

# JPEG XS – еще полшага в IP-будущее

Продолжение. Начало в № 5/2023

Часть 2

Схема процесса кодирования JPEG XS

Константин Гласман

**С**тандарт JPEG XS, в отличие от предыдущих стандартов видеокompрессии, разработанных группой JPEG (формально – ISO/IEC SC29WG1), относится к сжатию видеосигналов [1]. Цель его разработки заключалась в создании простого «мезонинного» кодера для передачи цифрового видеосигнала по развернутым, например, в вещательных студиях, IP-сетям. Проблема, которую достижение поставленной цели позволяло бы решить, связана с тем, что четкость ТВ-изображения в студиях, оснащенных IP-сетями, растет с первоначального размера кадра Full HD (1920×1080) до 4K и даже до 8K. Но обновление существующей сетевой инфраструктуры в вещательных студиях сравнительно дорого. Простое сжатие видеосигнала с небольшой степенью компрессии без визуальных потерь позволит существенно сократить расходы по внедрению форматов 4K и 8K.

На первый взгляд кажется, что поставленная цель может быть легко достигнута с помощью существующих технологий. Например, JPEG 2000 и даже более старый стандарт, такой как JPEG, безусловно, обеспечивают сжатие без визуальных потерь с заданными степенью компрессии и скоростью цифрового потока. Однако здесь есть подводный камень – повторяющиеся циклы кодирования-декодирования при видеокompрессии по этим стандартам в нескольких точках внутростудийной сети увеличат задержку и внесут дополнительные искажения и артефакты. И это закономерно, потому что основной целью проектирования систем JPEG и JPEG 2000 было достижение максимальной степени компрессии при заданном уровне качества после одного цикла.

В список требований при разработке JPEG XS были добавлены очень важные дополнительные ограничения. Во-первых, максимальная сквозная задержка не должна превышать 32 строк от входа кодера при сжатии до

выхода декодера при распаковке, и, во-вторых, ухудшение качества при повторных циклах кодирования-декодирования не должно быть заметно при большом числе циклов (не менее 10). К тому же кодек JPEG XS должен быть сравнительно простым и допускать реализацию на разных платформах, включая графические процессоры компьютеров и программируемые пользователем вентильные матрицы [2-7].

Кодер JPEG XS является классическим устройством сжатия статичных изображений на основе вейвлет-преобразования. В схеме кодера можно увидеть прямое дискретное вейвлет-преобразование (DWT), квантование и энтропийное кодирование (рис. 8). Но упомянутые выше дополнительные ограничения приводят к значительным отличиям от других известных кодеков на основе DWT, таких, например, как JPEG 2000. Последовательность обрабатываемых изображений – это кадры ТВ-изображения, поэтому скорость цифрового потока видеоданных на выходе кодера должна оставаться постоянной и равной выбранному значению. Поэтому в схеме присутствует блок, отвечающий за управление скоростью потока компрессированных данных. Он управляет работой квантователя, в котором квантуются коэффициенты вейвлет-преобразования и благодаря этому уменьшается скорость потока данных. В схеме присутствуют также некоторые дополнительные преобразования (предварительная обработка, обратимое цветовое преобразование), которые заслуживают специального рассмотрения.

## Предварительная обработка

Дискретное вейвлет-преобразование Ле Галла 5/3 является обратимым. Это означает, что, выполняя прямое преобразование исходных данных в виде массива целых чисел и применяя к результату обратное преобразование,

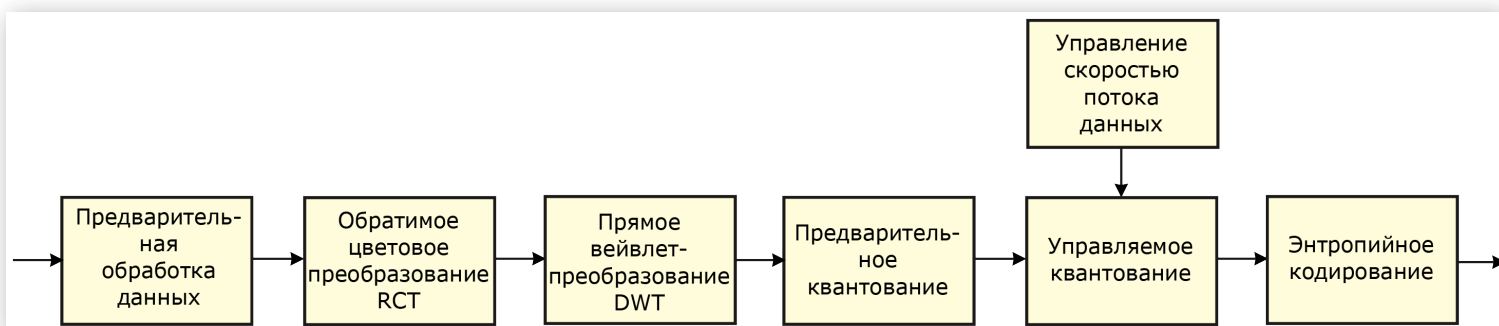


Рис. 8. Схема процесса кодирования JPEG XS

можно получить исходный массив. Но это справедливо только в том случае, если в системе цифровых фильтров нет округления результатов. Система становится нелинейной при округлении промежуточных результатов до целых чисел. Тогда при вычислениях накапливаются ошибки и дискретное вейвлет-преобразование перестает быть обратимым.

Кодер JPEG XS минимизирует нелинейность, увеличивая разрядность входных данных до 20 бит. Если входные данные представлены 8-битовыми числами, то увеличение разрядности может быть формально достигнуто умножением входных чисел на 212. Дальнейшие преобразования – обратимое цветовое преобразование и вейвлет-преобразование – будут выполняться с точностью, обеспечиваемой 20-битовыми числами. Это подразумевает однако, что декодер обязательно должен уменьшить разрядность выходных данных обратного преобразования с 20 битов до разрядности входных данных.

Вторая задача предварительной обработки – симметрирование номинального динамического диапазона входных данных относительно нуля. Если входные данные представлены P-битовыми числами без знака (что типично для трехкомпонентного цветного сигнала в формате RGB), то номинальный динамический диапазон определяется интервалом  $[0, 2^P-1]$ , и он явно не симметричен относительно нуля. Путем вычитания числа  $2^P-1$  из значений входных данных номинальный динамический диапазон трансформируется в интервал  $[-2^P-1, 2^P-1-1]$  и станет симметричным относительно нуля. Если, например,  $P = 8$ , то вычитание  $2^8 - 1 = 128$  из входных данных преобразует номинальный динамический диапазон из интервала  $[0, 255]$  в интервал  $[-128, 127]$ .

Если входные данные представлены числами со знаком (что типично для цветоразностных сигналов  $C_R$  и  $C_B$ ), то номинальный динамический диапазон уже симметричен относительно нуля и дополнительная обработка не требуется.

Симметрирование номинального динамического диапазона относительно нуля упрощает реализацию вейвлет-преобразования и преобразования цвета в арифметике с фиксированной запятой/точкой. Гарантируя, что номинальный динамический диапазон симметричен относительно нуля, можно сделать ряд упрощающих допущений при разработке кодека (например, в отношении контекстного моделирования, числового переполнения и т. д.).

Этап постобработки в декодере по существу отменяет эффекты предварительной обработки в кодере. Если значения компонента выходных данных не имеют знака, то восстанавливается исходный номинальный динамический диапазон.

Постобработка в декодере JPEG XS содержит один дополнительный этап, у которого нет эквивалента в кодере. Это ограничение уровней выходного сигнала сверху и снизу, позволяющее гарантировать, что значения выходного сигнала не выходят за пределы допустимого диапазона. Это необходимо, поскольку вейвлет-коэффициенты, рассчитываемые в декодере, не обязательно идентичны коэффициентам, определенным в кодере, и, следовательно, восстановленные в декодере значения видеосигнала могут выйти за пределы допустимого диапазона данных.

## Обратимое цветовое преобразование RCT

В кодере JPEG XS за стадией предварительной обработки следует обратимое цветовое преобразование RCT (Reversible Color Transform). Оно не является обязательным и используется в том случае, если на вход подан трехкомпонентный цифровой сигнал в формате RGB. Необходимость преобразования RCT в этом случае определяется следующими причинами. Компоненты R, G и B цветного телевизионного сигнала очень сильно коррелированы друг с другом, что делает избыточным поток цифровых данных из этих трех компонентов в исходной форме. RCT является межкомпонентным преобразованием, которое служит для уменьшения корреляции между компонентами, обеспечивая более эффективное кодирование. После преобразования компоненты сигнала обрабатываются независимо.

Фактически RCT – это преобразование цветового пространства RGB в другое цветовое пространство, в котором компоненты ТВ-сигнала коррелированы друг с другом в меньшей степени. Хорошо известно цветовое пространство  $YC_R C_B$ , определяемое Рекомендацией ITU BT.601. Преобразование в соответствии с Рекомендацией ITU BT.601 трансформирует компоненты телевизионного сигнала RGB в яркостный сигнал Y и два цветоразностных:  $C_B$  и  $C_R$ . В контексте вычислительных процедур в кодеках компрессии это преобразование является «необратимым» (обозначается как ICT – Irreversible Color Transform) в том смысле, что абсолютно точное восстановление исходных сигналов RGB в цикле «прямое преобразование и обратное преобразование  $YC_R C_B$  в RGB» невозможно, потому что неизбежное на практике округление результатов промежуточных вычислений вызывает накопление ошибок.

Преобразование RCT является целочисленным приближением к преобразованию ICT. При его вычислении значения компонентов сигнала умножаются или делятся на числа 2 в целой степени (такие операции просто реализуются в виде сдвига двоичных чисел в разрядной сетке в сторону старших или младших разрядов соответственно на величину, равную показателю степени числа 2), а потом складываются или вычитаются. Преобразование RCT обратимо в том смысле, что в цикле «прямое RCT – обратное RCT» исходные числа преобразуются в те же числа, если есть некоторый «запас» в разрядной сетке по отношению к разрядности исходных чисел. В кодере JPEG XS «запас» в разрядности создается в блоке предварительной обработки, описанном выше.

Прямое преобразование RCT определяется простыми формулами:

$$\ddot{Y} = \frac{R+2G+B}{4}, \quad \ddot{C}_R = R - G, \quad \ddot{C}_B = B - G,$$

а обратное – такими же простыми формулами:

$$G = \ddot{Y} - \frac{\ddot{C}_R + \ddot{C}_B}{4}, \quad R = \ddot{C}_R + G, \quad B = \ddot{C}_B + G.$$

Обозначения  $\ddot{Y}$ ,  $\ddot{C}_R$  и  $\ddot{C}_B$  в формулах фокусируют внимание на том, что эти величины не соответствуют компонентам, определяемым Рекомендацией ITU BT.601, а дают только целочисленное приближение к ним.

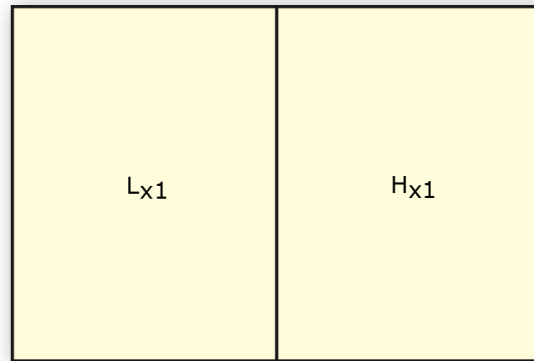
Как было отмечено выше, обратимое цветовое преобразование RCT используется в том случае, когда на вход подан трехкомпонентный цифровой сигнал в формате RGB. Если компоненты входного телевизионного сигнала представляют собой яркостный сигнал  $Y$  и два цветоразностных ( $C_B$  и  $C_R$ ), то необходимость в преобразовании RCT отпадает.

### Вейвлет-преобразование

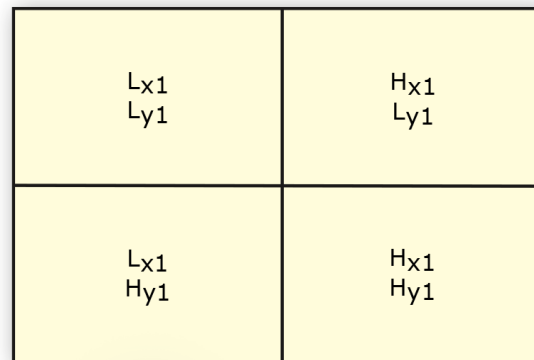
В кодеке JPEG XS используется асимметричное вейвлет-преобразование с одним или двумя уровнями декомпозиции по вертикали и пятью уровнями декомпозиции по горизонтали. Один уровень декомпозиции по вертикали сводит к минимуму количество строк, которые необходимо буферизовать для выполнения вейвлет-преобразования, и, следовательно, приводит к минимуму задержку. Количество декомпозиций по горизонтали практически не влияет на задержку. Расчет показывает, что один уровень вертикальной декомпозиции, основанной на вейвлет-преобразовании Ле Галла 5/3 (оно описано в первой части статьи), вызывает сквозную задержку на 3 строки, а два уровня – на 9 строк [4]. Допустимая задержка больше, поэтому кажется, что можно было бы использовать большее число уровней декомпозиции по вертикали. что привело бы к увеличению эффективности компрессии. Однако работа блока управления скоростью потока компрессированных данных также требует задержки, поскольку блоку надо «заглядывать» на несколько строк вперед для определения оптимальной стратегии работы квантователя.

Асимметричное вейвлет-преобразование с одним уровнем декомпозиции по вертикали и тремя уровнями декомпозиции по горизонтали иллюстрирует рис. 9. На рис. 9-а показан результат вейвлет-фильтрации и прореживания в горизонтальном направлении на первом уровне декомпозиции. В результате получены субматрицы коэффициентов  $Lx1$  ( $L$  символизирует низкочастотную фильтрацию,  $x$  – фильтрацию в горизонтальном направлении, 1 – первый уровень декомпозиции) и  $Hx1$  ( $H$  обозначает здесь фильтрацию верхних частот). На рис. 9-б показаны четыре субматрицы:  $Lx1Ly1$ ,  $Hx1Ly1$ ,  $Lx1Hy1$ ,  $Hx1Hy1$ , полученные в результате применения низкочастотной и высокочастотной вейвлет-фильтрации и прореживания в вертикальном направлении на первом уровне декомпозиции к субматрицам  $Lx1$  и  $Hx1$ . Затем субматрица  $Lx1Ly1$  нижних пространственных частот горизонтального и вертикального направлений разлагается на низкочастотные и высокочастотные составляющие горизонтального направления еще раз на втором уровне декомпозиции (рис. 9-в). Рис. 9-г иллюстрирует результат применения к субматрице  $Lx1Ly1Lx2$  (рис. 9-в) вейвлет-преобразования в горизонтальном направлении на третьем уровне декомпозиции.

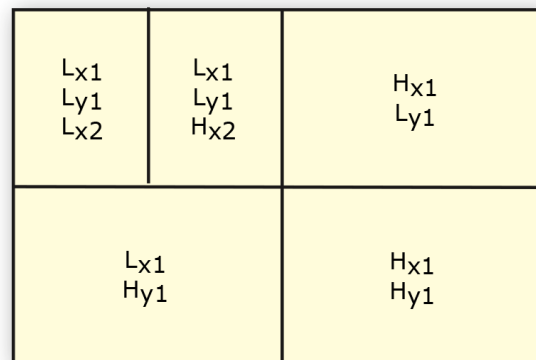
На рис. 9-г показан набор из 6 субматриц коэффициентов асимметричного двумерного вейвлет-преобразования телевизионного изображения (они перенумерованы для использования этих номеров в дальнейшем изложении), причем каждая субматрица – массив коэффициентов вейвлет-преобразования с разными параметрами частотных субполос горизонтального и



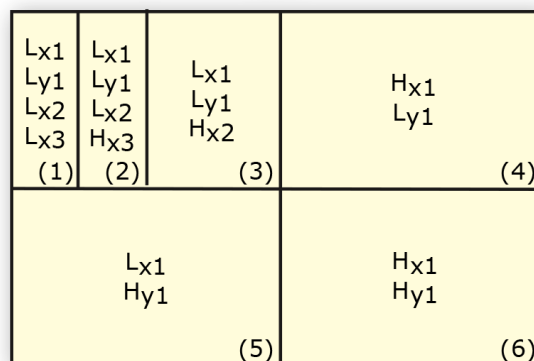
(а)



(б)



(в)



(г)

Рис. 9. Асимметричная вейвлет-декомпозиция изображения:  
 а – декомпозиция первого уровня в горизонтальном направлении;  
 б – декомпозиция первого уровня в вертикальном направлении;  
 в – декомпозиция второго уровня в горизонтальном направлении;  
 г – декомпозиция третьего уровня в горизонтальном направлении

вертикального направлений. Например, субматрица Lx1Ly1Lx2 Lx3 (она имеет номер 1) описывает низкочастотные компоненты изображения, причем в горизонтальном направлении ширина их спектра равна 1/8 от максимальной частоты, а в вертикальном – 1/2 от максимальной частоты спектра изображения. Описанному вейвлет-преобразованию независимо подвергаются все три компонента телевизионного сигнала: яркостный сигнал и два цветоразностных.

### Квантование как способ сокращения скорости потока данных

Операции, выполняемые в кодере до и после квантования, являются обратимыми и не вносят искажений, о чем было сказано выше при обсуждении цветового преобразования и вейвлет-преобразования. Квантователь представляет собой нелинейное устройство, в котором с передаваемыми данными выполняются необратимые преобразования. Результатом квантования является уменьшение объема передаваемых данных в единицу времени, то есть скорости потока цифровых данных. Но одновременно в передаваемый сигнал вносятся искажения. Задача квантователя вместе с остальными блоками кодера JPEG XS заключается в таком уменьшении скорости потока передаваемых данных, при котором будет достигнуто необходимое отношение компрессии, но при условии, что искажения будут визуально незаметными, то есть качество будет на уровне «без визуальных потерь» даже после большого числа циклов кодирования-декодирования.

Принцип квантования с целью уменьшения объема и скорости потока передаваемых данных иллюстрирует рис.10. Первая строка таблицы на рисунке – числа, представляющие собой набор четырех коэффициентов вейвлет-преобразования. Они записаны в столбцах таблицы как двоичные числа в формате «знак-величина». Первый бит – знак (знак плюс отображается в виде 0, минус – 1), остальные семь битов – абсолютная величина. Например, число +9 записывается как 00001001, число -26 – как 10011010. Старший разряд следует сразу за знаковым разрядом, в последней нижней строке таблицы находится младший разряд. Номер младшего разряда обычно принимается равным 0. Номер знакового разряда в данном примере равен 7.

Что можно сделать для уменьшения объема передаваемых данных? Можно не передавать нулевые биты до первой значащей единицы. Но для того, чтобы упростить отделение одного числа от другого при приеме, можно, например, сделать длину всех кодовых слов при передаче одинаковой и число отбрасываемых незначащих нулевых битов установить общим для всей группы чисел, зафиксировав позицию MSB (Most Significant Bit – наибольший значащий бит) как наименьшего номера разряда, в котором находится хотя бы одна значащая единица в числах группы. В данном примере можно отбросить 2 разряда (они выделены белым цветом). Но надо иметь в виду, что в этом случае есть накладные расходы – надо передать декодеру номер позиции MSB. Кроме того, если кодируемые числа велики и старшие разряды являются значащими, то есть содержат единицы, то выигрыш в сокращении объема данных будет невелик. Достоинством этого способа является то, что он обратим, – декодер может полностью восстановить передаваемые числа.

Радикальным способом уменьшения объема данных является отбрасывание некоторого количества младших разрядов. В примере рис. 10 позиция отсечки устанавливается на уровне разряда номер 0 и отбрасывается (отсекается) один младший разряд в двоичном коде каждого коэффициента (нижняя строка с отсекаемыми битами выделена красным цветом). Таким образом, в этом примере передаются 5 битов вместо 8 – 1 бит знака (выделен голубым цветом) и 4 бита величины (выделены зеленым цветом).

Число отсекаемых битов должно быть известно декодеру, тогда он сможет восстанавливать отправленные числа, но это восстановление будет приближенным из-за нелинейного характера процедуры отсечки. Вместо числа +9 после отсечки было передано число +4, вместо -26 – число -13 и т. д. Декодер, зная число отсеченных разрядов, сдвинет числа в сторону старших разрядов на одну позицию, что эквивалентно умножению принятой величины на 2. В результате вместо кодируемых чисел +9 и -15 будут восстановлены при декодировании числа +8 и -14, но числа -26 и +6 будут восстановлены точно. Все нечетные числа восстанавливаются с ошибкой, все четные – точно. Это и есть квантование с шагом 2. Числа 8 и 9 декодируются как 8, 14, а 15 – как 14 и т. д. и т. п.

Коэффициент	Знак	+9	-26	-15	6	Позиция MSB
		0	1	1	0	
Величина		0	0	0	0	Квантованная величина
		0	0	0	0	
		0	1	0	0	
		1	1	1	0	
		0	0	1	1	
		0	1	1	1	
		1	0	1	0	Позиция отсечки

Рис. 10. Квантование как способ сокращения объема данных

Свойства квантователя удобно иллюстрировать с помощью передаточной характеристики – зависимости уровня выходного сигнала от входного (рис. 11). В кодере JPEG XS доступны два режима квантования – равномерное квантование (рис. 11-а) и квантование с мертвой зоной (рис. 11-б). Передаточная характеристика равномерного квантователя представляет собой «лестницу» с постоянной шириной ступеньки, называемой шагом квантования и обозначаемой в описании стандарта JPEG XS как Δ. Если считать, что отметки горизонтальной оси на рис. 11-а следуют через единичное значение входного сигнала, то шаг квантования равен четырем единицам. Через каждые четыре единицы входного сигнала

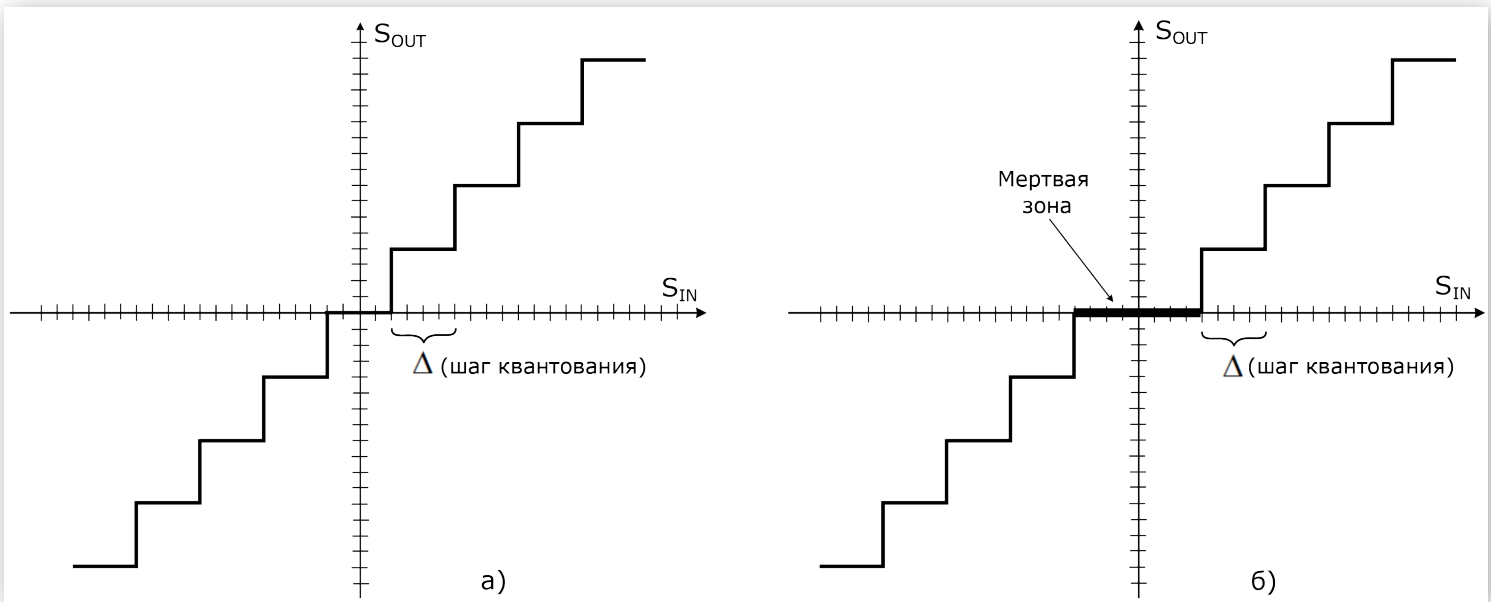


Рис. 11. Передаточные характеристики квантователей: а – равномерное квантование; б – квантование с мертвой зоной

ла значение выходного увеличивается скачком. Такое квантование также обычно используется в аналого-цифровых преобразователях телевизионного сигнала.

Квантование на рис. 11-б отличается наличием мертвой зоны в окрестности нулевого значения входного сигнала. В пределах мертвой зоны квантователь никак не откликается на изменения входного сигнала. Отсечка младшего разряда, описанная выше, является квантованием с мертвой зоной, причем ширина мертвой зоны равна удвоенному шагу квантования. Размер шага квантования связан с числом отброшенных двоичных разрядов  $T$  зависимостью:

$$\dots \Delta_T = 2^T.$$

Если опять считать, что отметки горизонтальной оси на рис. 11-б следуют через единичное значение входного сигнала, то шаг квантования равен четырем единицам. Передаточная характеристика такого квантователя описывает отсечку двух младших двоичных разрядов входного слова данных.

Окончание следует

## НОВОСТИ

### Zeiss покупает Ncam Technologies

Компания Zeiss приобрела одного из пионеров в сфере систем трекинга камер – Ncam Technologies. В результате этой сделки Zeiss расширяет свой технологический портфель специально для сфер кинематографа, вещания и визуальных эффектов. Гибридная технология трекинга камер Ncam, получившая большое число отраслевых наград, позволяет отслеживать положение камеры в пространстве, причем как в помещениях, так и под открытым небом, для чего применяются разные методы трекинга. Получаемые от системы трекинга данные важны для виртуального производства и композитинга в режиме реального времени, они позволяют существенно повысить эффективность процессора монтажа и обработки. Эта технология отлично дополняет линейку кинематографической продукции Zeiss, включая объективы Supreme Prime и Supreme Prime Radiance, Cinema Zoom и CP.3, а также недавно анонсированные сервисы обработки данных в рамках экосистемы CinCraft. Общая цель Zeiss и Ncam заключается в обеспечении высококачественных, простых в применении систем трекинга и визуальных эффектов для профессиональных рабочих процессов съемки.

«Мы рады, что можем объединить эффективную технологию трекинга Ncam с многолетним опытом Zeiss на рын-

ке кинематографических объективов и в работе с данными от объективов, – сказал ответственный за кинопродукцию Zeiss Кристоф Казенав. – Это позволяет нам выйти за пределы текущих возможностей трекинга камеры и предложить инновационные решения специально для визуальных эффектов, виртуального производства и других приложений».

Глава департамента исследований и разработок Ncam Брайс Мишо добавил: «Коллектив рад тому, что будет работать с новыми коллегами из Zeiss на следующем этапе совершенствования технологии трекинга, делая ее доступной большее широкому кругу пользователей, чем когда-либо ранее».

Компания Ncam была основана в 2012 году, она имеет офисы в Лондоне и Лос-Анджелесе. Слияние глобальных ресурсов Ncam и Zeiss позволит улучшить поддержку пользователей и добиться впечатляющих инноваций не только в кинематографе и вещании, но и за пределами этих отраслей.

В планах Zeiss – гарантировать, что каждая новая функция и каждое улучшение, привнесенные в технологию, станут доступны и для нынешних пользователей Ncam, так что компания собирается предложить им привлекательную программу модернизации. Первое объявление в этом смысле ожидается ближе к концу лета.