum_n h_n g_n$. Легко видеть, что

выбор $g_n = (-1)^n = h_{-n+2l+1}$ корректен для всех целых l .

Определения вейвлет-функций позволяют записать любую функцию $f(x) \in L^2(R)$ в виде суммы ее проекций на $W_j, j \in \mathbb{Z}$: $f(x) = \sum_{j=-\infty}^{\infty} e_j(x)$.

Если осуществить анализ до некоторого масштаба m , то $f_m(x)$ будет представлена суммой ее грубой аппроксимации $f_m(x) \in V_m$ и множества деталей $e_j \in W_j$:

$$f(x) = f_m(x) + \sum_{j=-\infty}^m e_j(x) = \sum_n c_{m,n} \phi_{m,n}(x) + \sum_{j=-\infty}^m d_{j,k} \psi_{j,k}(x).$$

Дискретное вейвлет-преобразование

Пусть имеется некоторый дискретный сигнал c_n . Интерпретируем его как коэффициенты разложения некоторой функции $f_0(x) \in V_0$ по базису масштабирующих функций подпространства V_0 :

$$f_0(x) = \sum_n c_{0,n} \phi_{0,n}(x), \text{ где } c_{0,n} = c_n,$$

тогда мы можем вычислить аппроксимации этой функции, принадлежащие пространствам V_1, V_2, \dots . Согласно идее кратномасштабного анализа, функция $f_0(x)$ раскладывается на сумму:

$$f_0(x) = f_1(x) + e_1(x) = \sum_k c_{1,k} \phi_{1,k}(x) + \sum_k d_{1,k} \psi_{1,k}(x).$$

Таким образом, мы получили две новые последовательности коэффициентов $c_{1,n}$ и $d_{1,n}$. Этот процесс может быть продолжен разложением $f_1(x), f_2(x), \dots$, и функция $f_0(x)$ (а следовательно, и исходный дискретный сигнал c_n) будет представлена совокупностью коэффициентов $d_{m,n}, m \in Z^+, n \in Z$. Так определяются дискретные ряды вейвлетов (DTWS).

Учитывая, что масштабирующие функции образуют базисы соответствующих пространств, можно показать, что

$$c_{j,k} = \sqrt{2} \sum_n c_{j-1,n} h_{n+2k} \quad \text{и} \quad d_{j,k} = \sqrt{2} \sum_n c_{j-1,n} g_{n+2k},$$

обратный процесс:
$$c_{j-1,n} = \sqrt{2} \sum_k c_{j,k} h_{n+2k} + \sqrt{2} \sum_k d_{j,k} g_{n+2k}.$$

Последовательности h_n и g_n называются фильтрами. Имеют место следующие соотношения на коэффициенты фильтров:

$$\begin{aligned} 2 \sum_k (h_{n+2k} h_{p+2k} + g_{n+2k} g_{p+2k}) &= \delta_{n,p}, \\ 2 \sum_n h_{n+2k} h_{n+2p} &= 2 \sum_n g_{n+2k} g_{n+2p} = \delta_{k,p}, \\ 2 \sum_n h_{n+2k} h_{n+2p} &= 0. \end{aligned}$$

На практике мы имеем дело с дискретным сигналом конечной длины. Следовательно, нужно скорректировать вейвлет-преобразование таким образом, чтобы из сигнала какой-то длины получать последовательность вейвлет-коэффициентов той же длины. Получившееся преобразование назовем дискретным вейвлет-преобразованием (DWT).

Обозначим через вектор v^j последовательность конечной длины $c_{j,n}$. Этот вектор преобразуется в вектор v^{j+1} , содержащий последовательности $c_{j+1,n}$ и $d_{j+1,n}$, каждая из которых вполнину короче, чем $c_{j,n}$: $v^{j+1} = M_j v^j$, где M_j — квадратная матрица, состоящая из нулей и элемен-

тов h_n умноженных на $\sqrt{2}$. В силу свойств последовательности h_n , матрица M_j является ортонормированной, следовательно $M_j^T = (M_j)^{-1}$.

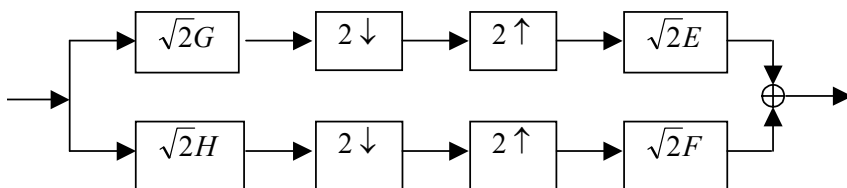
Таким образом, один шаг итерационной процедуры дискретного вейвлет-преобразования для сигнала длины $N = 8$ и фильтра длины $L = 4$ есть:

$$\begin{bmatrix} c_{1,0} \\ c_{1,1} \\ c_{1,2} \\ c_{1,3} \\ d_{1,0} \\ d_{1,1} \\ d_{1,2} \\ d_{1,3} \end{bmatrix} = \sqrt{2} \begin{bmatrix} h_0 & h_1 & h_2 & h_3 & & & & \\ & h_2 & h_3 & -h_2 & h_1 & -h_0 & & \\ h_2 & h_3 & -h_2 & h_1 & -h_0 & & & \\ h_3 & -h_2 & h_1 & -h_0 & h_1 & -h_0 & h_1 & -h_0 \\ h_1 & h_0 & & & h_3 & -h_2 & h_1 & -h_0 \end{bmatrix} \begin{bmatrix} c_{0,0} \\ c_{0,1} \\ c_{0,2} \\ c_{0,3} \\ c_{0,4} \\ c_{0,5} \\ c_{0,6} \\ c_{0,7} \end{bmatrix}$$

Полное преобразование состоит из $\log_2 N$ итераций.

Следует заметить, что в 4-й и 8-й строках последовательность h_n циклически сдвинута: коэффициенты, выходящие за пределы матрицы, помещены в ту же строку слева. Это означает, что DWT есть точно один период длины N DTWS сигнала \tilde{c}_n , получаемого путем бесконечного периодического продолжения сигнала c_n .

В теории вейвлет-преобразований принято описывать применение DWT блок-схемой банка фильтров:



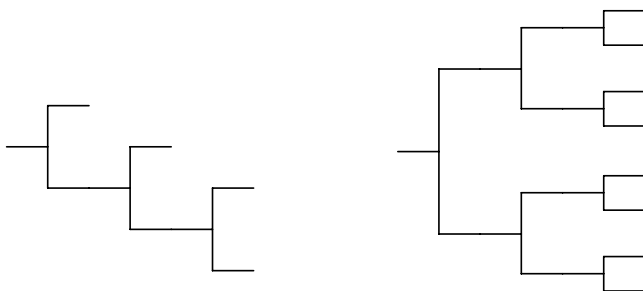
Фильтры G и H означают фильтрацию при помощи g_m и h_n , а фильтры E и F — фильтрацию при помощи g_{-m} и h_{-n} соответственно. Стрелка вниз с коэффициентом 2 обозначает децимацию результата в 2 раза, а стрелка вверх — умножение на 2.

В нижней ветви схемы осуществляется низкочастотная фильтрация. В результате получается аппроксимация исходного сигнала, лишенная деталей низкочастотная субполоса (НЧ). В верхней ветви выделяется высокочастотная составляющая сигнала (ВЧ). Дальнейшее вейвлет-преобразование получается путем рекурсивного применения того же банка фильтров к НЧ составляющей сигнала. При вейвлет-преобразовании изображения, преобразование последовательно применяется и к строкам, и к столбцам (строится так называемая пирамида Маллата).

Адаптивные вейвлет-преобразования

Как уже было сказано выше, вейвлет-преобразование изображения осуществляется путем каскадного соединения банков фильтров по НЧ составляющей сигнала. Такой подход применим при неявном предположении, что основная информация о сигнале содержится в его НЧ области. В общем случае это предположение неверно. Желательно было бы иметь схему способную адаптироваться к конкретным свойствам сигнала. Был разработан целый ряд таких схем, и они получили название «пакет вейвлетов».

Метод пакета вейвлетов основан на определении области, по которой выгоднее производить каскадирование банков фильтров. Для этого сначала проводится каскадирование по обеим полосам. Получается полное, сбалансированное дерево:



вейвлет-декомпозиция

полная декомпозиция

Далее, вводится некоторая функция стоимости, на основе которой определяется наилучший путь по этому дереву.

Пакеты вейвлетов были разработаны и изучены Р. Койфманом и М. Вилкерхаузером. В качестве стоимости они использовали энтропию, понимаемую ими, как концентрацию числа коэффициентов, требующихся для опи-

сания сигнала. Данная функция M будет большой, если коэффициенты примерно одной величины, и малой, если все, кроме нескольких коэффициентов, близки к нулю: $M = e^{-\sum_n p_n \log p_n}$, где $p_n = |x|^2 \|x\|^{-2}$. Энтропия вычисляется для каждого узла полного дерева. Далее, сравнивается энтропия двух потомков и их общего предка на дереве. Если энтропия предка оказалась меньше, то от декомпозиции отказываются. Алгоритм рекурсивно продолжается до достижения вершины дерева.

Для решения задачи сжатия изображений выбор энтропии в качестве функции стоимости, по-видимому, не является наилучшим. В некоторых работах в качестве функции стоимости используется функционал Лагранжа: $J = D + \lambda R$, где D — это искажение (средний квадрат ошибки), вносимое за счет непередачи коэффициента узла, R — количество бит, требуемых для описания коэффициента этого узла, λ — множитель Лагранжа. Алгоритм принятия решения такой же, как и в предыдущем методе. Этот алгоритм получил название одиночного (частотного) дерева.

Вейвлет-пакеты являются более гибким средством, чем вейвлет-преобразование, но, не смотря на это, они не изменяются в пространстве. Поскольку в данной статье рассматривается применение вейвлет-преобразований к кодированию видеопоследовательностей (и изображений, как их частного случая), то имеет смысл рассмотреть алгоритмы, учитывающие изменение исходного сигнала в пространстве. Такие методы получили название алгоритма двойного дерева.

Алгоритм двойного дерева основан на совместном поиске наилучшего разбиения исходного сигнала и в пространственной, и в частотной области. Для простоты, будем предполагать, что исходный сигнал является изображением и в пространственной области он разбивается на некоторые прямоугольные области, размерность которых кратна степени двойки. Далее, для каждой области строится разложение пакетами вейвлетов. Затем значения стоимостей Лагранжа каждой области записываются в виде бинарного дерева. К получившемуся дереву применяется алгоритм одиночного дерева для нахождения наилучшего разбиения исходного сигнала на области.

Алгоритм двойного дерева обладает некоторой асимметричностью. Действительно, деревья в частотной области строятся над пространственными областями, а не наоборот. Этот недостаток можно устранить, построив дерево, в котором кандидатом на дальнейшее разбиение будет являться как пространственная область, так и частотная субполоса. Это дерево (так называемое частотно-пространственное дерево), имеет структуру квадродерева. В самом деле, каждый родительский узел имеет двух пространственных

и двух частотных потомков. Обрезание такого дерева происходит путем сравнения стоимостей Лагранжа: сравниваются пространственные и частотные пары в направлении от листьев к корню дерева. В результате выполнения алгоритма получается оптимальное бинарное дерево разбиения по частоте и пространству полной глубины.

В заключение темы адаптивных вейвлет-преобразований, приведем значения вычислительной сложности всех трех алгоритмов:

Алгоритм	Сложность вычислений
одинокое дерево	$O(Nd)$
двойное дерево	$O(Nd^2)$
пространственно-частотное дерево	$O(N2^d)$

где N — размерность сигнала, d — максимальная высота дерева.

ПРИМЕНЕНИЕ ВЕЙВЛЕТ-ПРЕОБРАЗОВАНИЙ ДЛЯ КОДИРОВАНИЯ ИЗОБРАЖЕНИЙ

Вейвлет-кодер состоит из трех основных частей: декоррелирующее преобразование, процедура квантования и энтропийное кодирование. Рассмотрим каждую составляющую подробно.

Декоррелирующее преобразование

Выбор оптимального базиса вейвлетов чрезвычайно сложная и вряд ли решаемая задача. Известен ряд критериев построения «хороших» вейвлетов, среди которых наиболее важными являются: гладкость, точность аппроксимации, величина области определения, частотная избирательность. К сожалению, на данный момент наилучшая комбинация этих свойств неизвестна. Более того, в последних работах требование гладкости базиса подвергается сомнению.

Простейшим видом вейвлет-базиса для изображений является separable базис, получаемый сжатием и растяжением одномерных вейвлетов. Использование разделимого преобразования сводит проблему поиска эффективного базиса к одномерному случаю, что сильно упрощает задачу, и почти все известные кодеры используют этот подход. Однако неразделимые базисы могут быть более эффективными.

Прототипами базисных функций разделимого преобразования являются функции $\phi(x)\phi(y), \phi(x)(y), (x)\phi(y), (x)(y)$. На каждом шаге преобразования выполняется не одно, а два разбиения по частоте: сначала по строкам, затем по столбцам изображения. Предположим, изображение имеет размерность $N \times N$, тогда после каждого шага преобразования получаются 4 изображения размера $\frac{N}{2} \times \frac{N}{2}$ (см. рис.).

Д. Вилласенор систематически протестировал все биортогональные блоки фильтров минимального порядка с длиной фильтра не более 36. Наилучшим фильтром оказался сплайновый фильтр 7/9. Этот фильтр наиболее часто используется в вейвлет-кодерах изображений.

HCH_2	VCH_2	VCH_1
HCV_2	VCV_2	
HCV_1		VCV_1

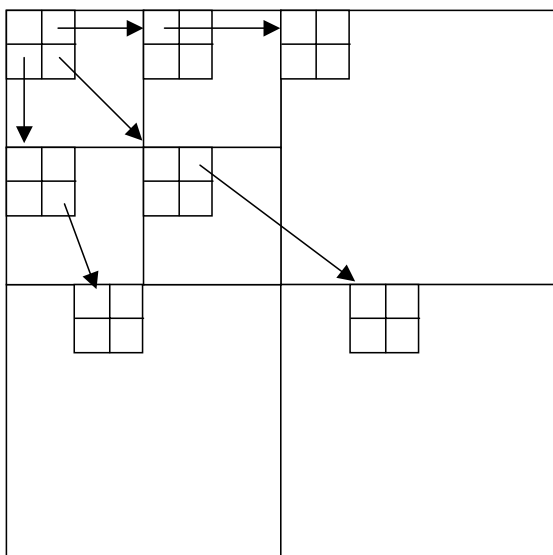
Два уровня вейвлет-преобразования изображения

Далее, традиционное кодирование вейвлет-коэффициентов основано на предположении, что большая часть энергии сосредоточена в малом числе коэффициентов, которые квантуются в соответствии с их значением. Как показывает практика, это предположение бывает неверно.

Традиционное кодирование может быть улучшено путем введения операторов выбора. Вместо квантования коэффициентов в заранее определенном порядке, кодер может выбирать нужные для кодирования элементы

(наиболее значимые). Наиболее значительным результатом этого подхода является разработка алгоритма нуль-дерева и его модификаций.

Впервые идея нуль-дерева была предложена Л. Льюисом и Г. Ноулесом. В их алгоритме применялась древовидная структура данных для описания вейвлет-коэффициентов. Такая структура получается в результате применения двухканального разделимого вейвлет-преобразования. Корневой узел дерева представляет коэффициент в самой НЧ области и имеет три потомка. Узлы дерева соответствуют вейвлет-коэффициентам масштаба, равного их высоте в дереве. Каждый узел имеет четыре потомка, соответствующих вейвлет-коэффициентам следующего уровня и того же пространственного расположения.



Зависимости между коэффициентами вейвлет-преобразования, используемые в алгоритме нуль-дерева

Для каждого из коэффициентов самой НЧ области существуют три таких дерева.

Квантование нуль-дерева основано на наблюдении, что если коэффициент мал, то и все его потомки на дереве зачастую тоже малы. Это объясняется тем, что значимые коэффициенты возникают вблизи контуров и тек-

стур, которые локальны. Дерево или поддерво, которое содержит только незначимые коэффициенты, называется нуль-деревом.

Далее используется следующий итеративный алгоритм квантования. Вначале каждый узел квантуется. Если значение квантованного узла меньше некоторого порога, поддерво, начинающееся с этого узла, объявляется нуль-деревом, и его потомки игнорируются. Эти потомки будут восстановлены декодером как нули. Иначе переходят к четырем потомкам узла, и процедура квантования повторяется. Если узел не имеет потомков, начинает обрабатываться следующий корневой узел.

Особую эффективность этому алгоритму придает совместное кодирование нулей при помощи кодера длин серий (run-length encoder). Для повышения эффективности на вход кодера коэффициенты должны подаваться в определенном порядке, обеспечивающем наибольшие длины последовательностей нулей. Например, в JPEG это зигзагообразное сканирование. Наиболее важным вкладом Льюиса и Ноулеса была демонстрация того, что область вейвлет-коэффициентов хорошо приспособлена для работы такого кодера. В самом деле, генерируются очень длинные последовательности нулей, и нет нужды передавать их длины, поскольку высота дерева известна.

Характеристики этого алгоритма незначительно превосходят JPEG, хотя визуальное качество изображения заметно лучше. Недостатком алгоритма является способ распознавания нуль-дерева. Как было отмечено, если коэффициент мал, то и его потомки малы. В случае если это не так, данный алгоритм ведет к большим искажениям. Преимуществом алгоритма является его вычислительная простота.

Следующее поколение кодеров, в частности алгоритмы Шапиро и Саида–Перельмана, улучшают способ распознавания нуль-дерева.

Шапиро разработал так называемый алгоритм вложенного нуль-дерева (Embedded Zerotree Wavelet encoder — EZW). Он основан на передаче и ненулевых данных, и некоторой карты значений. Если имеется незначимый родительский узел, то очень вероятно, что его потомки тоже будут незначимы. Так что в большинстве случаев генерируется символ нуль-дерева. Если один или больше потомков незначимого узла являются значимыми, то генерируется символ «изолированного нуля». Вероятность этого события ниже, следовательно, для кодирования требуется меньше бит.

Алгоритм EZW генерирует вложенный, иерархический код. Это позволяет осуществить прогрессивную передачу изображения последовательным уточнением при приеме. Подобные коды имеют большой практический интерес по следующим причинам:

- 1) возможность точного регулирования скорости передачи;
- 2) возможность восстановления всего изображения при прекращении приема декодером в любой точке. При этом изображение будет максимально хорошего качества для данного числа принятых бит;
- 3) возможность быстрого просмотра изображения в удаленной базе данных. Для поиска достаточно и грубой копии, а при нахождении нужного изображения оно декодируется полностью.

Алгоритм EZW генерирует вложенный код побитово следующим образом. В начале выполняется частичное упорядочение вейвлет-коэффициентов путем сравнения каждого коэффициента (ВК) с некоторым пороговым значением T . Если $ВК < T$, то коэффициент считаем значимым, иначе — незначимым. Сканирование производится от НЧ полос к ВЧ. Для кодирования знака и позиции ВК используется двухбитный символ:

- « \pm » — знак ВК,
- «0» — показывает, что ВК незначимый,
- «корень нуль-дерева» — показывает, что ВК незначимый со всеми своими потомками (ВК — является корнем нуль-дерева),

т.е. используется межполосная, пространственная корреляция ВК. После генерации и передачи карты значений для значащих коэффициентов должны быть переданы биты, уточняющие их значение («карта данных»). Далее карта значений и карта данных сжимаются арифметическим кодером. В том случае, если не исчерпан ресурс скорости передачи, порог T делится на 2 и процесс повторяется.

На начальных итерациях (при большом T) в карте значений будет встречаться много нулей. Роль нуль-дерева заключается в предотвращении передачи лишних нулей. Символ нуль-дерева может снова и снова генерироваться для данного ВК, пока он не станет больше порогового значения, и тогда будет передано квантованное значение ВК.

А. Саид и В. Перельман улучшили алгоритм EZW. Их версия кодера получила название «установка подразделений в иерархических деревьях» (Set Partition In Hierarchical Trees — SPIHT). Он основан на обнаружении схожих фрагментов в различных поддеревьях вейвлет-коэффициентов.

Квантование

Обычно в алгоритмах кодирования изображений используется скалярное квантование. Числовая прямая разбивается на отрезки равной длины (неравномерная сетка порождает неравномерное скалярное квантование).

Далее, квантуемый элемент, попадающий в какой-то отрезок, заменяется центром этого отрезка.

Более эффективным методом квантования является векторное квантование. Этот алгоритм состоит из двух этапов: составления кодовой книги и собственно кодирования. В кодовой книге хранятся взвешенные комбинации двумерных блоков коэффициентов. Каждому блоку приписан его код, получаемый посредством алгоритма Хаффмана. Кодирование состоит из нахождения для каждого кодируемого блока ближайшего блока из кодовой книги относительно некоторой меры, например среднеквадратичной разности. В выходной поток записываются соответствующие коды ближайших блоков. Недостаток этого метода в том, что он требует хранения кодовой книги. Однако, если кодовая книга составлена из элементов самого изображения, то затраты на передачу кодовой книги — минимальны.

Энтропийное кодирование

Субоптимальное энтропийное кодирование можно реализовать при помощи алгоритма арифметического кодирования. Кодеру требуется оценить распределение квантованных коэффициентов. Эта оценка получается путем аппроксимации распределения коэффициентов гауссовской плотностью и вычислением параметров распределения. Оценка параметров может производиться и в процессе работы, «на ходу». Такой подход позволяет учитывать локальные изменения статистики изображения.

Ввиду того, что изображение не является случайным гауссовским источником, коэффициенты преобразования обладают определенной структурой. Энтропийный кодер может каким-то образом использовать эту структуру, осуществляя предсказание параметров. В ряде работ отмечалось, что это приводит к незначительному улучшению эффективности.

На практике вместо арифметического кодера используется кодер Хаффмана. Причина заключается в меньшей вычислительной сложности алгоритма Хаффмана, а также в том, что большинство вариаций алгоритма арифметического кодирования запатентованы. Однако, поскольку патент имеет ограниченный срок действия (а у большей части патентов этот срок истек в 2001 году), в скором времени возможно появление целой серии кодеров, использующих арифметическое кодирование.

ЗАКЛЮЧЕНИЕ

Современные исследования в области сжатия изображений ведутся по разным направлениям: нелинейные вейвлет-преобразования, помехозащищенные кодеры и т.д.

Особый интерес представляет адаптация вейвлет-кодирования изображений для кодирования видео. Здесь можно сочетать внутрикадровое кодирование с межкадровым предсказанием, аналогично стандарту MPEG4. Можно также рассматривать видеопоследовательность как трехмерный сигнал и применять трехмерный вейвлет-анализ. Здесь возникают трудности с представлением получающейся древовидной структуры, но работы в этом направлении активно ведутся.

В заключение стоит отметить, что вейвлеты и сопутствующие им идеи внесли существенный вклад в теорию и практику кодирования изображений и, по-видимому, будут оставаться основным направлением исследований в этой области в ближайшем будущем.

СПИСОК ЛИТЕРАТУРЫ

1. **Воробьев В.И., Грибунин В.Г.** Теория и практика вейвлет-преобразования. — СПб: ВУС, 1999.
2. **Mallat S.** A Theory for Multiresolution Signal Decomposition: The Wavelet Representation // IEEE Transactions on Pattern Analysis and Machine Intelligence, 1989. — Vol. 11. — P. 674–693.
3. **Shapiro J.** Embedded Image Coding Using Zerotrees of Wavelet Coefficients // IEEE Transactions on Signal Processing, 1993. — Vol. 41, No. 12.
4. **Said A., Pearlman W.** A New Fast and Efficient Image Codec Based on Set Partitioning in Hierarchical Trees // IEEE Transactions on Circuits and Systems for Video Technology, 1996. — Vol. 6. — P. 243–250.
5. **Antonini M., Barlaud M., Mathieu P., Daubechies I.** Image coding using wavelet transform // IEEE Transactions On Image Processing, 1992. — Vol. 1, № 2. — P. 205–220.