JPEG XS – еще полшага в IP-будущее

Часть 1

UHD и IP: тенденции и проблемы развития индустрии телевизионного вещания и видеопроизводства

Константин Гласман

Сегодня в индустрии телевизионного вещания и видеопроизводства можно наблюдать две основные тенденции развития. Первая из них связана с увеличением четкости изображения до 4K (UHD-1) и даже 8K (UHD-2) в сочетании с расширенным динамическим диапазоном (HDR), ростом частоты кадров и количества передаваемых и обрабатываемых потоков данных. Проявлением этой тенденции является значительное увеличение объема данных в изображении и рост требований к пропускной способности каналов связи. Вторая тенденция – стремление избавиться от специализированных кабелей и инфраструктуры, основанной на последовательном цифровом интерфейсе (SDI), и использовать IP-инфраструктуру информационных сетей с технологией Ethernet на физическом уровне.

Пропускная способность каналов связи увеличивается, но гораздо медленнее, чем растут требования, и, что существенно, за счет огромных инвестиций, которые будут окупаться в течение ряда лет. В вещательных студиях пока развернуты главным образом инфраструктуры SDI, в основном HD-SDI (1,5 Гбит/с) или 3G-SDI (3 Гбит/с). Однако вещатели постепенно переходят на инфраструктуру IP. В настоящее время предпочтение отдается каналам 1 Gigabit Ethernet (1 GbE) для дистанционного производства и инфраструктурам 10 GbE для межстудийного обмена. Переход к форматам 4K и 8K требует использования технологий 25-гигабитного и 100-гигабитного Ethernet на физическом и канальном уровнях IP-инфраструктуры телевизионных центров и компаний (табл.1).

Но каналы передачи данных следующего поколения – 25, 40 или 100 GbE – пока не развернуты или еще слишком дороги и не могут быть рентабельными. Следовательно, передача некомпрессированного видео в реальном масштабе времени становится практически невозможной в рамках существующих систем и инфраструктур.

Таблица 1. Физические каналы, необходимые для передачи потоков некомпрессированного видео разных форматов

Формат видеоданных	Физический канал			
	1 GbE	10 GbE	25 GbE	100 GbE
2K 60p 422 10 бит	✗	✓	✓	✓
4K 60p 444 12 бит	✗	✗	✓	✓
8K 120p 422 10 бит	✗	✗	✗	✓

Решение проблем – видеокомпрессия! Но какая?

В реалиях телевидения сегодняшнего дня только видеокомпрессия, или сжатие потоков видеоданных, позволяет упростить и сгладить последовательный переход от одного поколения форматов, протоколов и инфраструктур к следующим. К настоящему времени предложено огромное число алгоритмов компрессии, некоторые из которых были доведены до практической реализации и международной стандартизации. Но к видеокомпрессии, назначение которой – обеспечить передачу высокоскоростных потоков видеоданных по каналам связи со сравнительно невысокой пропускной способностью в условиях телевизионной студии или системы дистанционного производства программ, предъявляются особые требования. По этой причине в 2016 году комитет JPEG, официально называемый ISO/IEC SC29 WG1, начал работу по созданию нового кодека, в качестве первого шага определив требования, наиболее актуальными из которых являются:

- форматы изображения RGB 444 и YCbCr 444/422 с разрядностью кодирования каждого компонента до 12 битов на отсчет;
- компрессия «визуально без потерь», то есть без видимых искажений;
- сквозная задержка в одном цикле кодирования-декодирования не более 32 строк;
- внутрикадровое кодирование – отдельные кадры должны кодироваться и декодироваться независимо друг от друга, что устраняет необходимость во внешней буферной памяти кадров;
- отсутствие заметного накопления искажений в течение нескольких циклов кодирования-декодирования (до 10 циклов);
- возможность программной реализации в режиме реального времени для форматов сверхвысокой четкости UHD на стандартных современных компьютерах;
- поддержка нескольких платформ для построения кодеков: центральный процессор (CPU) компьютера, графический процессор (GPU), специализированная интегральная схема (ASIC), программируемая пользователем вентильная матрица (FPGA);
- невысокая сложность кодека, определяемая как максимальный процент использования ресурсов некоторой целевой программируемой матрицы FPGA.

Комитет JPEG в 2016 году получил следующие предложения: упрощенная версия системы компрессии HEVC (High Efficiency Video Coding), модифицированные версии систем JPEG 2000, VC-2 (SMPTE 2042 Video Compression),

JPEG-LS (Low Complexity Lossless), а также кодеки видеокомпрессии на основе дискретного косинусного преобразования DCT и на основе дискретного вейвлет-преобразования DWT с использования банков фильтров Хаара и Ле Галла 5/3. Комитет JPEG счел сложность предложений на основе JPEG 2000 и HEVC слишком высокой для интеграции в целевую архитектуру FPGA. Предложение на основе VC-2 было отклонено из-за низкого уровня качественных показателей, хотя его сложность вполне соответствовала требованиям. Предложение на основе JPEG-LS не смогло достичь очень высокого целевого качества во всем наборе тестов. Предложение кодека на основе DWT с использованием банков фильтров Хаара было отозвано его авторами. После долгих обсуждений осталось два предложения: кодеки видеокомпрессии на основе дискретного вейвлет-преобразования DWT с использованием банков фильтров Ле Галла 5/3 и на основе дискретного косинусного преобразования DCT. Они обеспечивали стабильно высокие качественные показатели и достаточно низкую сложность, а потому соответствовали требованиям.

Мезонинная видеокомпрессия

В итоге комитет JPEG принял решение объединить достоинства обоих отобранных кодеков и попросить авторов предложить соответствующую технологию кодирования для новой мезонинной видеокомпрессии с малой задержкой, без визуальных потерь и невысокой алгоритмической сложностью для приложений типа VoIP (видео поверх IP) на основе дискретного вейвлет-преобразования DWT с использованием банков фильтров Ле Галла 5/3.

В табл. 2 перечислены некоторые из вариантов использования разрабатываемой технологии для передачи видеопотоков по существующим и разворачиваемым сейчас и в близком будущем инфраструктурам. В таблице приведены формат видео и соответствующая формату скорость потока видеоданных, а также целевой физический канал и его доступная пропускная способность. Исходя из этого, можно оценить требуемое отношение компрессии (округленное значение приведено в последнем столбце таблицы). Как видно, максимально необходимое для этих вариантов отношение достигает 6:1, что эквивалентно шести передаваемым битам на один цветной элемент изображения (6 bpp) для 36-разрядного потока некомпрессированных видеоданных в формате RGB 444 12 бит или 4 bpp для формата RGB 444 8 бит.

Соответствующий данным табл. 2 кодек видеокомпрессии должен позволять увеличивать четкость, динамический диапазон, частоту кадров и количество потоков, используя существующие каналы связи и первые элементы новой сетевой IP-инфраструктуры, практически сохраняя при этом достоинства несжатого потока, то есть совместимость, качество «визуально без потерь», низкую задержку при кодировании и декодировании, простоту реализации и быстрое программное обеспечение, работающее на процессорах общего назначения и программируемых интегральных схемах.

Приведенные соображения помогают понять название этой системы видеокомпрессии – мезонинная. Как известно, мезонин (*итал.* mezzanino – «промежуточный») – это надстройка, полуэтаж, главное назначение которого – расширить жилую площадь. Мезонинная система видеокомпрессии – это промежуточная надстройка, упрощающая переход от существующего поколения форматов, протоколов и инфраструктур к следующему и делающая на этом пути половину шага, поскольку полный шаг затруднен по экономическим причинам.

Работа по созданию мезонинной видеокомпрессии была продолжена. Ключевые технологические решения были найдены институтом Fraunhofer IIS и компанией intoPIX. В 2019 году эта видеокомпрессия была стандартизована. Новому стандарту ISO/IEC 21122 было дано название JPEG XS (XS – сокращение от eXtra Speed и eXtra Small, фокусирующее внимание на главных особенностях кодека). В настоящее время доступна уже вторая редакция стандарта.

JPEG XS – это отличное технологическое решение для передачи видеоданных поверх IP, разработанное для удовлетворения требований рабочих процессов цифрового телевизионного производства в условиях сегодняшнего дня и ближайшего будущего. Но оно, несомненно, найдет также применение в цифровом кино, в профессиональной фотографии, в Pro-AV, в автомобильной индустрии, в сфере виртуальной и дополненной реальности (VR и AR) и других областях.

Таблица 2. Расчет необходимого отношения компрессии для разных вариантов использования разрабатываемой технологии

Формат видеоданных	Скорость потока видеоданных, Гбит/с	Целевой физический канал	Доступная пропускная способность, Гбит/с	Отношение компрессии
2K 60p 422 10 бит	2,7	HD-SDI	1,33	2:1
2K 120p 422 10 бит	5,4	HD-SDI	1,33	4:1
4K 60p 422 10 бит	10,8	3G-SDI	2,65	4:1
2K 60p 422 10 бит	2,7	1GbE	0,85	3:1
2K 60p 444 12 бит	4,8	1 GbE	0,85	6:1
4K 60p 444 12 бит	19	10 GbE	8,5	2,2:1
2×(4K 60p 444 12 бит)	37,9	10 GbE	8,5	4,5:1
3×(4K 60p 422 10 бит)	32,4	10 GbE	8,5	3,8:1
8K 60p 422 10 бит	85	25 GbE	21,25	4:1

Вейвлет – что это?

JPEG XS — это классический видеокодек на основе вейвлет-преобразования. Но что такое вейвлет? Вейвлет-преобразование сигнала можно рассматривать как представление сигнала в виде суперпозиции некоторых базисных функций – волновых пакетов (Wavelet – маленькая волна). Особенностью этих волновых пакетов является то, что все они получены из одной прототипной волны путем растяжения (или сжатия) и смещения. Прототипная волна может рассматриваться как импульсная реакция базового фильтра. Тогда вейвлет-преобразование сводится к совокупности процессов фильтрации и децимации (рис.1).

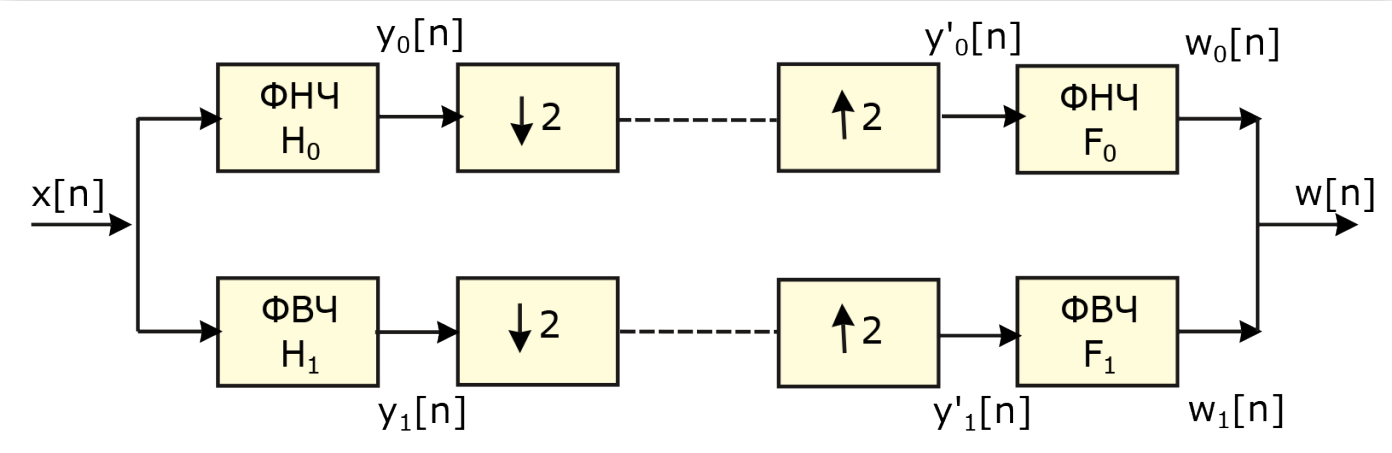

Рис. 1. Прямое и обратное вейвлет-преобразование

Преобразуемый сигнал подвергается фильтрации с помощью фильтров нижних (ФНЧ) и верхних (ФВЧ) частот, которые делят диапазон частот исходного сигнала на две половины. И низкочастотный $y_0[n]$, и высокочастотный $y_1[n]$ компоненты сигнала, полученные при фильтрации сигнала x[n], имеют в два раза более узкую полосу частотных составляющих, чем исходный сигнал. Поэтому в соответствии с теоремой Котельникова-Шеннона они могут быть дискретизированы с частотой, равной половине частоты дискретизации исходного сигнала. Входной сигнал схемы x[n] предполагается цифровым, поэтому после низкочастотной и высокочастотной фильтрации с помощью цифровых фильтров может быть просто исключен каждый второй отсчет, что означает децимацию, или прореживание, выполняемое в схеме рис.1 после фильтрации. Благодаря прореживанию суммарное число отсчетов сигналов $y_0[n]$ и $y_1[n]$ на некотором временном интервале оказывается равным числу отсчетов сигнала x[n] на том же интервале. Такое преобразование называется прямым дискретным вейвлет-преобразованием DWT (Discreet Wavelet Transform). Фильтры ФНЧ и ФВЧ с частотными характеристиками $H_0(z)$ и $H_1(z)$ соответственно часто называют фильтрами анализа сигнала.

Обратное преобразование выполняется с помощью фильтров синтеза нижних и верхних частот с частотными характеристиками $F_0(z)$ и $F_1(z)$ соответственно. Сначала удваивается частота следования отсчетов, и после каждого отсчета вставляется дополнительный нулевой. Недостающие отсчеты восстанавливаются путем интерполяции, а ФНЧ и ФВЧ играют роль интерполирующих фильтров. Затем сигналы с выходов фильтров ФНЧ и ФВЧ складываются. Какими должны быть частотные характеристики фильтров анализа и синтеза для того, чтобы сигнал на выходе устройства синтеза был как можно более точной копией сигнала на входе устройства анализа? Ответ на вопрос дает теория квадратурных зеркальных фильтров QMF (Quadrature Mirror Filter) анализа и синтеза, которые обеспечивают совершенное восстановление исходного сигнала. В отсутствии шума канала и квантования такие фильтры обеспечивают идеальную реконструкцию входного сигнала без искажений.

В системе видеокомпрессии JPEG XS используются банки фильтров Ле Галла 5/3, разработанные в соответствии с концепцией QMF. Выходные сигналы фильтров ФНЧ и ФВЧ прямого преобразования (фильтров анализа) находятся в результате суммирования с весовыми коэффициентами соответственно 5 и 3 последовательных отсчетов входного сигнала:

$$y_0[n] = -\frac{1}{8}x[n+2] + \frac{1}{4}x[n+1] + \frac{3}{4}x[n] + \frac{1}{4}x[n-1] - \frac{1}{8}x[n-2],$$

$$y_1[n] = -\frac{1}{2}x[n+1] + x[n] - \frac{1}{2}x[n-1].$$

Выходные сигналы фильтров ФНЧ и ФВЧ обратного преобразования (фильтров синтеза) находятся в результате суммирования с весовыми коэффициентами соответственно 3 и 5 последовательных отсчетов сигналов y_0 и y_1:

$$w_0[n] = \frac{1}{2}y_0[n+1] + y_0[n] + \frac{1}{2}y_0[n-1],$$

$$w_1[n] = -\frac{1}{8}y_1[n+2] - \frac{1}{4}y_1[n+1] + \frac{3}{4}y_1[n] - \frac{1}{4}y_1[n-1] - \frac{1}{8}y_1[n-2].$$

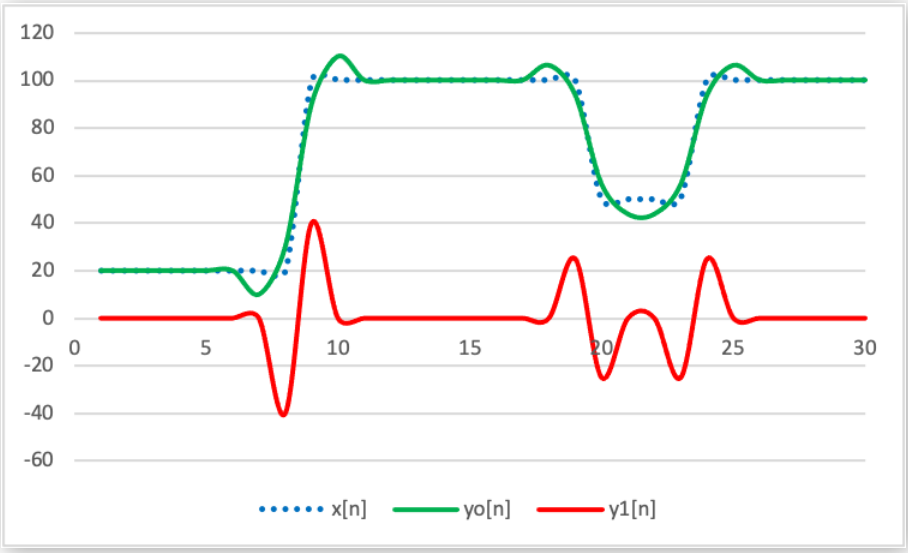

Рис. 2. Частотный вейвлет-анализ сигнала

Рис. 2 иллюстрирует типичные результаты частотного анализа DWT для импульсного сигнала x[n] (синяя пунктирная линия). Зеленая линия показывает сигнал после ФНЧ, красная – после ФВЧ. Как видно, низкочастотный сигнал представляет собой сглаженную версию входного, а высокочастотный отображает перепады входного сигнала (для наглядности в представлении взаимосвязи диаграмм не показана задержка сигналов в фильтрах).

Можно сказать, что прямое преобразование DWT выполняет декомпозицию сигнала на низкочастотную и высокочастотную составляющие. К низкочастотной составляющей можно еще раз применить прямое вейвлет-преобразование и разделить ее на две составляющие с помощью двух дополнительных фильтров ФНЧ и ФВЧ. Такая процедура, показанная на рис. 3, выполняет декомпозицию второго уровня. Применять DWT к низкочастотной составляющей предыдущего вейвлет-преобразования можно до любого уровня, что приведет к пирамидальной декомпозиции входного сигнала.

Вейвлет для ТВ-изображения

Применение вейвлет-преобразования к телевизионному изображению имеет особенности, связанные с двумерной структурой телевизионного растра. Сначала фильтрация нижних частот применяется к каждой строке данных телевизионного растра, что дает низкочастотные компоненты строк. Прореживание полученного растра дает прямоугольную матрицу, ширина которой равна половине ширины исходного изображения. Обозначим эту матрицу как L_{x1} (L символизирует низкочастотную фильтрацию, x – фильтрацию в горизонтальном направлении, 1– первый уровень

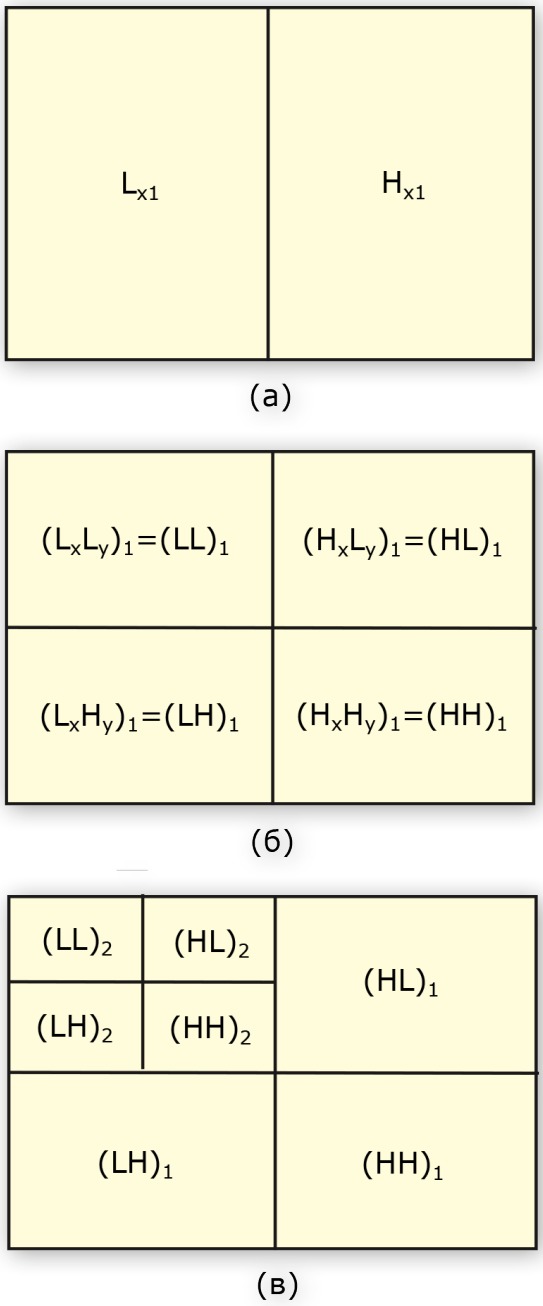

Рис. 4. Схема декомпозиции изображения: а – фильтрация и прореживание в горизонтальном направлении; б – вейвлет-декомпозиция изображения первого уровня; в - вейвлет-декомпозиция изображения второго уровня

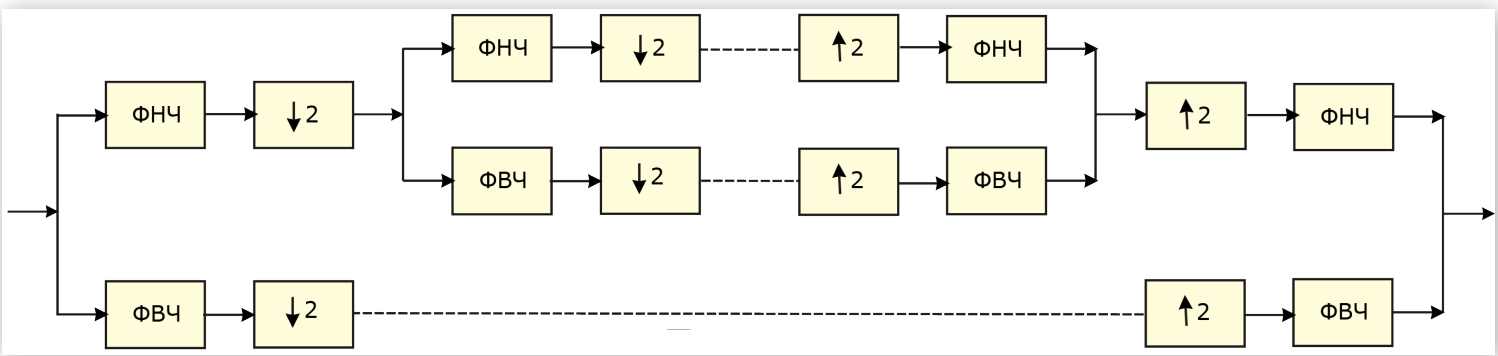

Рис. 3. Вейвлет-декомпозиция сигнала второго уровня

преобразования). Фильтрация верхних частот каждой строки и прореживание (это можно выполнять параллельно с фильтрацией нижних частот) дает матрицу данных H_{x1} (H – фильтр верхних частот), ширина которой также равна половине ширины исходного изображения. Матрицы L_{x1} и H_{x1} занимают место матрицы одного телевизионного кадра на рис. 4а. Таким образом выполнена декомпозиция изображения на составляющие нижних и верхних частот горизонтального направления.

Далее выполняется фильтрация нижних частот вертикального направления для каждого столбца матрицы промежуточных данных рис. 4а. После прореживания это дает две субматрицы: $(L_xL_y)_1$ и $(H_xL_y)_1$, высота которых в два раза меньше матрицы целого кадра (рис. 4б). Они обозначены для краткости записи как $(LL)_1$ и $(HL)_1$ (индексы x и y могут быть опущены, поскольку известно, что сначала выполняется фильтрация в горизонтальном направлении, а потом – в вертикальном). Фильтрация и прореживание верхних частот в вертикальном направлении дает субматрицы $(L_xH_y)_1=(LH)_1$ и $(H_xH_y)_1=(HH)_1$ (рис.4 б).

Субматрица $(LL)_1$ нижних пространственных частот горизонтального и вертикального направлений может быть разложена на составляющие еще раз таким же образом. Вместо нее появится четыре субматрицы второго уровня с половинными горизонтальными и вертикальными размерами (рис. 4в). Затем процесс можно повторить до любого уровня, что приведет к пирамидальной декомпозиции изображения.

Применение к телевизионному изображению одномерной фильтрации в горизонтальном направлении и прореживания (показано на рис. 5). Левое изображение – оригинал, подвергающийся вейвлет-преобразованию. В левой части правого изображения находится отфильтрованный и прореженный низкочастотный компонент изображения, в правой – высокочастотный. Так как после прореживания число отсчетов в каждой составляющей сокращается вдвое в каждой телевизионной строке, оба компонента размещаются на площади исходного изображения. Это визуальный эквивалент рис. 4а.

Рис. 5. Частотная фильтрация и прореживание изображения в горизонтальном направлении

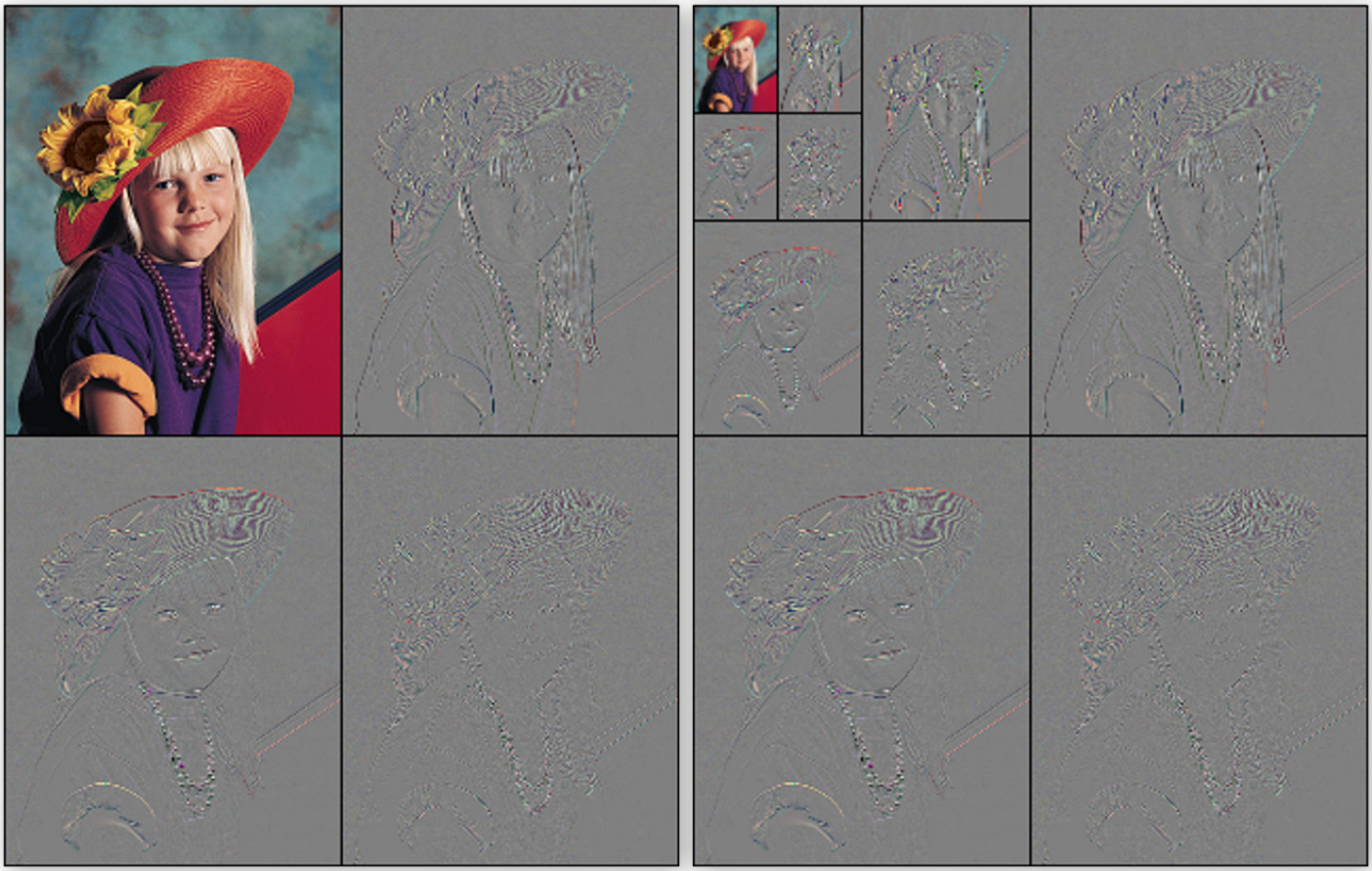

Рис. 6. Декомпозиция изображения первого и третьего уровня

Применение к изображению на правой картинке рис. 5 одномерной фильтрации в вертикальном направлении и прореживания показано на рис. 6. Изображение разделилось на одну низкочастотную и три высокочастотных субматрицы (это визуальный образ рис. 4б). Субматрица в левом верхнем углу дает сглаженную в горизонтальном и вертикальном направлениях версию исходного изображения, четкость которой (число элементов) уменьшилась после прореживания в четыре раза. Субматрица в правом верхнем углу отображает горизонтальные высокочастотные составляющие изображения, сглаженные по вертикали. Она лучше всего рисует границы вертикальных яркостных переходов и вертикальные линии. Второй высокочастотный компонент – субматрица в левом нижнем углу. Она отображает вертикальные высокочастотные составляющие, сглаженные по горизонтали, например, горизонтальные яркостные переходы и горизонтальные линии. В правом нижнем углу располагается третий высокочастотный компонент – субматрица изображения, полученная в результате фильтрации верхних частот горизонтального и вертикального направлений. Она отображает мелкие детали и диагональные яркостные переходы исходного изображения.

На втором уровне вейвлет-преобразования субматрица составляющих изображения нижних горизонтальных и вертикальных пространственных частот вновь разделяется на низкочастотную и три высокочастотных субматрицы с помощью фильтров такого же типа, как на первом уровне. Результаты двумерной вейвлет-декомпозиции изображения третьего уровня показаны на правой картинке рис. 6. Изображение разделилось на одну субматрицу нижних пространственных частот горизонтального и вертикального направлений и девять субматриц с составляющими верхних пространственных частот разных направлений с различными четкостями (различными полосами частот). Следует обратить внимание на то, что после третьего этапа четкость низкочастотного компонента, располагающегося в верхнем левом углу правой картинки рис. 6, в 8 раз меньше четкости исходного изображения (полоса частот каждого компонента, полученного на третьем этапе, равна 1/8 полосы исходного сигнала).

Видеокомпоресссия на основе DWT

Прямое DWT эффективно декоррелирует сигнал исходного изображения, в результате чего большая часть мощности преобразованного сигнала концентрируется в меньшем количестве отсчетов, чем во входном сигнале. Это открывает широкие возможности для применения вейвлет-преобразования в системах видеокомпрессии. Видеокомпрессия на базе вейвлет-преобразования в принципе осуществляется так же, как во всех системах компрессии с использованием унитарных преобразований, например, в компрессии на базе дискретного косинусного преобразования. Компоненты видеосигнала, полученного после вейвлет-преобразования, подвергаются квантованию и энтропийному кодированию (рис. 7).

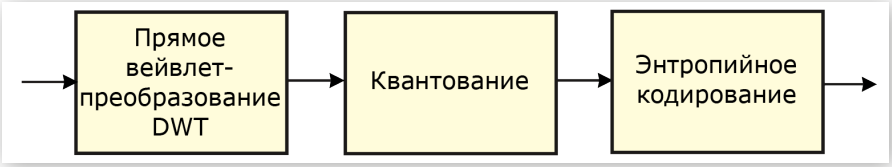

Рис. 7. Общая схема компрессии на основе вейвлет-преобразования

Как было отмечено выше, уже при использовании третьего уровня вейвлет-декомпозиции изображения бо**о**льшая часть матрицы отсчетов кадра занята частотными составляющими верхних пространственных частот. Квантование составляющих с высокими пространственными частотами, описывающих мелкие детали изображения, не приводит к заметным артефактам. Это обстоятельство позволяет добиваться значительных отношений компрессии при сохранении качества восстановленного изображения на уровне «визуально без потерь».

Восстановление исходного сигнала выполняется в обратном порядке с использованием обратных преобразований: декодирование, инверсное квантование и обратное вейвлет-преобразование. На стадии обратного вейвлет-преобразования каждый компонент преобразованного сигнала изображения сначала растягивается в два раза, то есть после каждого отсчета вставляется дополнительный нулевой. Растянутый компонент подвергается фильтрации, в результате которой на место нулевых отсчетов помещаются интерполированные величины. Эта процедура полностью соответствует схеме рис. 3, но выполняется для двух направлений пространственных частот – горизонтального и вертикального.

Вейвлет-преобразование не связано с формированием блоков, поэтому артефакты видеокомпрессии на его основе более «естественны», проще говоря, выглядят менее чужеродными на типичных изображениях, чем, например, блочная структура в виде просвечивающей через изображение шахматной доски, которая появляется в результате компрессии на базе DCT.

Принципиальное отличие вейвлет-компрессии от компрессии на базе дискретного косинусного преобразования DCT заключается в способе получения частотных компонентов изображения. DCT позволяет получать частотные компоненты, занимающие равные полосы при всех средних частотах (например, 1/8 от максимальной частоты сигнала). Вейвлет-преобразование дает компоненты, полосы частот которых уменьшаются в два раза по мере уменьшения средней частоты (например, 1/2, 1/4, 1/8 от максимальной частоты сигнала и т. д.). Такой способ получения частотных компонентов в большей мере соответствует особенностям субъективного восприятия.

Подробности технологических решений JPEG XS будут рассмотрены в следующей части.

Окончание следует