Прикладная теория кодирования и сжатия информации

№ 4(26)

DOI 10.17223/20710410/26/9

УДК 004.627

2014

ЭКСПЕРИМЕНТАЛЬНЫЙ АНАЛИЗ ВНУТРИКАДРОВОГО ПРЕДСКАЗАНИЯ В H.265/HEVC¹

Р. И. Черняк

Национальный исследовательский Томский государственный университет, Томский государственный университет систем управления и радиоэлектроники, г. Томск. Россия

E-mail: roman.chernyak@elecard.ru

Рассмотрена одна из двух основных составляющих сжатия видеоинформации в новом стандарте H.265/HEVC — внутрикадровое кодирование. Проведена серия экспериментов, результаты которых позволяют оценить практическую эффективность используемых в HEVC решений.

Ключевые слова: сжатие видео, внутрикадровое предсказание, H.265/HEVC, кодирование режимов предсказания.

Введение

В связи с бурным ростом телекоммуникаций в современном мире всё более актуальной становится задача компактного представления информации; особенно остро она стоит в видеоиндустрии. Согласно статистике, доля видеоконтента составила 51% всего мобильного трафика в 2012 г. и будет увеличиваться в дальнейшем [1]. В связи с этим в апреле 2013 г. группой экспертов по видеокодированию ITU-T Video Coding Experts Group совместно с экспертной группой по движущимся изображениям Moving Picture Experts Group предложен новый стандарт сжатия видеоданных H.265/HEVC [2]. В стандарт включено множество алгоритмических улучшений, которые позволили добиться существенного увеличения степени сжатия при прежнем качестве. В работе [3] приведён подробный анализ преимуществ нового стандарта H.265 по сравнению с уже устаревшими на сегодня стандартами H.263 [4] и MPEG-4 [5] и с текущим индустриальным стандартом H.264/AVC [6].

1. Краткое описание H.265/HEVC

В рамках данного стандарта кодирование видеоданных проходит последовательно по отдельным кадрам. Каждый кадр разбивается на блоки, называемые Coding Units — блоками кодирования; последующее кодирование происходит поблочно. При кодировании каждого блока кодер выполняет процедуру предсказания этого блока. Под предсказанием понимается нахождение блока, наиболее похожего на данный. В зависимости от настроек кодер может осуществлять межкадровое (интер-) или внутрикадровое (интра-) предсказание. В первом случае происходит поиск похожего участка в соседних кадрах, во втором — используется специальная процедура построения блока в рамках текущего кадра. Наличие двух способов поиска предсказанного блока обусловлено двумя типами избыточности в видеоматериале: временной и пространственной. Для устранения временной избыточности используется механизм межкадрового предсказания,

 $^{^{1}}$ Работа выполнена в рамках комплексного проекта «Предоставление услуг мультимедийного вещания в сетях общего пользования Интернет, основанных на технологиях пиринговых сетей и адаптивной передачи потоков данных» при финансовой поддержке Министерства образования и науки Российской Федерации.

пространственной — внутрикадрового. При кодировании для каждого кадра определяется тип допустимых в нём предсказаний. Всего возможны три вида кадров: I (Intra), P (Predicted) и B (Bidirectional). Для I-кадра возможны только интрапредсказания, для P и B— как интра, так и интер. Различие между P- и B-кадрами состоит в том, что P-кадры могут быть предсказаны только от хронологически предшествующих им кадров, а B-кадры — как от предшествующих, так и от следующих за ними. Для обеспечения такой возможности входящие кадры внутри кодера переупорядочиваются.

HEVC допускает 35 различных режимов внутрикадрового предсказания, среди которых два «плоских» — 0 (INTRA_PLANAR) и 1 (INTRA_DC) и 33 угловых — $2, \ldots, 34$ (INTRA_ANGULAR2, ..., INTRA_ANGULAR34).

2. Эксперименты

Проведена серия экспериментов на входных видеоданных различных длительностей и разрешений с тем, чтобы собрать статистические данные о частоте применения тех или иных режимов интрапредсказаний. Помимо этого, собраны статистики «попадания» режима в массив MPM (Most Probable Modes) наиболее вероятных режимов из трёх элементов MPM[0], MPM[1], MPM[2]. Эксперименты проводились на видеоданных, закодированных эталонным кодером, свободно доступном на ресурсе [7]. В качестве исходного материала использовались некоторые из видеопоследовательностей, рекомендованных экспертными группами VCEG и MPEG для тестирования утилит видеокодирования. Согласно [8], все тестовые данные разделены на классы в зависимости от их разрешений и характера представленного контента, где классам А–D соответствуют снятые на видеокамеру сцены «реальной жизни» в разрешениях от WQXGA до WQVGA, а классу Е—видеоконференции с разрешением HD.

Видеоданные кодировались со следующими стандартными конфигурационными файлами: intra_main, lowdelay_main, lowdelay_P_main, randomaccess_main.

Конфигурация intra_main предполагает для всех кадров последовательности кодирование типа интра. Конфигурации lowdelay_main и lowdelay_P_main предполагают кодирование типа интра только для первого кадра последовательности, а остальных — в режимах P или B. Конфигурация randomaccess_main предполагает использование периодических последовательностей из I- и B-кадров вида IBB...BI.

Тестовое видео Traffic относится к классу A — материалов с наиболее высоким разрешением. Для него характерно интенсивное движение объектов на неподвижном фоне.

Для данной последовательности распределения режимов внутрикадрового предсказания меняются несущественно при разных конфигурациях кодирования. Двумя наиболее частыми режимами, вне зависимости от конфигурации, стали режимы 0 и 1, следом за ними, чередуясь, с близкими результатами идут режимы 10 и 26.

Рассматривая статистики попадания режимов интрапредсказания в массив МРМ для видеопоследовательности Traffic, можно сделать следующие выводы. Во всех четырёх случаях наиболее часты ситуации «МРМ[0]» и «вне МРМ». При этом самый благополучный с точки зрения эффективности кодирования случай имел место на конфигурации intra_main — доля нулевого элемента в МРМ при этом максимальна.

2.2. People On Street

Видеопоследовательность People On Street также относится к классу A и имеет разрешение 2560×1600 пикселей. Для неё, как и для последовательности Traffic, характерно интенсивное движение объектов на неподвижном фоне.

Для данной видеопоследовательности проявилась бо́льшая зависимость распределения режимов в зависимости от конфигурации кодирования. Во всех случаях тремя наиболее применимыми стали режимы 0,1 и 26 в порядке убывания частоты. Следующие за ними режимы встречаются существенно реже, и их порядок меняется в зависимости от конфигурации.

Во всех конфигурациях наиболее частой стала ситуация «вне MPM». При этом на конфигурациях lowdelay_main и lowdelay_P_main её доля близка к половине всех случаев. Далее, с большим отставанием, идёт ситуация попадания режима в MPM[0].

2.3. Kimono

Видеопоследовательность Kimono относится к классу В и имеет разрешение FullHD (1920 × 1080 пикселей). Для нее характерно движение как объекта, так и видеокамеры вслед за объектом, что влечёт изменение фона. Интенсивность движения в обоих случаях умеренная. Результаты эксперимента показывают, что для данного видеоматериала характерно большое количество нулевого режима интрапредсказания во всех конфигурациях. Следом за ним неизменно идут режимы 1 и 26. Частоты остальных режимов выражены менее существенно и незначительно изменяются при изменении конфигурации.

Степень попадания режимов в нулевой элемент массива MPM для данного видео довольно высока — около $40\,\%$; непопадание режима в MPM относительно редко — около $30\,\%$.

2.4. Cactus

Видеопоследовательность Cactus также имеет разрешение FullHD и относится к тестовому классу В. Для неё характерно умеренное движение объектов при неподвижном фоне. Как и прежде, режимы 0, 1 и 26 наиболее применимы для всех конфигураций. Частоты появления других режимов существенно меньше и незначительно изменяются при разных конфигурациях.

Для данного видео частоты наиболее и наименее благоприятных случаев попадания режимов в MPM примерно равны — около $34\,\%$. При изменении конфигураций они изменяются несущественно.

2.5. Basketball Drill

Данная видеопоследовательность имеет разрешение 832×480 пикселей и относится к классу С. Для неё характерно интенсивное движение на неподвижном однородном фоне с ярко выраженными угловыми текстурами.

Помимо типичных INTRA_PLANAR и INTRA_DC, для данного видео характерно большое количество режимов 18 и 19 во всех конфигурациях. Это объясняется характером материала— неподвижный фон может быть хорошо предсказан в направлениях 18 и 19. В результате доля вертикального режима 26 меньше, чем доля угловых режимов 18—20.

С точки зрения попаданий режимов в массив MPM для данного материала имеет место различие между конфигурацией intra_main и остальными. В первом случае доли попадания и непопадания в нулевой элемент массива MPM примерно равны. Во втором—ситуация «вне MPM» существенно более частая. Такое различие объясня-

ется тем, что в случае межкадровых предсказаний доля интрарежимов относительно невелика. Иными словами, кодер чаще принимает решение закодировать тот или иной блок, используя межкадровые, а не внутрикадровые связи. В случае недоступности соседнего блока в MPM попадут наиболее вероятные в общем случае режимы — 0, 1 и 26, которые, согласно результатам эксперимента, являются неоптимальными для данной видеопоследовательности.

2.6. Blowing Bubbles

Видеопоследовательность Blowing Bubbles относится к классу D и имеет разрешение 416×240 пикселей. Она содержит интенсивное движение объектов и умеренное движение фона; для неё характерно типичное распределение режимов внутрикадрового предсказания— режимы 0, 1 и 26 являются наиболее частыми, существенно опережая остальные.

Анализируя статистики попадания режимов в MPM, можно увидеть, что во всех конфигурациях наиболее частым сценарием является непопадание режима в MPM. Его доля близка к половине всех случаев. Следом с большим отставанием идёт сценарий попадания режима в MPM[0].

2.7. Four People

Видеопоследовательность Four People имеет размеры 1280×720 , относится к классу E и представляет собой видеоконференцию с участием четырёх человек; характеризуется умеренным движением на неподвижном фоне. На всех конфигурациях тройка наиболее частых режимов выглядит одинаково — 0, 1, 26.

Несмотря на то, что статистики распределения режимов внутрикадрового предсказания на данном видео меняются несущественно при изменении конфигураций кодирования, данные о попадании режимов в массив MPM различаются. Наиболее благоприятная ситуация — попадание режима в MPM[0] — происходит наиболее часто при конфигурации intra_main. Её доля в этом случае составляет около 36%, а доля ситуации «вне MPM» — 60,5%. Наименее благоприятная ситуация достигается при конфигурации lowdelay_P_main. В этом случае частоты сценариев попадания режима в MPM[0] и непопадания равны соответственно 30,4 и 43%. Такое различие объясняется особенностями алгоритма, в котором при отсутствии соседнего блока используется наиболее вероятный.

2.8. Резюме

Анализируя результаты экспериментов, можно выделить следующие общие закономерности.

Наиболее частым, независимо от характера контента и конфигурации кодирования, является режим INTRA_PLANAR. Его доля изменяется от 13,38% на последовательности Basketball Drill при конфигурации intra_main до 43,19% на последовательности Kimono при конфигурации lowdelay_P_main. При этом из всех разностей частот соседних режимов разность, соответствующая режиму INTRA_PLANAR, максимальна и существенно превышает все остальные.

Распределение остальных режимов меняется в зависимости от последовательности. В большинстве случаев на втором и третьем местах идут режимы INTRA_DC и INTRA_ANGULAR_26 соответственно. Как правило, режим INTRA_DC используется чаще, однако на последовательностях People On Street при конфигурации lowdelay_P_main и Four People при конфигурациях intra_main и lowdelay_main их частоты примерно совпадают.

Отдельно следует выделить видеопоследовательность Basketball Drill. Вследствие специфического характера контента кодер более эффективно кодирует фон, используя действительное направление его текстур. Из результатов экспериментов видно, что угловые режимы интрапредсказания 18–20 применяются намного чаще, чем в других последовательностях. Такая ситуация обусловлена тем, что режимы 18–20 наиболее точно описывают направление изменения фоновых текстур в кадрах последовательности. В большей степени это выражено при конфигурации intra_main, поскольку в этом случае используются только внутрикадровые предсказания и, следовательно, индуцированные ими эффекты проявляются наиболее ярко. Нетипичный характер контента также изменяет распределение режимов 0 и 1: частота наиболее применимого режима INTRA_PLANAR и её отрыв от следующего режима на данной последовательности минимальны; частота режима INTRA_DC находится на третьем месте после режима INTRA_ANGULAR 18.

В целом, для усреднённых по конфигурациям типичных видеопоследовательностей в множестве наиболее применимых режимов, помимо INTRA_PLANAR и INTRA_DC, характерно наличие вертикального и горизонтального угловых режимов—INTRA_ANGULAR_26 и INTRA_ANGULAR_10. Этот факт обусловливается большим количеством вертикально и горизонтально направленных текстур в кадрах типичных видеопоследовательностей.

Рассмотрим статистики попадания режимов интрапредсказания в разные позиции массива MPM.

Из результатов экспериментов видно, что наиболее частыми ситуациями являются попадание режима в нулевой элемент массива и непопадание в массив вовсе. Распределения при этом меняются довольно существенно как между последовательностями, так и внутри одной последовательности между конфигурациями. Частота попадания режима в элемент MPM[0] варьируется от 26,6 % на видео Kimono при конфигурации intra_main до 48,7 % на видео Blowing Bubbles при конфигурации randomaccess_main. Частота непопадания режима в массив изменяется от 26,5 % на последовательности Blowing Bubbles при конфигурации randomaccess_main до 44,1 % на последовательности Kimono при конфигурации lowdelay_P_main. Попадания режимов в MPM[1] и MPM[2] происходят с близкими частотами; при этом во всех экспериментах доля попаданий режима в MPM[1] несколько выше, чем в MPM[2]. Наибольшая частота попаданий в MPM[1] составляет 21,4 % для последовательности Kimono при конфигурации intra_main.

Следует отметить, что во всех экспериментах массив MPM, содержащий частоты попадания режимов, оказался упорядочен по убыванию. Такая ситуация объясняется особенностями алгоритма построения MPM. Действительно, для внутрикадрового предсказания характерна высокая степень корреляции соседних блоков между собой, а, согласно алгоритму, режимы соседних блоков попадают в нулевой и первый элементы массива MPM. Второй элемент определяется, исходя из статистических данных, по остаточному принципу. Отметим также, что в случае недоступности соседних блоков MPM определяется режимами 0, 1, 26, которые, согласно проведённым экспериментам, в общем случае являются наиболее вероятными.

Заключение

В ходе проведённого исследования собраны статистические данные о внутрикадровом предсказании в новейшем стандарте сжатия видеоданных H.265/HEVC. Полученные результаты позволяют сделать вывод о том, что используемая методика в малой

степени учитывает характер сжимаемого материала, из-за чего возможны ситуации неэффективного кодирования передаваемых данных. Ещё одним отрицательным следствием текущего подхода является большое количество ситуаций непопадания режима внутрикадрового предсказания в массив наиболее вероятных режимов. Поскольку попадание в тот или иной элемент массива непосредственно определяет затраты на передачу режима, актуальной является задача построения такой процедуры кодирования, при которой три наиболее вероятных режима чаще всего попадали бы в массив, а ситуации непопадания были бы минимальны. В HEVC это свойство часто нарушается—во всех экспериментах элементы MPM[1] и MPM[2] встречались реже, чем ситуация непопадания режима в MPM. Более гибкий подход к выбору массива наиболее вероятных режимов, возможно, учитывающий характер видеоматериала, может стать предметом дальнейшего развития данного направления видеокодирования.

Следует отметить, что полученные в рамках данного исследования статистики учитывают логику принятия решений кодером, основанную на стандарте HEVC. Иными словами, при принятии решения о кодировании очередного блока тем или иным режимом кодер принимает во внимание различную стоимость этого кодирования с учётом построенного массива наиболее вероятных режимов для данного блока. Таким образом, полученные статистики являются скорее не объективными характеристиками тестовых видеопоследовательностей, а лишь иллюстрацией используемой в настоящее время практики внутрикадрового кодирования. Представляет интерес получение объективных статистических данных, с учётом которых текущий подход может быть улучшен.

ЛИТЕРАТУРА

- 1. http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.html—Cisco Visual Networking Index: Global Data Traffic Forecast Update, 2012-2017. White Paper, February, 2013.
- 2. ITU-T Rec. H.265 and ISO/IEC 23008-2: High efficiency video coding. ITU-T and ISO/IEC JTC 1. Version 1. 2014.
- 3. Ohm J-R., Sullivan G. J., Schwarz H., et al. Comparison of the coding efficiency of video coding standards-including high efficiency video coding (HEVC) // IEEE Trans. Circuits and Systems for Video Technology. 2012. V. 22. No 12. P. 1669–1684.
- 4. ITU-T Rec. H.263: Video Coding for Low Bitrate Communication. ITU-T. Version 1-1995, version 2-1998, version 3-2000.
- 5. ISO/IEC 14496-2 (MPEG-4 Visual): Coding of Audio-Visual Objects. P. 2: Visual. ISO/IEC JTC 1. Version 1-1999, version 2-2000, version 3-2004.
- 6. ITU-T Rec. H.264 and ISO/IEC 14496-10 (AVC): Advanced Video Coding for Generic Audiovisual Services. ITU-T and ISO/IEC JTC 1. Version 1-2003, version 2-2004, versions 3,4-2005, versions 5,6-2006, versions 7,8-2007, versions 9,10,11-2009, versions 12,13-2010, versions 14,15-2011, version 16-2012.
- 7. http://hevc.hhi.fraunhofer.de/ High Efficiency Video Coding (HEVC). 2014.
- 8. JCTVC-K1100: Common test conditions and software reference configurations // Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11. 11th Meeting: Shanghai, CN, 10–19 October 2012.