



Субъективное Сравнение Современных Видеокодеков

Руководитель проекта: Дмитрий Ватолин

Проведение сравнения: Александр Паршин, Олег Петров

Обработка, перевод: Олег Петров

Проверка: Александр Паршин, Артем Титаренко

Кодеки:

- DivX 6.0
- XviD 1.1.0
- x264 svn-352
- WMV 9.0

Число экспертов: 50

Метод сравнения: SAMVIQ

Январь 2006

CS MSU Graphics&Media Lab

Video Group

<http://www.compression.ru/video>

Содержание

Благодарности	3
Аннотация.....	4
Обзор	5
Кодеки.....	5
Настройки кодеков	5
Настройки декодеров.....	5
Тестовые видео	5
Задачи и правила тестирования.....	6
Задачи субъективного тестирования кодеков	6
Выбор последовательностей и битрейтов.....	6
Кодирование последовательностей	6
Метод субъективного тестирования.....	8
Прохождение тестирования	10
Последовательности, использовавшиеся в тестировании	11
Battle	11
Rancho	12
Matrix sc.1	13
Matrix sc.2	14
Кодеки, использовавшиеся в тестировании	15
DivX 6.0.....	15
XviD 1.1.0-beta2	15
x264 svn-352.....	16
WMV 9.0	17
ffdshow.....	17
Результаты тестирования	18
Анализ субъективных результатов.....	18
Объективные данные.....	19
Графики MOS+PSNR/bitrate	20
Battle	20
Rancho	21
Matrix sc.1	22
Matrix sc.2	23
Графики MOS+PSNR, сгруппированные по битрейту.....	24
Корреляция объективных метрик и субъективных оценок.....	26
Результаты сравнения	30
Общие выводы.....	32
Литература	33

Благодарности

Авторы хотят выразить признательность Стасу Солдатову и Алексею Москвину за помощь в проверке и подготовке отчета, а также всем экспертам, участвовавшим в субъективном сравнении, за их неоценимую помощь в создании этого сравнения.

Аннотация

В этом сравнении оценивается качество четырех популярных кодеков на двух битрейтах (690 и 1024 kbps) с использованием субъективного сравнения. Среднее субъективное мнение о качестве сжатых последовательностей можно считать самой главной характеристикой работы кодека. В качестве метода субъективного сравнения использован метод SAMVIQ. В нашем субъективном сравнении приняло участие **50 экспертов**; их оценки формируют результирующий рейтинг кодеков. Были измерены три объективные метрики (PSNR, VQM, SSIM) и проведен анализ их соответствия субъективному мнению. Результаты субъективного сравнения подтверждают, что кодек х264 стандарта H.264 предоставляет значительно лучшее субъективное качество, чем другие популярные кодеки, участвовавшие в сравнении. Также показано, что метрика PSNR не всегда может быть надежным показателем качества видео на реальных последовательностях (т.е. не созданных специально для демонстрации неадекватности метрики).

Обзор

Кодеки

Кодек	Изготовитель	Версия
DivX	DivXNetworks	6.0 b1571-CenterOfTheSun
XviD	Кодек с открытыми исходниками	1.1.-125 ("xvid-1.1.0-beta2")
x264	Кодек с открытыми исходниками	Core 48 svn-352M by Sharktooth
WMV	Microsoft Corporation	9.0

Настройки кодеков

Кодек	Параметр	Значения
DivX	Bitrate	690 kbps, 1024 kbps
XviD	Target bitrate	690 kbps, 1024 kbps
x264	Average Bitrate	690 kbps, 1024kbps
WMV	Bit rate	700000 bps, 1048576 bps

Остальные настройки кодеков оставались без изменений после их установки. Их значения можно увидеть на скриншотах интерфейсов кодеков.

Настройки декодеров

Для декодирования полученных последовательностей для всех кодеков, кроме x264, использовались декодеры, поставляемые вместе с соответствующими кодеками. В качестве декодера для x264 мы использовали ffdshow, версия ff_x264.dll 33, от Milan Cutca.

Тестовые видео

В следующей последовательности указаны характеристики последовательностей, использованных в нашем сравнении (также см. ["Кодирование последовательностей"](#)).

Название	Длина [кадры]	Длина [секунды]	Разрешение	Источник
Battle	257 кадров	10.71	704x288	MPEG2 (DVD)
Rancho	240 кадров	10.01	704x288	MPEG2 (DVD)
Matrix sc.1	250 кадров	10.00	720x416	MPEG2 (DVD)
Matrix sc.2	250 кадров	10.00	720x416	MPEG2 (DVD)

Задачи и правила тестирования

Задачи субъективного тестирования кодеков

Нашей лабораторией на протяжении долгого времени проводились различные тестирования кодеков для видео, изображений и звука. Все они использовали **объективные** метрики, такие как PSNR или VQM и SSIM, что часто вызывало вопрос о том, насколько адекватны объективные замеры субъективно воспринимаемому качеству.

Крупные организации, такие как VQEG и ITU, на протяжении многих лет проводили субъективные сравнения видео и оценивали точность предсказания различными метриками мнения пользователя, получив в результате множество ценных результатов [5]. Но в основном эти сравнения были ориентированы на телевизионный материал и видеокодеки стандарта MPEG2, и только недавно начали появляться сравнения кодеков современных стандартов, способных обеспечивать приемлемое качество на более низких битрейтах и предназначенных для просмотра видео на компьютере.

Задачей нашего тестирования было **субъективное сравнение новых версий популярных кодеков**, сравнение результатов с данными объективных метрик и отработка технологии субъективного тестирования.

Выбор последовательностей и битрейтов

Мы были ограничены в количестве выбранных битовых скоростей и фильмов, так как в процессе тестирования каждый эксперт должен был просмотреть каждую последовательность, сжатую с каждым битрейтом, как минимум один раз (подробнее см. ниже "[Метод субъективного тестирования](#)"). Поэтому было решено сосредоточиться на одной важной области применения тестируемых видеокодеков – сжатии фильмов.

Были выбраны четыре сцены из фильмов "Терминатор 2" и "Матрица": две со средним и две с очень быстрым движением. Именно искажения при движении часто являются причиной дискомфорта зрителя при просмотре сжатого видео. Также были выбраны два битрейта, типичные для сжатия фильмов – 690 kbps и 1024 kbps¹.

Кодирование последовательностей

Последовательности были закодированы в один проход без тонкой настройки параметров кодека (большинству настроек было оставлено значение по умолчанию), изменялся только битрейт. Так как длина последовательностей около 10 секунд, для стабилизации работы кодека и точного удержания заданного битрейта при сжатии был применен следующий метод: каждая последовательность повторялась пять раз, затем сжималась, от сжатого видео отрезался последний повтор, который и использовался для тестирования.

¹ 690 kbps – один фильм в среднем качестве на CD, 1024 kbps – один фильм с высоким качеством на два CD или два фильма на один DVD

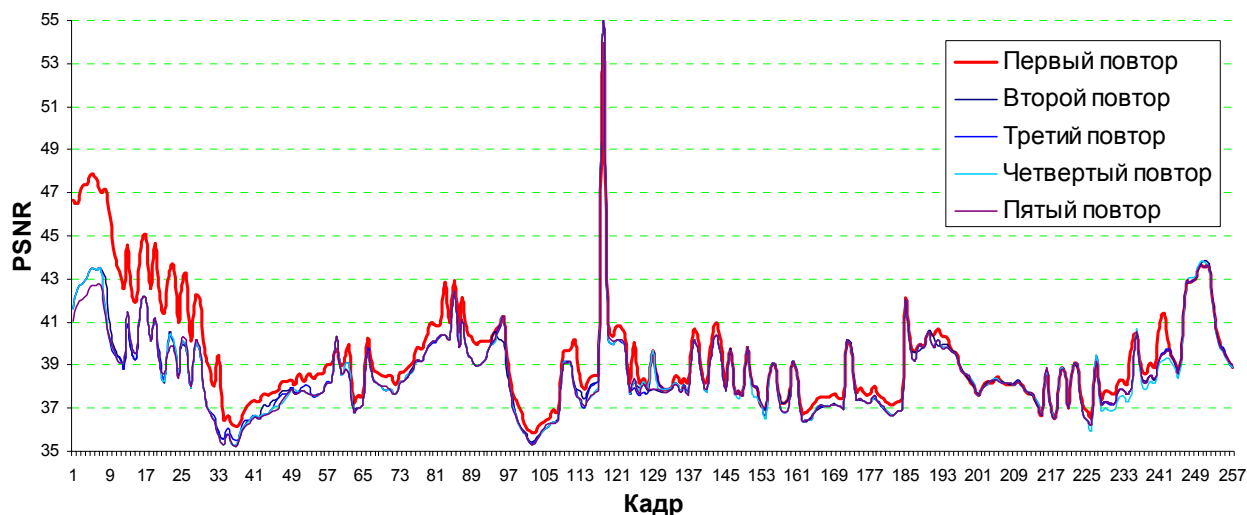


Рисунок 1. График PSNR для повтора "Battle", WMV 690 kbps

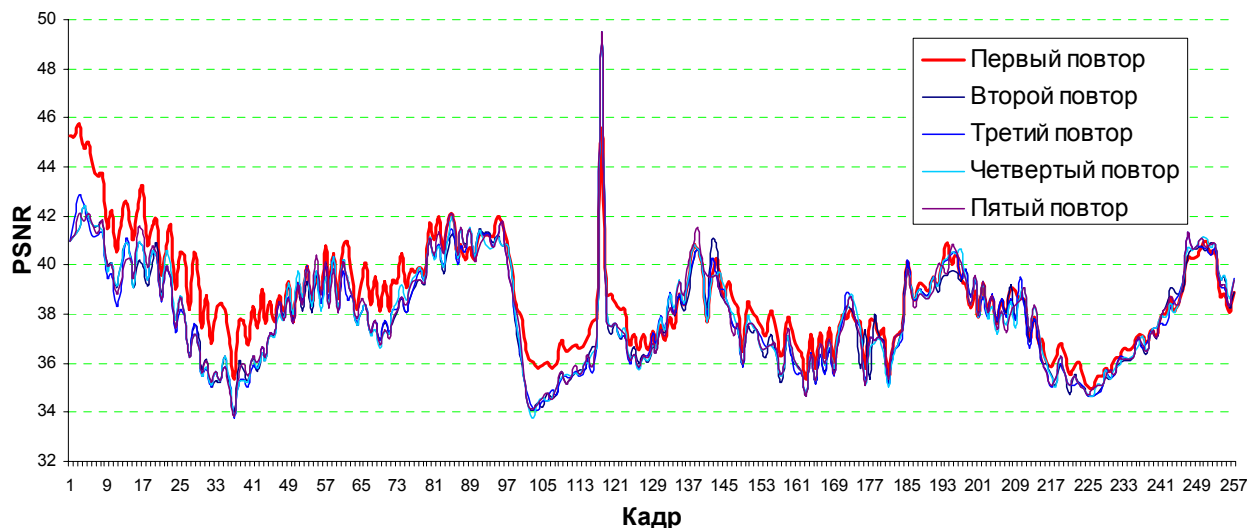


Рисунок 2. График PSNR для повтора "Battle", XviD 690 kbps

На графиках PSNR для повторов фильма "Battle" видно, что уже после второго повтора кодеки начинают выдавать стабильный результат, а результат первого повтора слишком оптимистичен.

Для вырезания последнего повтора использовалась функция "Direct stream copy" в программе VirtualDub[9]. Для всех последовательностей удалось корректно отрезать последний повтор - все кодеки делали ключевым один и тот же кадр (отрезать закодированный поток можно только начиная с ключевого кадра). На всех фильмах, кроме Battle, этот кадр совпал с началом повтора. На Battle, видимо, из-за частой смены сцен, все кодеки сделали ключевым кадр, отстающий на 10 кадров от начала повтора.

Как демонстрирует следующая таблица, кодеки достаточно точно уложились в требуемый битрейт (показан битрейт последнего повтора, который непосредственно демонстрировался экспертам).

	690 kbps				1024 kbps			
	Battle	Matrix sc.1	Matrix sc.2	Rancho	Battle	Matrix sc.1	Matrix sc.2	Rancho
DivX	692	690	687	688	1044	1020	1026	1028
XviD	681	685	680	687	1013	1034	1025	1017
x264	688	603	687	697	1021	907	954	1031
WMV	692	693	667	697	1034	1049	1046	1045

Метод субъективного тестирования

Идея субъективного тестирования состоит в том, что видео, пропущенное через сравниваемые системы, демонстрируется группе экспертов, которые выставляют оценки, основываясь на своих впечатлениях от качества. Существует много методов демонстрации последовательностей и сбора оценок, многие из них описаны в рекомендациях ITU [1]. К сожалению, в основном они рассчитаны на сравнение видео в телевизионном формате, и не очень удобны для проведения тестирования на PC.

Для нашего тестирования мы выбрали метод SAMVIQ (Subjective Assessment of Multimedia Video Quality, [2]), недавно разработанный EBU для сравнения видеокодеков и проходящий стандартизацию в ITU. Он был разработан специально для проведения сравнений на PC, и более удобен и прост в использовании, чем другие методы субъективного сравнения. Как и многие другие методы, SAMVIQ реализован в MSU Perceptual Video Quality tool [6].

Этапы тестирования:

1. Эксперт вводит свое имя (любая уникальная последовательность символов).



Рисунок 3. Диалог ввода имени

2. Тест на цветовосприятие (используются стандартные таблицы Ишихары).

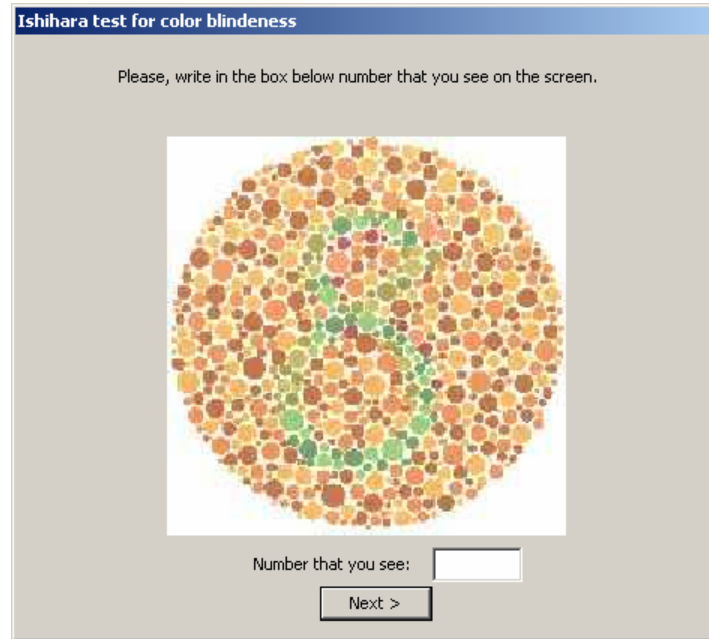


Рисунок 4. Пример теста на цветовосприятие

3. Для каждой тестовой последовательности:

1) Демонстрируется референсное (исходное) видео.



Рисунок 5. Вид окна проигрывания

2) До тех пор, пока есть не просмотренные сжатые варианты этого видео, эксперт выбирает вариант видео, смотрит его и выставляет оценку. **Оценка для фильма принадлежит отрезку от 0 до 100, чем выше, тем лучше.** Оценка уже просмотренных вариантов последовательности может быть изменена в любой момент, также возможно пересмотреть любой из вариантов.

Рисунок 6. Диалог оценки

- 3) Если все варианты видео просмотрены, то эксперт может перейти к следующей тестовой последовательности.

Разные варианты сжатой последовательности скрыты за буквенными обозначениями, поэтому **эксперт не знает, какой кодек он оценивает в данный момент.**

Исходное видео доступно в явном виде² (кнопка “Reference sequence”), также оно скрыто под одним из буквенных обозначений и оценивается наравне со сжатыми. Всего эксперт должен был оценить 9 версий каждой последовательности (4 кодека * 2 битрейта + 1 исходное видео).

Прохождение тестирования

Субъективное тестирование проходило в течении трех дней. Всего в нем приняло участие **50 человек**, из них 14 – специалисты по видео. Использовались три типа мониторов: 6 x 15” CRT Dell, 1 x 17” CRT Samsung и 2 x 17” LCD Samsung, они находились в двух различных помещениях (как утверждается в [8], тип монитора не оказывает существенного влияния на результаты тестирования). Одновременно в тестировании принимало участие до 9 человек. Во время тестирования была обеспечена тихая обстановка, не отвлекающая экспертов, настройки мониторов были отрегулированы. Предварительно все эксперты были проинструктированы о способе прохождения тестирования.

² Исходное видео было сжато без потерь кодеком Huffuv v2.1.1 для обеспечения возможности проигрывания на компьютерах экспертов.

Последовательности, использовавшиеся в тестировании

Battle



Battle, кадр 215

Название	Battle
Разрешение	704x288
Особенности	Отрывок из фильма "Терминатор-2". Сложна для сжатия: яркость постоянно изменяется, движение очень сильное, постоянные смены сцен.

Rancho



Rancho, кадр 149

Название	Rancho
Разрешение	704x288
Особенности	Также из фильма "Терминатор-2". Сначала движение слабое, затем происходит смена сцены и начинается сильное движение, которое затем успокаивается.

Matrix sc.1



Matrix sc.1, кадр 178

Название	Matrix sc.1
Разрешение	720x416
Особенности	Отрывок из фильма "Матрица". Состоит из двух частей: в первой камера следит за ногами вбегающих солдат, во второй камера вращается. Движение не быстрое, но имеет сложный характер.

Matrix sc.2



Matrix sc.2, кадр 52

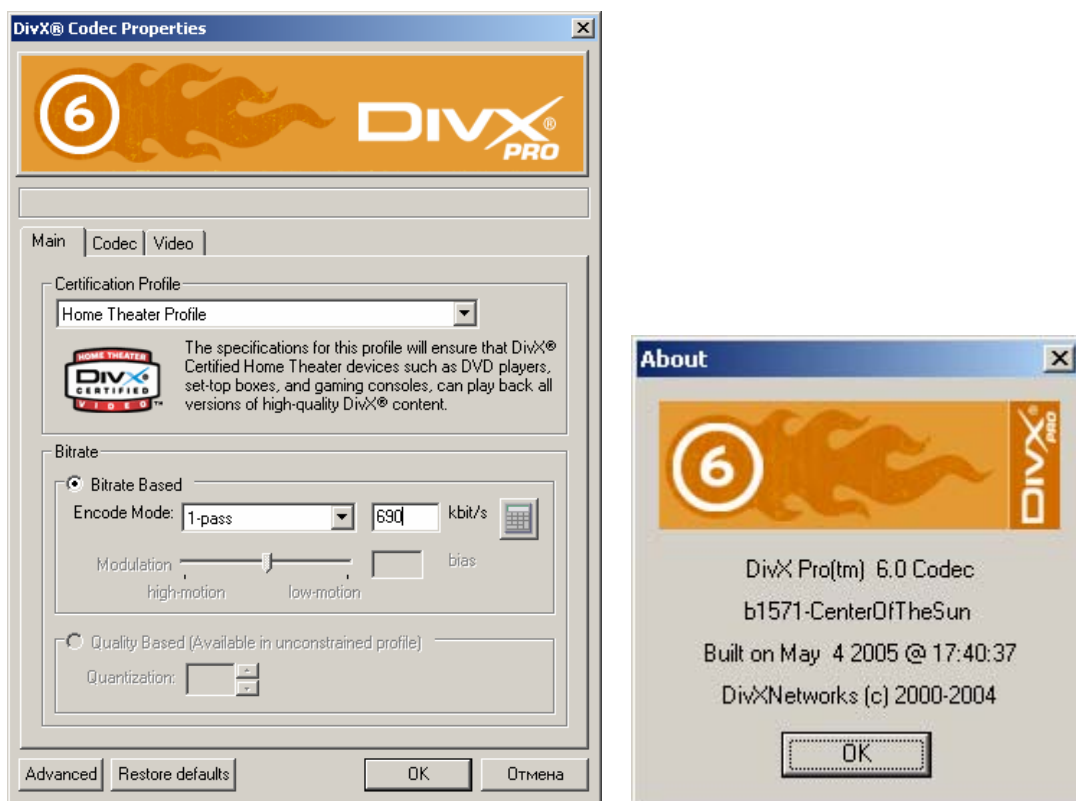
Название	Matrix sc.2
Разрешение	720x416
Особенности	Фрагмент из фильма "Матрица". Кадр заполнен отлетающими обломками от стен, которые двигаются по всевозможным траекториям. Движение не очень быстрое, но его трудно предсказать.

Кодеки, использовавшиеся в тестировании

DivX 6.0

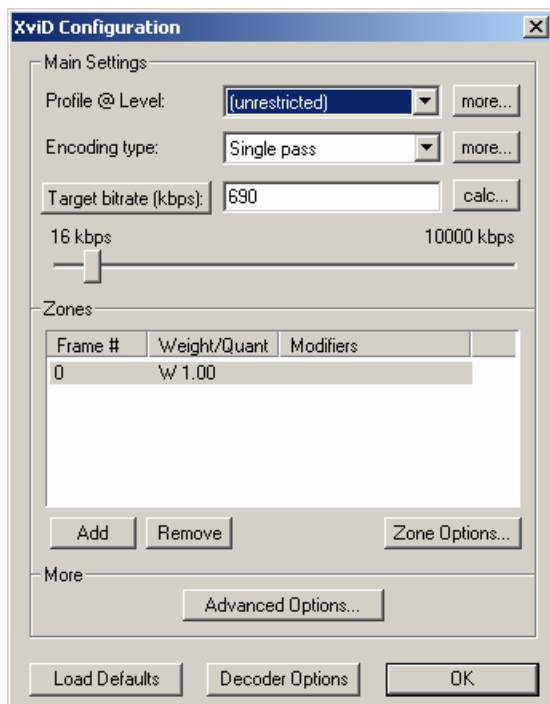
Кодек формата MPEG4-ASP. Для сжатия и замера объективных метрик применялась "Create Bundle" версия кодека, при субъективном тестировании использовалась бесплатный пакет "Play Bundle".

При кодировании кодек ставился на чистую систему, никакие настройки, кроме "Bitrate"- "kbit/s", не изменялись.



XviD 1.1.0-beta2

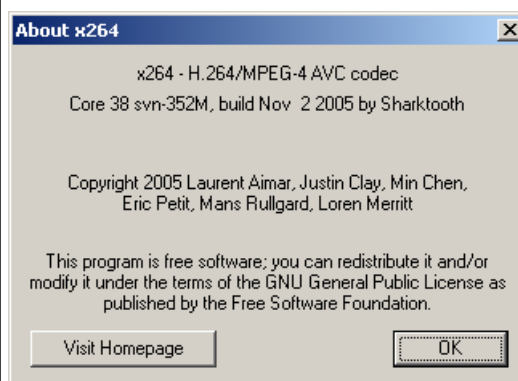
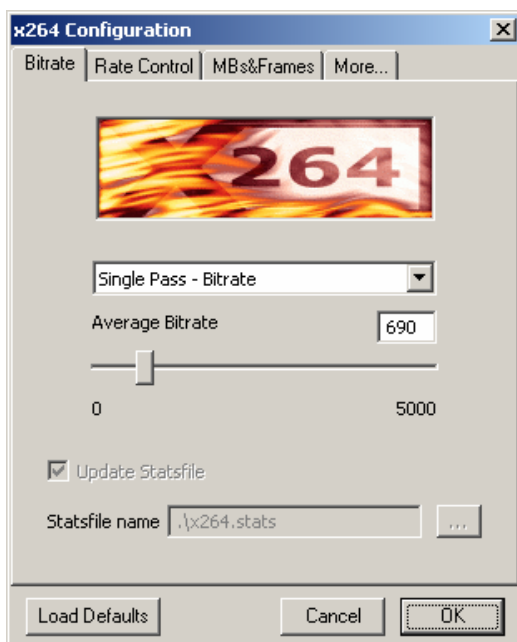
Бесплатный кодек формата MPEG4-ASP с открытыми исходными текстами. Использовалась самая свежая на момент тестирования версия кодека. Несмотря на то, что она еще находится в статусе "beta", никаких проблем во время тестирования обнаружено не было, и так как ее превосходство над предыдущей стабильной версией достигает 1 дБ, было решено использовать ее в тестировании. Изменялся параметр "Target bitrate".



К сожалению, в декодере этой версии XviD'a по умолчанию отключен деблокинг. Он не был включен в процессе тестирования и замеров объективных метрик в связи с политикой невмешательства в параметры декодера – именно такое качество получит средний пользователь при просмотре закодированного видео.

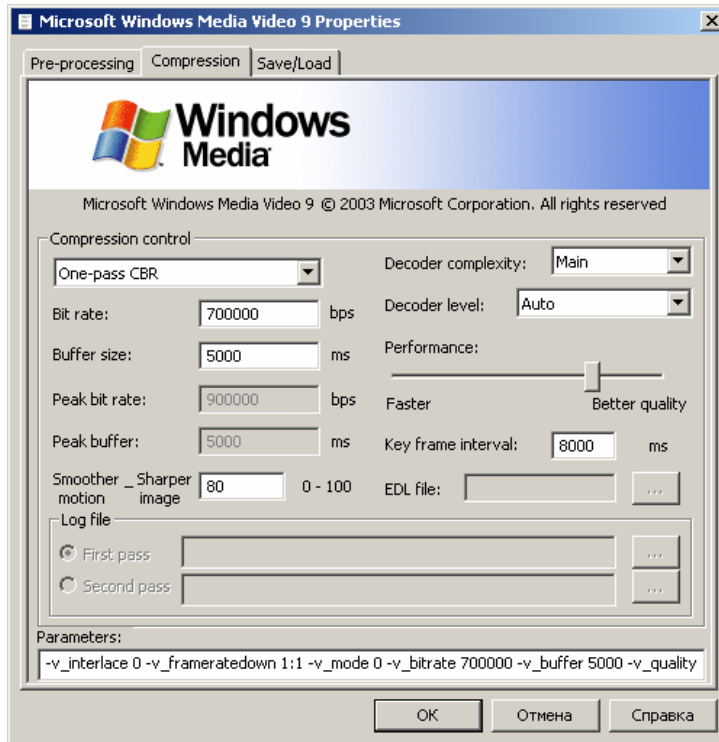
x264 svn-352

Бесплатный кодек формата H.264 с открытыми исходными текстами, использовался только для получения сжатых последовательностей. В качестве декодера был использован ffdshow. Изменялся параметр "Average Bitrate".



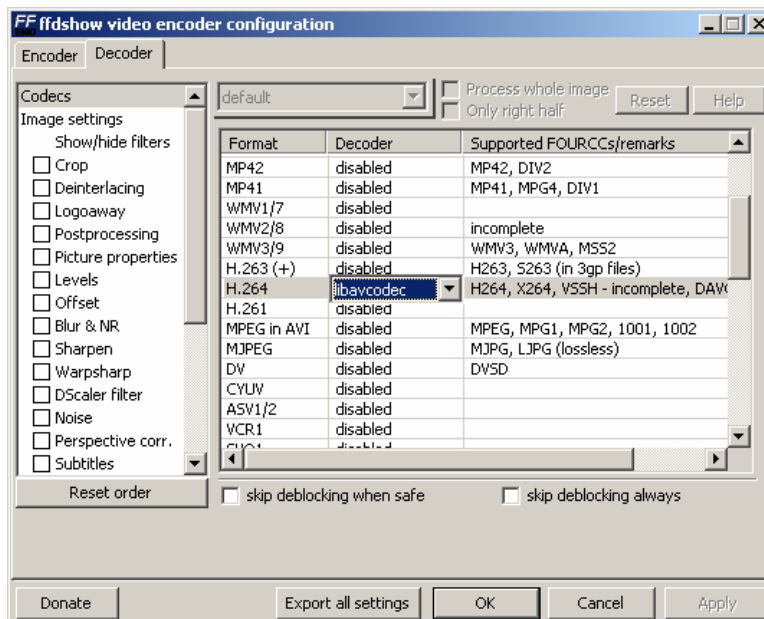
WMV 9.0

Бесплатный кодек проприетарного формата. Декодер для него встроен по умолчанию в последние версии Windows. Изменялся параметр “Bit rate”



ffdshow

Использовался только для декодирования x264, версия декодера ff_x264.dll 33, build data Aug 10 2005 16:33:17



Результаты тестирования

Анализ субъективных результатов

Средняя субъективная оценка последовательности называется MOS (Mean Opinion Score). Эти оценки для каждого фильма и битрейта были получены простым усреднением мнений всех экспертов:

$$MOS_k = \sum_{i=1}^{\text{experts_num}} \frac{\text{mark}_{i,k}}{\text{experts_num}}$$

Где k - номер фильма, для которого вычисляется оценка;
 $\text{mark}_{i,k}$ – оценка, поставленная i -м экспертом k -ому фильму;
 experts_num – общее количество экспертов.

Также были вычислены верхние и нижние границы 95% доверительного интервала, что позволяет оценить разброс мнений экспертов о данной последовательности.

Чтобы посчитать вероятность того, что эксперты действительно смогли различить два кодека на этой последовательности, для каждой пары кодеков и битрейтов вычислялось значение z -теста. Для расчетов использовалась следующая формула:

Пусть x_1 и x_2 – субъективные оценки для двух фильмов. Тогда

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\text{Var}(x_1) + \text{Var}(x_2)}{\text{experts_num}}}}$$

Где \bar{x}_1 и \bar{x}_2 - MOS для первого и второго фильма;
 $\text{Var}(x_1)$ и $\text{Var}(x_2)$ - вариации субъективных оценок;
 experts_num – общее количество экспертов.

Искомая вероятность равна

$$p(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-z^2/2} dz$$

Объективные данные

Для всех последовательностей были замерены объективные метрики PSNR, VQM и SSIM. Замеры проводились при помощи MSU Video Quality Measurement Tool [7].

PSNR – самая часто используемая на практике метрика. По смыслу она близка к среднеквадратичной ошибке, но более удобна из-за логарифмической шкалы.

$$PSNR(x, y) = 10 \cdot \log_{10} \frac{255^2 \cdot n \cdot m}{\sum_{i=1}^n \sum_{j=1}^m (x_{i,j} - y_{i,j})^2}$$

Известно множество примеров искажений, при которых значение PSNR не отражает субъективно видимой разницы между изображениями.

VQM[3] и SSIM[4] – относительно новые метрики, претендующие на лучшее, чем у PSNR, соответствие субъективному мнению.

Для того, чтобы сравнивать метрики, их результаты должны быть приведены к единой шкале. В соответствии с процедурой, описанной в [1], их данные были приведены к диапазону субъективных результатов по следующей формуле (далее “функция подгонки”, fitting function):

$$O^{\text{fitted}} = \frac{1}{1 + e^{g \cdot O + d}}$$

Где

O – исходные объективные данные;

O^{fitted} - приведенные объективные данные;

g и d – параметры.

Параметры g и d были подобраны такими, чтобы минимизировать сумму квадратов разностей между O^{fitted} и субъективными данными:

$$g, d : \sum (O^{\text{fitted}} - S)^2 \rightarrow \min$$

Где S – субъективные данные.

Полученные приведенные данные можно считать предсказанием субъективного мнения, предоставленным объективной метрикой.

Графики MOS+PSNR/bitrate

На следующих графиках представлены субъективные данные для каждого фильма, их 95% доверительные интервалы и значение MOS, предсказанное PSNR. Оно было получено после применения [функции подгонки](#) для каждого фильма в отдельности.

Battle

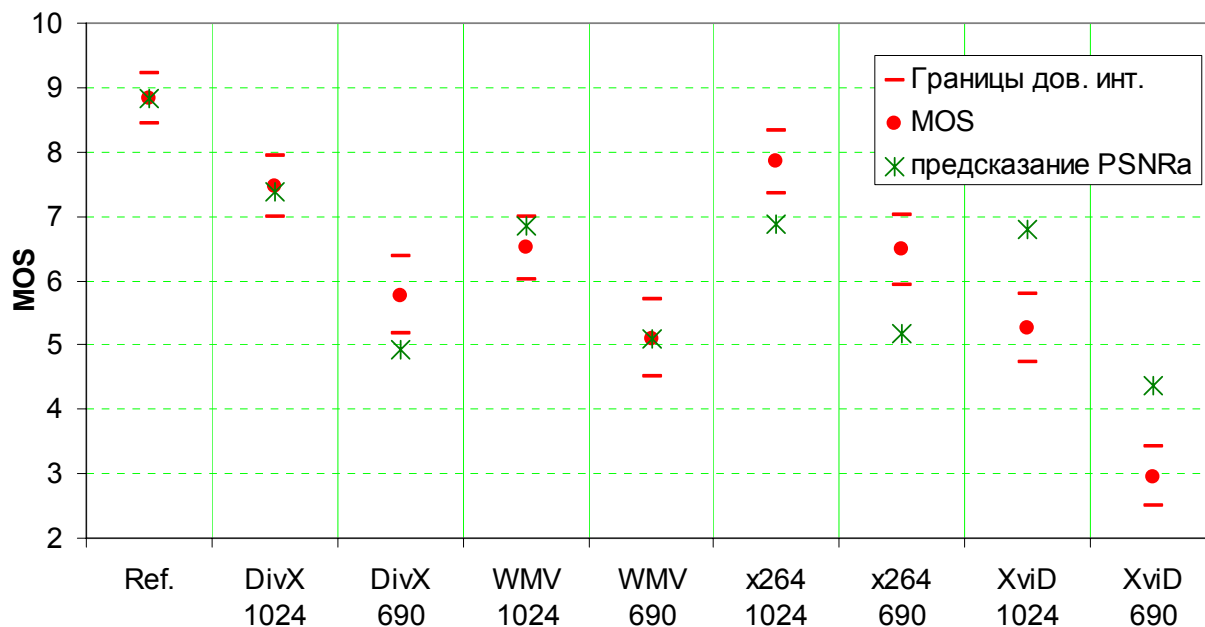


Рисунок 7. Battle

Последовательность Battle – самая сложная для кодеков. PSNR неадекватен на ней сразу в нескольких точках: например, для x264 690 и XviD 1024 показания PSNR полностью противоречат субъективному мнению. Лидер на данной последовательности – x264, он лучший на всех битрейтах. За ним DivX, WMV и XviD.

Ниже приведена таблица результатов z-теста (вероятность того, что эксперты различили два фильма).

Battle	Ref.	DivX 1024	DivX 690	WMV 1024	WMV 690	x264 1024	x264 690	XviD 1024	XviD 690
Ref.	1	1	1	1	1	1	1	1	1
DivX 1024	1	1	1	1	1	0.87	1	1	1
DivX 690	1	1	1	0.97	0.94	1	0.95	0.89	1
WMV 1024	1	1	0.97	1	1	1	0.53	1	1
WMV 690	1	1	0.94	1	1	1	1	0.65	1
x264 1024	1	0.87	1	1	1	1	1	1	1
x264 690	1	1	0.95	0.53	1	1	1	1	1
XviD 1024	1	1	0.89	1	0.65	1	1	1	1
XviD 690	1	1	1	1	1	1	1	1	1

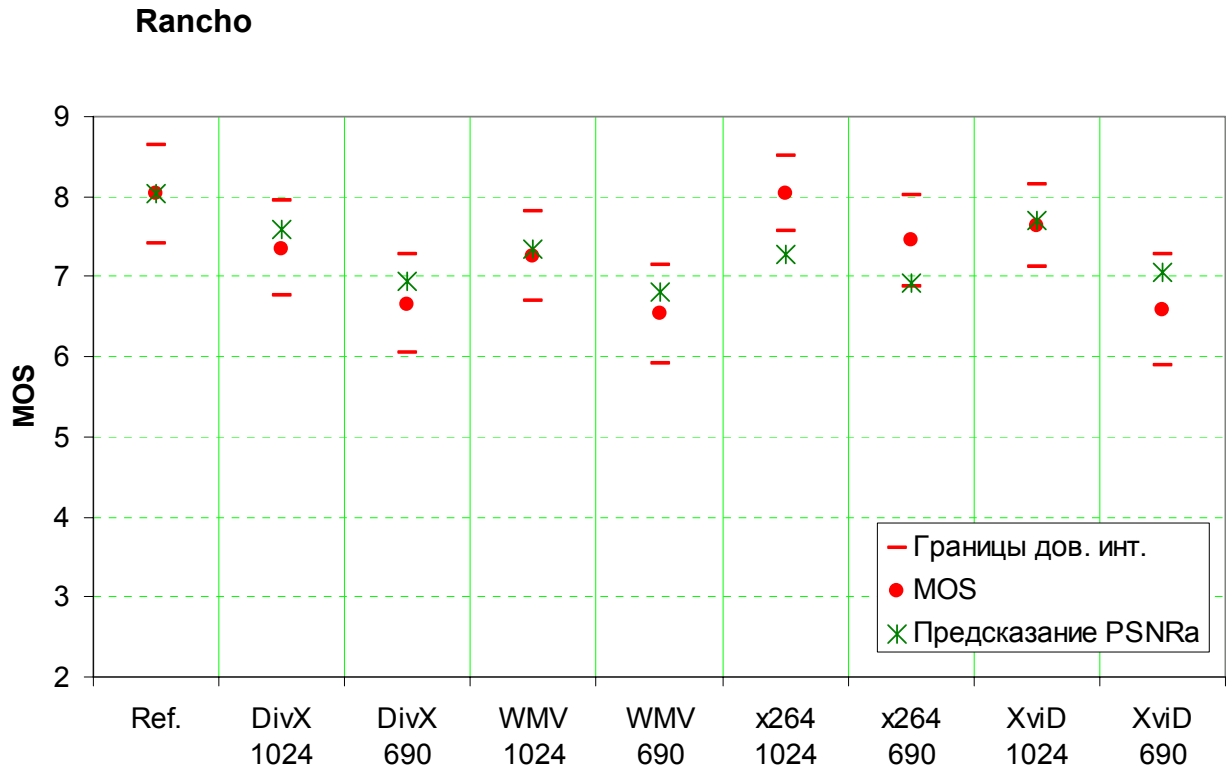


Рисунок 8. Rancho

С последовательностью Rancho все кодеки справились одинаково, разброс между субъективными оценками очень мал. По-прежнему немного впереди всех кодексов x264: на битрейте 1024 его оценка равна оценке несжатой последовательности.

Rancho	Ref.	DivX 1024	DivX 690	WMV 1024	WMV 690	x264 1024	x264 690	XviD 1024	XviD 690
Ref.	1	0.94	1	0.96	1	0.51	0.91	0.83	1
DivX 1024	0.94	1	0.94	0.59	0.97	0.96	0.59	0.76	0.95
DivX 690	1	0.94	1	0.92	0.61	1	0.97	0.99	0.56
WMV 1024	0.96	0.59	0.92	1	0.96	0.98	0.68	0.83	0.93
WMV 690	1	0.97	0.61	0.96	1	1	0.98	1	0.54
x264 1024	0.51	0.96	1	0.98	1	1	0.94	0.87	1
x264 690	0.91	0.59	0.97	0.68	0.98	0.94	1	0.69	0.97
XviD 1024	0.83	0.76	0.99	0.83	1	0.87	0.69	1	0.99
XviD 690	1	0.95	0.56	0.93	0.54	1	0.97	0.99	1

Matrix sc.1

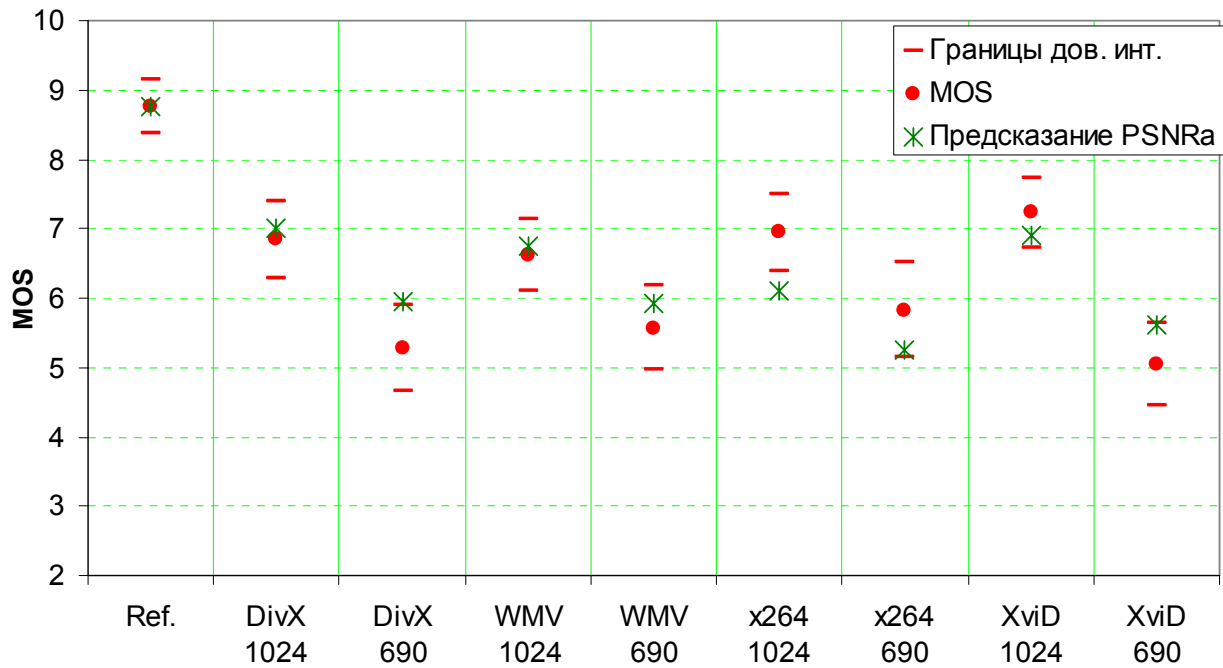


Рисунок 9. Matrix sc.1

На этом фильме в лидеры вышел XviD на битрейте 1024 kbps, но преимущество его незначительно. PSNR для этой последовательности, кроме x264 на 1024 kbps, адекватен субъективной метрике.

Matrix sc.1	Ref.	DivX 1024	DivX 690	WMV 1024	WMV 690	x264 1024	x264 690	XviD 1024	XviD 690
Ref.	1	1	1	1	1	1	1	1	1
DivX 1024	1	1	1	0.71	1	0.6	0.99	0.85	1
DivX 690	1	1	1	1	0.74	1	0.88	1	0.7
WMV 1024	1	0.71	1	1	1	0.79	0.97	0.95	1
WMV 690	1	1	0.74	1	1	1	0.71	1	0.88
x264 1024	1	0.6	1	0.79	1	1	0.99	0.78	1
x264 690	1	0.99	0.88	0.97	0.71	0.99	1	1	0.95
XviD 1024	1	0.85	1	0.95	1	0.78	1	1	1
XviD 690	1	1	0.7	1	0.88	1	0.95	1	1

Matrix sc.2

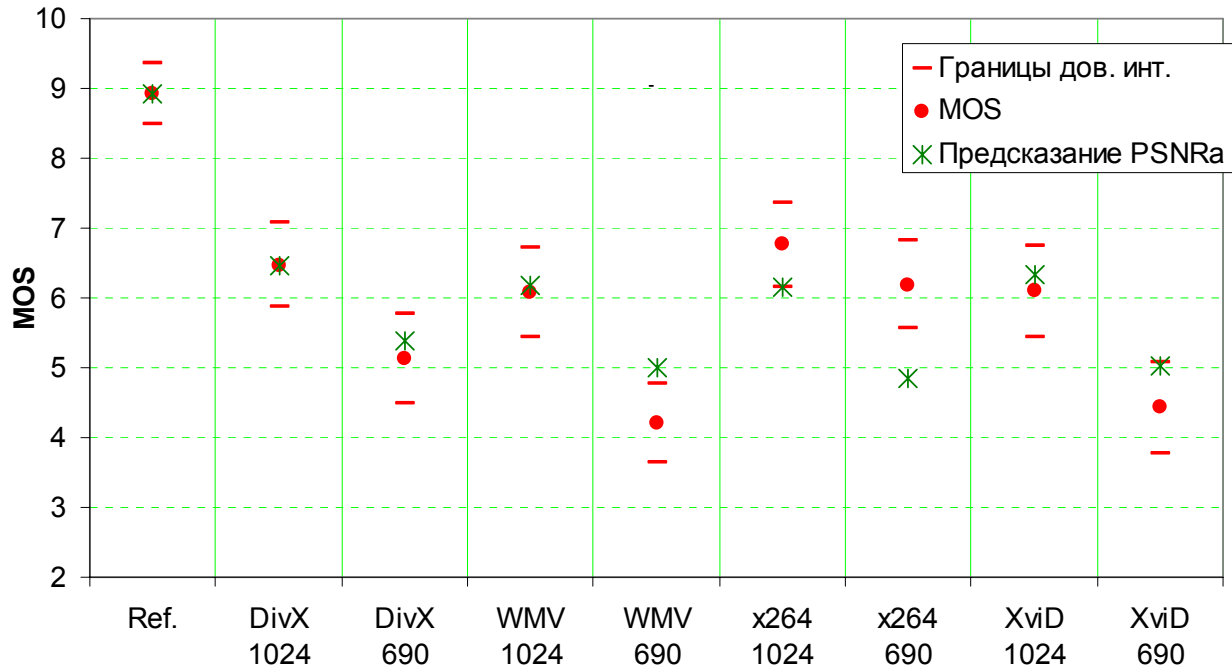


Рисунок 10. Matrix sc.2

На этой последовательности снова очень хорошо показал себя кодек x264. PSNR принимает почти одинаковые значения на DivX, WMV, x264 и XviD на бит-рейте 690 kbps, при этом субъективные мнения о них отличаются.

Matrix sc.2	Ref.	DivX 1024	DivX 690	WMV 1024	WMV 690	x264 1024	x264 690	XviD 1024	XviD 690
Ref.	1	1	1	1	1	1	1	1	1
DivX 1024	1	1	1	0.82	1	0.75	0.74	0.8	1
DivX 690	1	1	1	0.98	0.98	1	0.99	0.98	0.94
WMV 1024	1	0.82	0.98	1	1	0.94	0.6	0.52	1
WMV 690	1	1	0.98	1	1	1	1	1	0.69
x264 1024	1	0.75	1	0.94	1	1	0.9	0.93	1
x264 690	1	0.74	0.99	0.6	1	0.9	1	0.58	1
XviD 1024	1	0.8	0.98	0.52	1	0.93	0.58	1	1
XviD 690	1	1	0.94	1	0.69	1	1	1	1

Графики MOS+PSNR, сгруппированные по битрейту

Для простоты сравнения кодеков отдельно на разных битрейтах, мы приводим те же графики, что и в предыдущем параграфе, но сгруппированные по битрейту.

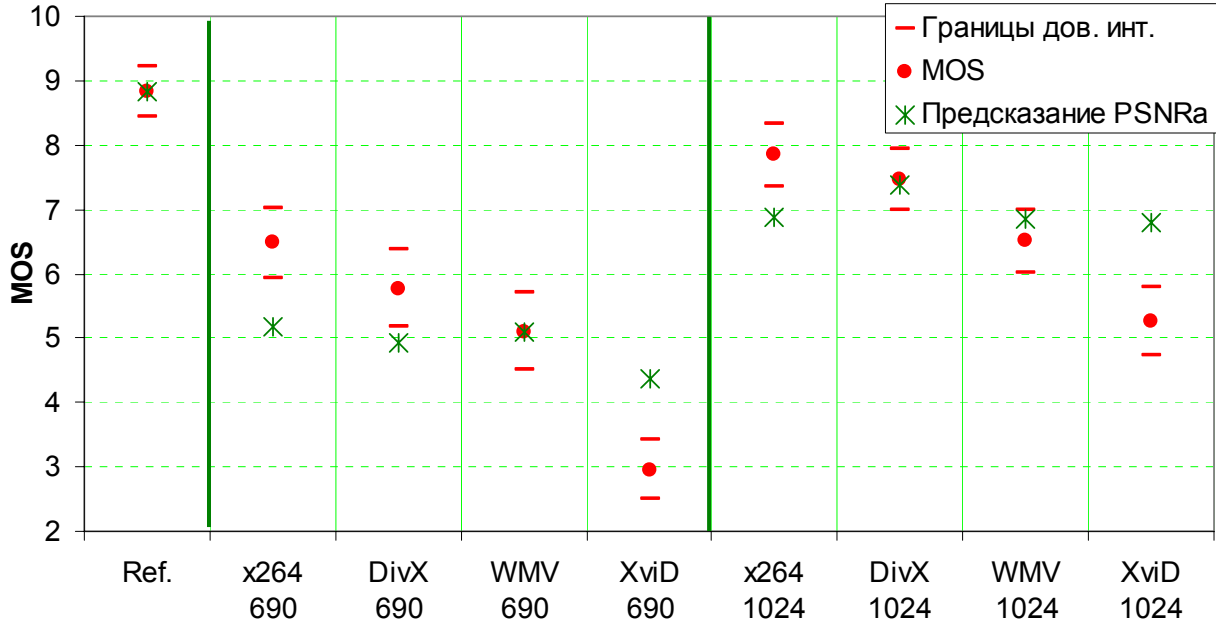


Рисунок 11. Battle

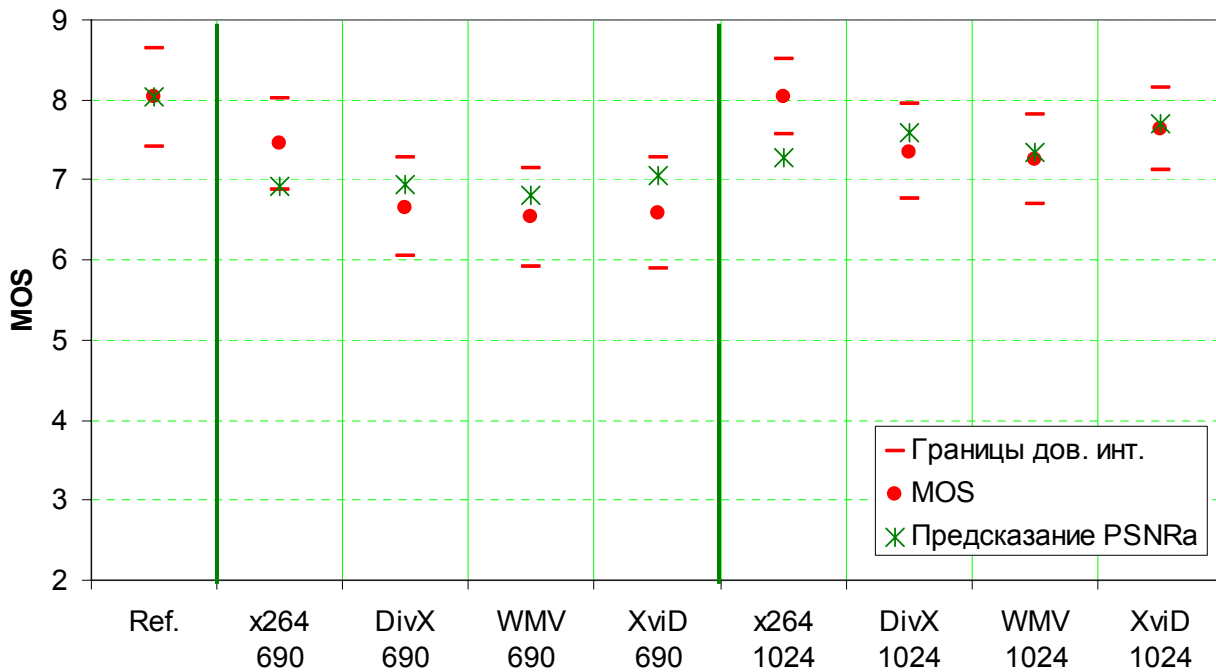


Рисунок 12. Rancho

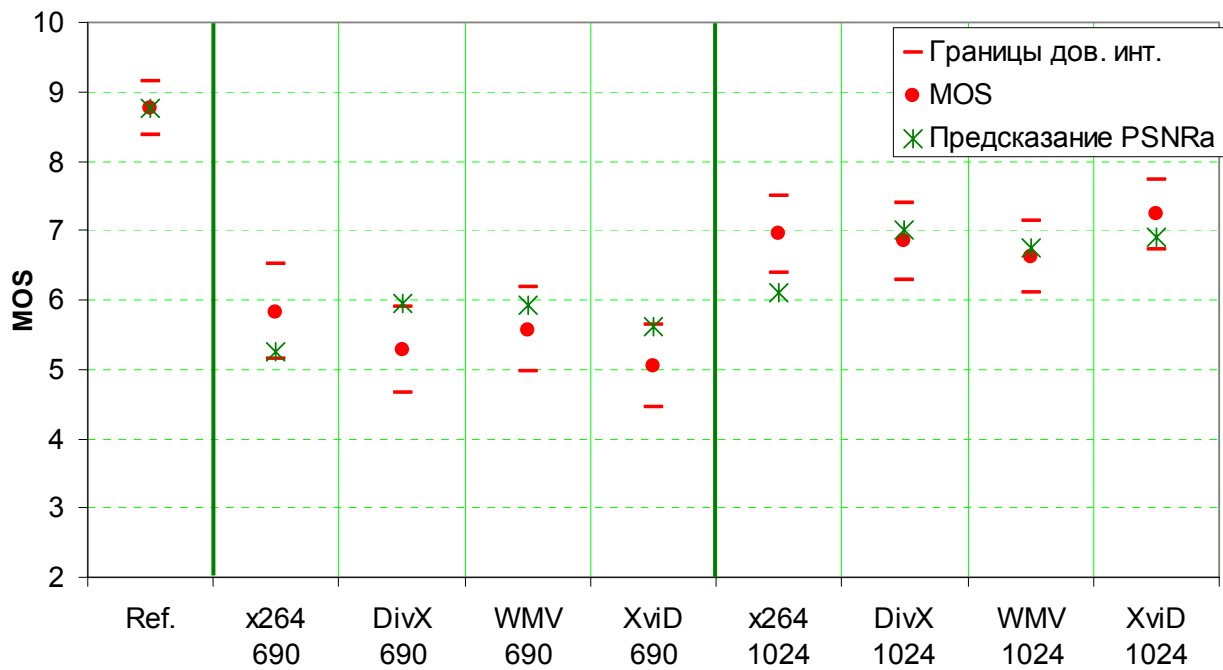


Рисунок 13. Matrix sc.1

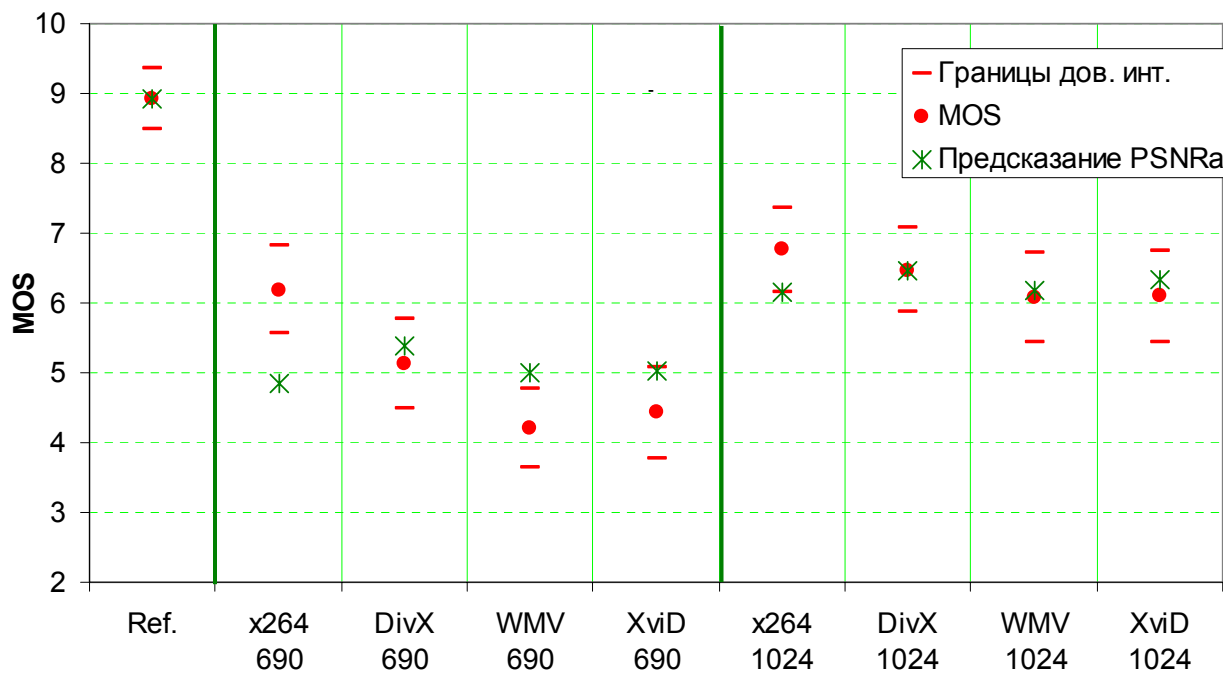


Рисунок 14. Matrix sc.2

Корреляция объективных метрик и субъективных оценок

Ниже представлены два типа графиков: первый из них – величина субъективной оценки при данной величине субъективной метрики.

Для каждой последовательности такой график нужно рассматривать отдельно, так как величина субъективной оценки зависит от контекста этой последовательности (субъективная оценка выставляется экспертом относительно остальных ее версий). На одной диаграмме они объединены для удобства восприятия.

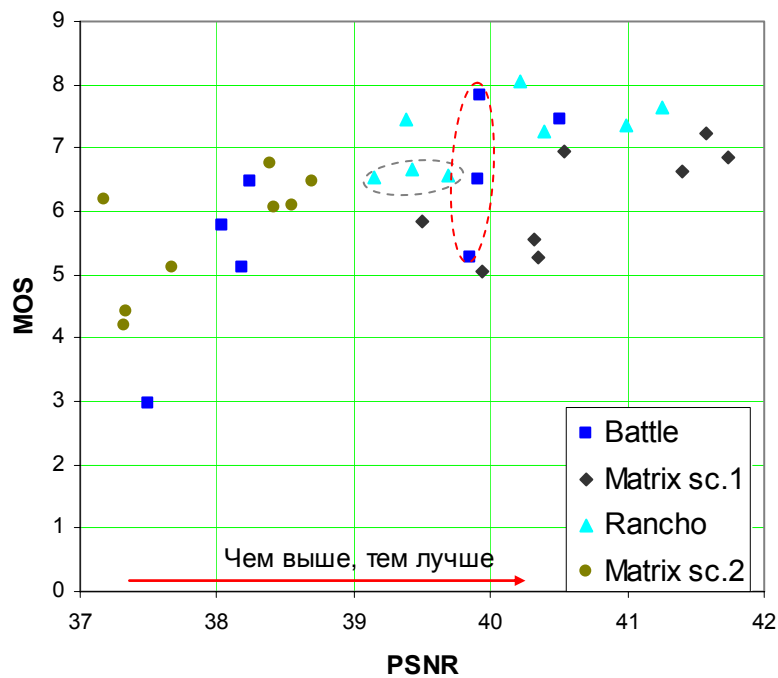


Рисунок 15. Соотношение между PSNR и MOS

Как видно, PSNR довольно ограниченно отражает реальное качество видео. Можно видеть ситуацию, когда одному значению PSNR соответствуют абсолютно разные субъективные мнения (отмечено красным овалом), и наоборот (отмечено серым овалом)

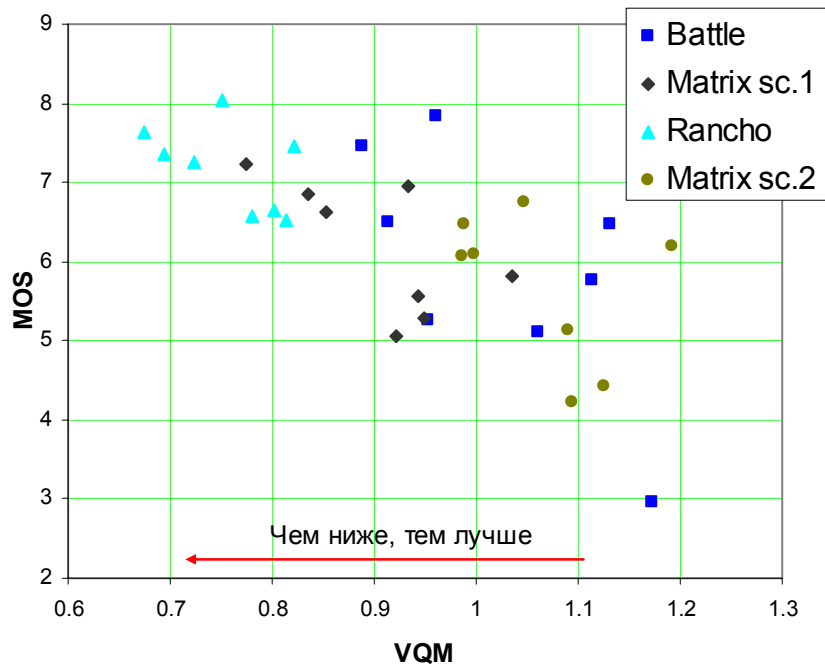


Рисунок 16. Соотношение между VQM и MOS

На нашем тестовом наборе VQM не более адекватен, чем PSNR: в общих чертах качество предсказано, но часто лучшему субъективному качеству соответствует худшее значение метрики.

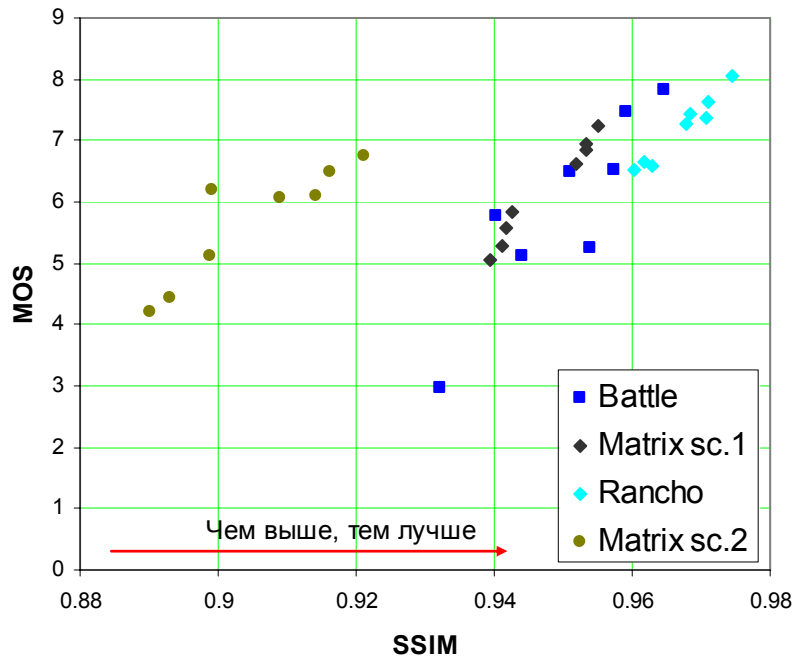


Рисунок 17. Соотношение между SSIM и MOS

SSIM предсказывает субъективные мнения с очень хорошей точностью: для каждого фильма его данные близки к прямой.

Второй тип графиков – величина оценки, предсказанной объективной метрикой, и реальной субъективной оценкой. Предсказанная оценка была получена при помощи применения [функции подгонки](#) для каждой последовательности в отдельности. Такие графики позволяют визуально оценить качество предсказания субъективной метрики (чем ближе к прямой, тем лучше):

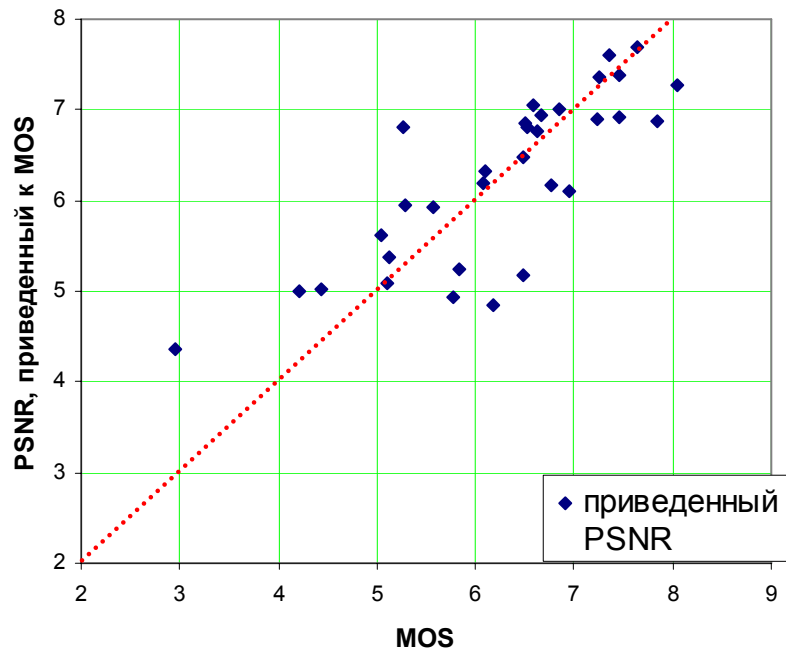


Рисунок 18. Значения PSNR, приведенные к MOS для каждого фильма

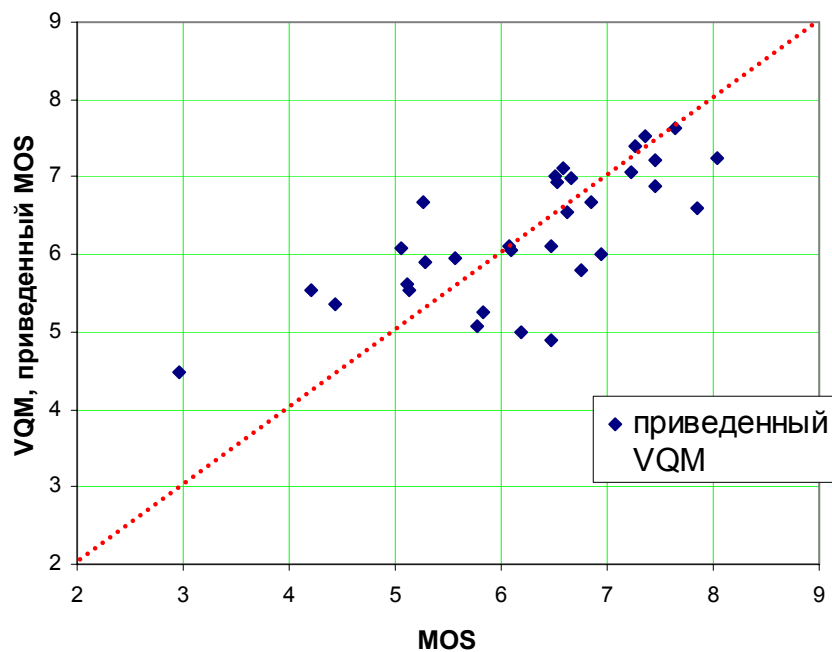


Рисунок 19. Значения VQM, приведенные к MOS для каждого фильма

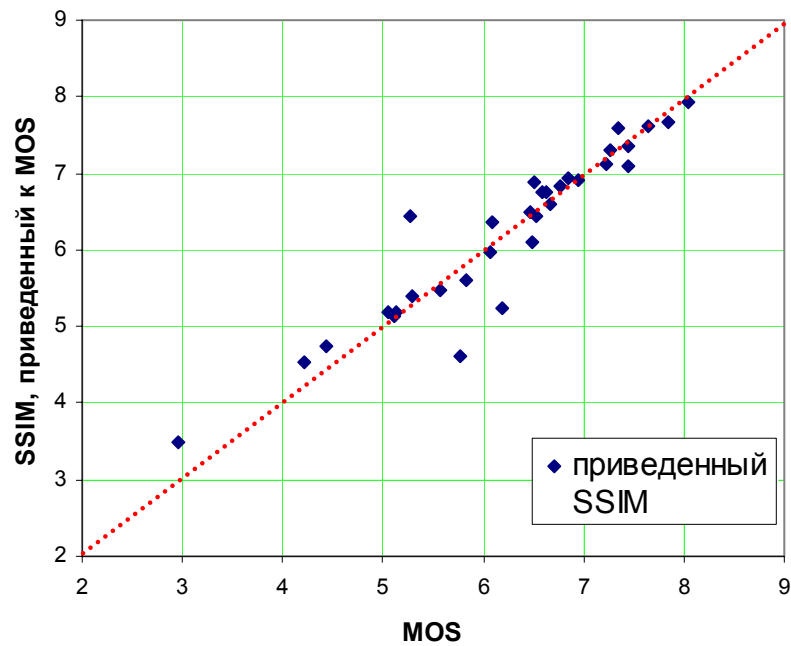


Рисунок 20. Значения SSIM, приведенные к MOS для каждого фильма

Как можно видеть, на нашем тестовом наборе PSNR и VQM были почти одинаково эффективны, а наиболее точное предсказание обеспечил SSIM.

Для того, чтобы количественно оценить предсказание объективной метрики, мы вычислили коэффициент корреляции Пирсона между каждой объективной метрикой (после применения [функции подгонки](#) к каждому из фильмов в отдельности) и субъективными оценками. Этот коэффициент принадлежит отрезку от -1 до 1 и позволяет оценить степень зависимости между величинами (чем модуль коэффициента больше, тем зависимость сильнее).

Метрика	Корреляция
PSNR	0.802
VQM	0.729
SSIM	0.937

Результаты сравнения

На следующем графике представлены значения MOS после усреднения по всем последовательностям и битрейтам.

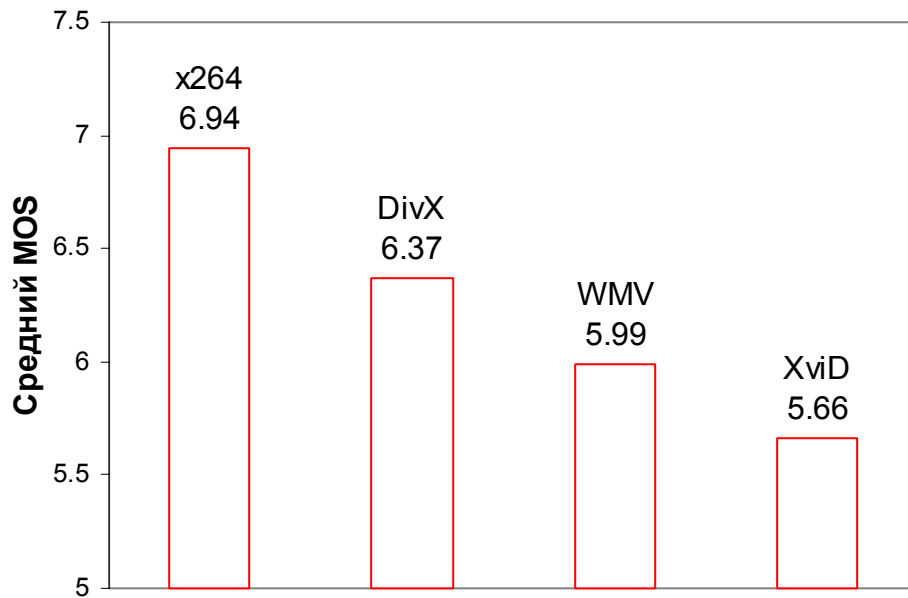


Рисунок 21. Средний MOS для всех кодеков

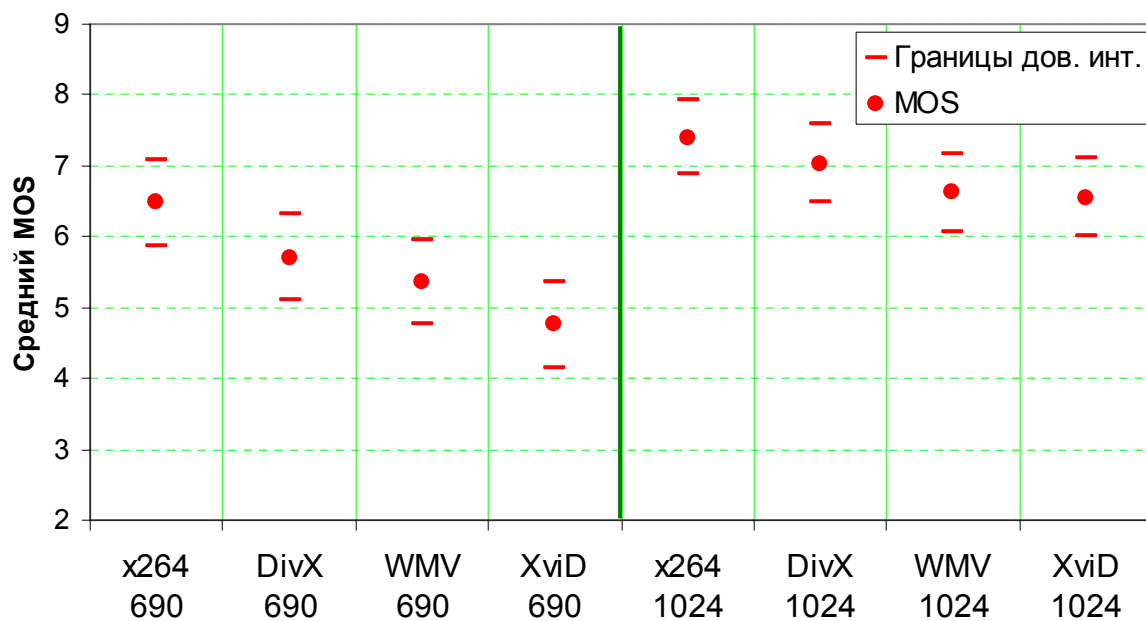


Рисунок 22. Средний MOS для всех кодеков и битрейтов

Графики из предыдущей части сравнения отражают среднее субъективное мнение о кодеках на различных последовательностях и с разными битрейтами. Для того чтобы определить, какое место занял каждый кодек с конкретным битрейтом на всем тестовом наборе, мы применим следующий метод: упорядочим значения MOS для каждой последовательности, расставим рейтинг и просуммируем его (столбец "Overall"). Отсортировав этот столбец, получим, какое место кодек на этом битрейте занял в сравнении (Ref. – несжатая версия последовательности, тоже оценивалась экспертами):

	Battle	Ranch o	Matrix sc.1	Matrix sc.2	Всего	Место
Ref.	1	1	2	1	5	-
x264 1024	2	3	1	2	8	1
DivX 1024	3	4	5	3	15	2
XviD 1024	7	2	3	5	17	3
x264 690	5	6	4	4	19	4
WMV 1024	4	5	6	6	21	5
DivX 690	6	8	7	7	28	6
WMV 690	8	7	9	9	33	7
XviD 690	9	9	8	8	34	8

Очевидно значительное превосходство кодека x264 на обоих битрейтах, причем на 690 kbps он получил более высокую оценку, чем WMV на 1024 kbps. DivX демонстрирует второй результат после x264. Аутсайдеры – кодеки WMV и XviD, на 690 kbps они почти равны, а на 1024 kbps WMV отстает. Плохой результат кодека XviD может быть объяснен [отсутствием деблокинга](#) во время тестирования.

Общие выводы

- Свободно доступный кодек х264 нового стандарта H.264 оказался лучше развитых проприетарных решений.
- Субъективное сравнение является стабильным и продуктивным методом сравнения видеосистем, при условии тщательного соблюдения условий тестирования.
- Результаты субъективного тестирования не всегда могут быть с достаточной точностью предсказаны объективными метриками. Лучшей метрикой на нашем наборе оказалась SSIM.

Литература

1. International Telecommunication Union, "Methodology for the subjective assessment of the quality of television pictures (ITU-R BT.500-11)" (2002)
2. European Broadcasting Union, "SAMVIQ – a New EBU Methodology for Video Quality Evaluations In Multimedia", web site www.ebu.ch
3. Feng Xiao, "DCT-based Video Quality Evaluation", Final Project for EE392J (2000)
4. Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity", IEEE Transactions on Image Processing, vol. 13, no. 4 (2004)
5. VQEG, "Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment" (2000), web site www.vqeg.org
6. MSU Graphics & Media Lab (Video Group), "MSU Perceptual Video Quality Tool" www.compression.ru/video/quality_measure/perceptual_video_quality_tool_en.html
7. MSU Graphics & Media Lab (Video Group), "MSU Video Quality Measurement Tool" www.compression.ru/video/quality_measure/video_measurement_tool_en.html
8. M.H. Prinson, S. Wolf, "The Impact of Monitor Resolution and Type on Subjective Video Quality Testing", NTIA TM-04-412
9. Avery Lee, "VirtualDub", web site <http://www.virtualdub.org>