

## ОСОБЕННОСТИ УСТРАНЕНИЯ МЕЖКАДРОВОЙ ИЗБЫТОЧНОСТИ В СТАНДАРТЕ MPEG-4/H.264

Стандарт видеокompрессии MPEG-4 Part 10 / H.264 был принят в конце 2003 года как логическое усовершенствование стандартов MPEG-4 и H.263. Стандарт H.264 также включает в себя множество усовершенствований по сравнению со стандартом compрессии MPEG-2 [3].

В рекомендации H.264, как и в предшествующих стандартах, отсутствует явное описание передающей и приёмной сторон, т.е. пары кодер-декодер, а имеется явное описание синтаксиса кодированного битового видеопотока совместно с методом его декодирования. Также, как и в предыдущих стандартах, H.264 включает в себя такие основные элементы, как прогноз, преобразование, квантование и энтропийное кодирование, но существуют также и элементы, присущие лишь этому стандарту, например деблокирующий фильтр и др. Основное отличие MPEG-4 от более ранних стандартов - это объектно-ориентированное представление медиainформации. В стандарте любая сцена разделяется на объекты, которые могут быть описаны отдельными элементарными потоками, соотнесенными во времени и пространстве.

Можно выделить несколько основных сфер применения стандарта видеокodирования H.264:

- DVD-диски с ТВЧ контентом (Blu-Ray),
- ТВЧ телевизионное вещание в Европе,
- Мобильное ТВ вещание,
- Мобильное и Интернет потоковое видео,
- Видеоконференции,
- Бытовые камкордеры.

В некоторых случаях предсказание по интерполированным позициям сэмплов в опорном кадре может дать лучшее предсказание компенсации движения. Пиксели опорной области интерполируются на полупиксельные позиции, и становится возможным найти лучшее совпадение для текущего макроблока при поиске по интерполированным сэмплам. Субпиксельная оценка и компенсация движения включает поиск позиций интерполированных субпикселей, также как и поиск по целым позициям пикселей и выбор позиции, дающей лучшее совпадение и минимизирующей энергию остатка. Рисунок 3 показывает четвертьпиксельную оценку движения. На первом этапе оценка движения находит лучшее совпадение на сетке целопиксельных значений (круги). Далее кодер ищет лучшее полупиксельное совпадение (квадраты), чтобы проверить, может ли совпадение быть улучшено, и если требуется, далее выполняется поиск по четвертьпиксельным позициям (треугольники). Окончательное совпадение, цело-, полу- или четвертьпиксельное, вычитается из текущего блока или макроблока.



Рис. 3. Цело-, полу- и четвертьпиксельная оценка движения.

В общем случае, «лучшая» интерполяция дает лучшее качество компенсации движения, дающее меньший остаток ценой увеличенной сложности [3]. Выигрыш в качестве убывает с ростом шагов интерполяции. Полупиксельная интерполяция дает значительный выигрыш над целопиксельной, четвертьпиксельная дает среднее дальнейшее улучшение, восьмикратная интерполяция дает маленькое дальнейшее улучшение и т.д.

Поиск подходящих блоков 4x4 при четвертьпиксельной интерполяции более сложно, чем поиск блоков 16x16 без интерполяции. В дополнение к вычислительной сложности, существуют ошибки кодирования, поскольку вектор каждого блока должен кодироваться и передаваться приёмнику с целью корректного восстановления изображения. С уменьшением размера блока количество векторов растёт. Больше бит требуется для представления полу- и четвертьпиксельных векторов, поскольку дробная часть вектора, т.е. 0,25 или 0,5, должна кодироваться, также как и целая часть.

Каждый P- макроблок соотносится с определённым разбиением макроблока на блочные разделы, которые уже используются для предсказания компенсации движения. Поддерживаются разделы с размерами блоков яркости 16x16, 16x8, 8x16 и 8x8 отсчётов. В случае если выбраны разделы с размерами 8x8 отсчётов, передаётся один синтаксический элемент для каждого блока 8x8. Этот синтаксический элемент определяет, будет ли соответствующий раздел 8x8 далее делиться на подразделы 8x4, 4x8 или 4x4 отсчётов яркости и соответствующие подразделы по направлению цветности. Рис. 4 показывает такое разбиение.

Сигнал предсказания для каждого блока размером MxN кодируемого с предсказанием получается при помощи перемещения области соответствующего опорного изображения, которое определено поступательным вектором движения и индексом опорного изображения. Таким образом, если макроблок кодируется с использованием четырёх разделов 8x8 и каждый из них далее делится на четыре раздела 4x4, для одного P-макроблока могут быть переданы максимальное значение в 16 векторов движения.

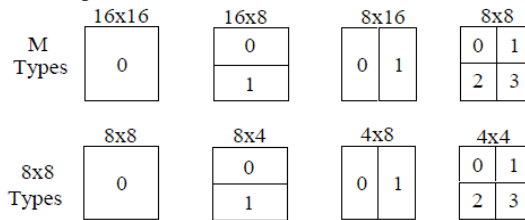


Рис. 4. Разбиение макроблока для компенсации движения. Сверху: разбиение макроблоков, снизу: разбиение 8x8 разделов.

Точность компенсации движения определяется в единицах до одной четвёртой от расстояния между отсчётами яркости.

H.264 предлагает большое количество различных опций для выбора опорных изображений для inter предсказания. Кодер обычно использует опорные изображения структурированным образом. Далее приведём несколько примеров структур опорных изображений.

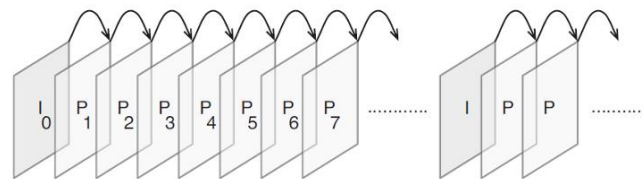


Рис. 5. Структура предсказания с малой задержкой.

Структура предсказания, показанная на рис. 5, использует только I и P кадры. Она совместима с Baseline Profile и Constrained Baseline Profile H.264 [4], которые не допускают B кадров, и будет применимой для применения, требующего малой задержки и/или минимальной памяти декодера. Первый кадр кодируется как I кадр, и последующие кадры кодируются как P кадры, предсказанные по предыдущему кадру. Эффективность предсказания относительно мала, поскольку для каждого кадра предусмотрены только одно направление предсказания и один опорный кадр. Тем не менее, декодер может отобразить каждый кадр, как только он декодирован, с минимальной задержкой, и декодеру нужно запоминать только один опорный кадр, что минимизирует требования к памяти. Этот тип структуры будет подходить для таких приложений, как видеоконференции, где задержка должна оставаться минимальной. Заметим, что кадры I могут добавляться к потоку с определенным интервалом, чтобы ограничить распространение ошибок передачи и разрешить произвольный доступ к закодированной последовательности.

Ранние стандарты, такие как MPEG-1 и MPEG-2, использовали группу структур изображений, показанную на рис. 6. Группа изображений (GOP, Group of Pictures) начинается с I кадра. P кадры добавляются с определенными интервалами, с B кадрами между I и P кадрами. Кадры I и P используются как опорные; кадры B не используются как опорные, т.е. все изображения предсказываются по кадрам I и P. Каждый кадр P предсказывается по предыдущему кадру I или P, и каждый кадр B предсказывается по I и/или P кадрам с любой стороны от него.

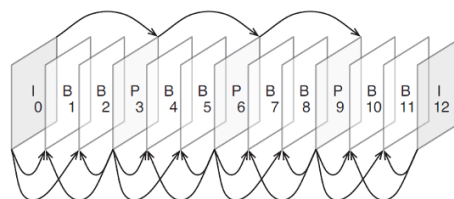


Рис. 6. «Классическая» структура предсказания группы изображений

Эта структура дает лучшую эффективность сжатия, чем структура, показанная на рис. 5, поскольку предсказание кадра В обычно более эффективно, чем предсказание кадра Р, вследствие того, что доступны два опорных изображения. Недостатком является увеличенная задержка и большие требования к памяти для хранения кадров..

Н.264 делает возможным создание других структур предсказания, к примеру использование для предсказания множественных опорных кадров. Рис. 7 показывает структуру предсказания IPPP..., в которой все ранее закодированные кадры доступны как опорные. Кадр P1 предсказывается по кадру I0; кадр P2 предсказывается по кадрам I0 и P1; кадр P3 предсказывается по кадрам I0, P1 и P2 и т.д. Это означает, что кодер может проверять до N опорных кадров для поиска лучшего совпадения для каждого Р макроблока. Это может увеличить эффективность сжатия за счёт: а) увеличения вычислительной сложности кодера и б) увеличения памяти кодера и декодера, поскольку должны сохраняться N опорных кадров.

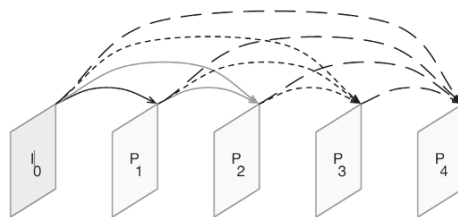


Рис. 7. IPPP с множественными опорными изображениями

Рисунок 8 показывает пример иерархической структуры GOP [5]. Этот тип структуры становится возможным благодаря гибким опциям предсказаний Н.264. GOP начинается и заканчивается I кадрами 0 и 12. Далее, кадр В6 предсказывается с использованием I0 и I12 как опорных; заметим, что В6 находится посередине между I0 и I12 в порядке отображения. В3 предсказывается по I0 и В6; В9 предсказывается по В6 и I12. В1 и В2 предсказываются по I0 и В3 и т.д. GOP составляется из серий слоёв, двухэлементным или пирамидальным разложением. Этот тип структуры может дать улучшенное качество сжатия, если параметры квантования слоёв тщательно контролируются таким образом, что параметр квантования QP растёт с увеличением количества слоёв. Иерархические или пирамидальные структуры GOP были впервые предложены как метод, дающий временную масштабируемость, но также они могут использоваться в любом Н.264 кодеке, поддерживающем В кадры.

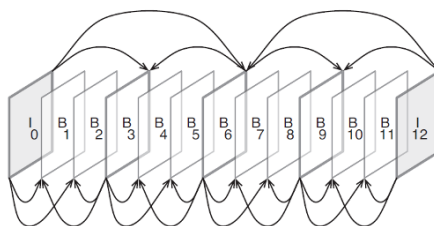


Рис. 8. Иерархическая структура группы изображений - GOP

Новый стандарт видеокodирования Н.264/AVC предоставляет значительные преимущества в области улучшения эффективности сжатия и гибкости для эффективного использования в широком диапазоне типов сетей и областей применения.

#### Библиографический список

1. Р.Гонсалес, Р.Вудс Цифровая обработка изображений. –М.: Техносфера, 2006.
2. Ян Ричардсон Видеокodирование. H.264 и MPEG-4 – стандарты нового поколения. –М.: Техносфера, 2005.
3. ITU-T and ISO/IEC JTC 1, "Advanced video coding for generic audiovisual services," ITU-T Rec. H.264 and ISO/IEC 14496-10 AVC, 2003.
4. Gary J. Sullivan, Thomas Wiegand Video Compression - From Concepts to the H.264/AVC Standard. // Proc. of the IEEE, vol.93, pages: 18-31, 2004
5. Быков Р.Е., Фрайер Р., Иванов К.В., Манцветов А.А. Цифровое преобразование изображений. –М: Горячая линия-Телеком, 2003. -228 с.

**Е.О. Korzhikhin**

#### FEATURES ELIMINATION OF INTERFRAME REDUNDANCY IN THE STANDARD MPEG-4/H.264

H.264/AVC video coding level is based on a hybrid concept of the traditional block coding with motion compensation, but with some important differences from the previous standards, which gives about 50% reduction in bit rate for the same visual quality as compared to previous standards.