

СИСТЕМНЫЙ АНАЛИЗ, УПРАВЛЕНИЕ И ОБРАБОТКА ИНФОРМАЦИИ/SYSTEM ANALYSIS,  
MANAGEMENT AND PROCESSING OF INFORMATIONDOI: <https://doi.org/10.60797/IRJ.2025.160.54>

## СРАВНЕНИЕ ОБЪЕКТИВНЫХ ОЦЕНОК НАБОРА МЕТРИК С СУБЪЕКТИВНЫМИ ОЦЕНКАМИ

Научная статья

Токтасынов С.Э.<sup>1,\*</sup>, Калиберда Е.А.<sup>2</sup><sup>1</sup>ORCID : 0009-0007-9538-3487;<sup>2</sup>ORCID : 0000-0003-2796-5913;<sup>1,2</sup> Омский государственный технический университет, Омск, Российская Федерация

\* Корреспондирующий автор (stalkervelikaia[at]gmail.com)

## Аннотация

Субъективная оценка качества является фундаментальной темой исследований в области оценки качества видео, которая в настоящее время привлекает повышенное внимание всего научного сообщества. Несмотря на то, что данный тип оценки является трудновоспроизводимым, он остается до сих пор востребованным. Существует большое количество исследований, по объективной оценке, качества видео, но остается мало понимания того насколько хорошо существующие методы объективной оценки сравнимы с субъективной при измерении воспринимаемого качества. В работе с помощью выбранного набора метрик были исследованы две базы данных субъективных оценок и проведено экспериментальное сравнение результатов объективных и субъективных оценок, в ходе которого были выявлены расхождения результатов, полученных для одних и тех же кадров. На основе полученных результатов сделан вывод о необходимости разработки решения, альтернативного эталонным метрикам качества.

**Ключевые слова:** объективная оценка, субъективная оценка, качество видео, эксперты, база данных.

## COMPARISON OF OBJECTIVE ASSESSMENTS OF A SET OF METRICS WITH SUBJECTIVE ASSESSMENTS

Research article

Toktassynov S.E.<sup>1,\*</sup>, Kaliberda E.A.<sup>2</sup><sup>1</sup>ORCID : 0009-0007-9538-3487;<sup>2</sup>ORCID : 0000-0003-2796-5913;<sup>1,2</sup> Omsk State Technical University, Omsk, Russian Federation

\* Corresponding author (stalkervelikaia[at]gmail.com)

## Abstract

Subjective quality assessment is a fundamental topic of research in the field of video quality assessment, which is currently attracting increased attention from the entire scientific community. Although this type of assessment is difficult to reproduce, it remains in demand. There is a large number of studies on objective video quality assessment, but there is still little understanding of how well existing objective assessment methods compare to subjective ones when measuring perceived quality. In this work, two databases of subjective assessments were examined using a selected set of metrics, and an experimental comparison of the results of objective and subjective assessments was carried out, during which discrepancies in the results obtained for the same frames were identified. Based on the results, it is concluded that there is a necessity to develop a solution that is an alternative to the standard quality metrics.

**Keywords:** objective assessment, subjective assessment, video quality, experts, database.

## Введение

Качество видеоконтента представляет собой набор параметров, определяющих восприятие обработанного видео, человеком, просматривающим данный контент [1]. Оценка качества обязательно проводится на разных этапах обработки видео. Данная оценка может быть субъективной или объективной. Субъективная оценка качества проводится группой экспертов, в соответствии с определенными рекомендациями и методологиями [2], и представляет собой экспертное мнение о качестве видео в заранее определенных тестовых условиях. Несмотря на то, что проведение субъективной оценки сопряжено со значительными трудностями, данный метод остается одним из самых надежных и актуальных методов оценки качества [3], а также является лучшим методом оценки качества в режиме реального времени. Данный тип оценки широко применяется в медиаиндустрии, при обработке и передаче видео, при имитации компьютерного зрения, также используется с целью обеспечения хорошего пользовательского опыта [4].

Объективная оценка качества проводится метриками, которые базируются на математических алгоритмах с использованием показателей, которые пытаются отразить механизмы восприятия зрительной системы человека для сравнения эталонного и искаженного изображения [3]. Одной из проблем использования данного типа оценки является то, что некоторые метрики не учитывают особенности визуального восприятия видеоконтента человеком, а базируются, например, на таких критериях как структура и текстура, которые слабо различимы человеческим зрением, особенно при низком битрейте, когда текстурных и структурных искажений возникает больше.

Ко второй проблеме, по мнению авторов данной статьи, стоит отнести некоторую односторонность оценок качества видеоконтента с помощью существующих объективных метрик. Метрики, как правило, оценивают один параметр качества видео, без учета взаимного влияния разных параметров друг на друга.

Авторами статьи «Набор объективных метрик качества, ориентированных на особенности восприятия видеосигнала человеком» [5] было проведено исследование объективных метрик качества с целью выявления набора метрик наиболее соответствующего субъективной оценке. В результате чего был выделен набор из 3 метрик, включающий в себя PSNR [6], SSIM [7] и DISTS [8].

Задачей данной статьи является объективная оценка с помощью трех обозначенных метрик открытых баз видео с разными типами искажений и сравнение полученных результатов работы метрик из предложенного набора с результатами субъективной оценки на основе их относительных отклонений.

### Основные результаты

Для проведения исследования были выбраны искаженные видео с субъективными оценками из открытых баз данных BVI-FVI [9] и MCL-V [10].

База данных BVI-VFI, содержит 36 эталонных видео в разрешении 1920x1080, обладающих разной частотой кадров (30, 60 и 120 кадров в секунду). Каждое видео обработано с пониженной дискретизацией и интерполяцией пятью методами VFI (Video Frame Interpolation) [11], из чего получено 180 искаженных видеороликов. К данным видам искажений можно отнести:

- 1) Repeat (повторение) — метод, основанный на многократном дублировании последовательности кадров;
- 2) Average (усреднение) — метод, основанный на усреднении кадров, где средний кадр генерируется путем усреднения двух смежных кадров;
- 3) DVF (Deep Voxel Flow) — метод, основанный на глубокой нейронной сети, которая синтезирует видеокадры, извлекая значения вокселей из существующих;
- 4) QVI (Quadratic Video Interpolation) — метод, основанный на многослойной структуре, где присутствуют слои обратного и прямого потока для аппроксимации движения пикселей и создания дополнительных кадров;
- 5) ST-MFNet (Spatio-Temporal Multi-Flow Network) — метод, основанный на пространственно-временной многопоточной архитектуре, с помощью которой происходит прогнозирование дополнительных кадров и последующее их синтезирование.

Указанные методы, генерируя дополнительные видеокадры, позволяют создавать уникальные типы артефактов, такие как, дрожание и размытие, что непосредственно влияет на текстуру и структуру видеофайла.

При субъективной оценке кадров использовалась методология попарной оценки сходства, основанная на выборе пользователем лучшего из двух изображений и оценке степени различия между изображениями [2]. Вывод средних субъективных оценок проводился по методу DMOS (Differential Mean Opinion Score) [12].

База данных MCL-V содержит 12 эталонных видео с разрешением 1920x1080 и частотой кадров 30 кадров в секунду. Каждое видео обработано в сервисах потоковой передачи видео (сжатие [13] и масштабирование [14] изображения) при помощи двух типов искажений с четырьмя уровнями сжатия, что суммарно дает 96 искаженных видео. К данным типам искажений можно отнести:

- 1) Сжатие с использованием стандарта H.264/AVC;
- 2) Сжатие с последующим масштабированием (изменение изображения при помощи каскада операций: понижающей дискретизации, кодирования, а затем изменения размера до исходного разрешения).

При сжатии изображения использовалось 4 уровня сжатия:

- 1) H.264\_1 — метод, основанный на сжатии видео при ограничении пропускной способности до 2000 кбит в секунду;
- 2) H.264\_2 — метод, основанный на сжатии видео при ограничении пропускной способности до 3000 кбит в секунду;
- 3) H.264\_3 — метод, основанный на сжатии видео при ограничении пропускной способности до 4000 кбит в секунду;
- 4) H.264\_4 — метод, основанный на сжатии видео при ограничении пропускной способности до 5000 кбит в секунду.

При сжатии видеофайла появляются такие артефакты сжатия, как, шумы, зернистость, что негативно влияет на текстуру и структуру.

При субъективной оценке кадров использовалась методология категориальной оценки ухудшения качества, основанная на поочередном представлении эталонного и искаженного изображения пользователю с последующей оценкой степени ухудшения качества изображения [2]. Вывод средних субъективных оценок проводился по методу MOS (Mean Opinion Score) [15].

Исследование искажений и изменяемых параметров видеофайла двух баз данных показало, что главным изменяемым параметром является структура и текстура видеофайла. Далее необходимо сравнить результаты субъективной и объективной оценок и оценить результаты сравнения.

Сложность сравнения состоит в том, что каждая метрика имеет свою шкалу оценивания качества видеоконтента, в соответствии с используемым алгоритмом. Поэтому необходимо приведение получаемых оценок к единому виду. Для чего применялась шкала перевода оценок, представленная в таблице 1.

Таблица 1 - Шкала перевода оценок

DOI: <https://doi.org/10.60797/IRJ.2025.160.54.1>

Цветовая идентификация	Субъективная идентификация	Балльная шкала	SSIM	PSNR	DISTS	DMOS	MOS
#D9EAD3	отличное	5	0,93 и выше	38 и выше	0 - 0,21	100	8-10
#FFF2CC	хорошее	4	0,88-0,93	35-38	0,21 - 0,28	75-100	6-8
#C9DAF8	среднее	3	0,84-0,88	33-35	0,28 - 0,33	50-75	4-6
#F4CCCC	низкое	2	0,78-0,84	30-33	0,33 - 0,43	25-50	2-4
#EA9999	плохое	1	0,78 и ниже	30 и ниже	0,43 и выше	0-25	0-2

Для упрощения чтения таблиц, были введены сокращения, представленные в таблице 2.

Таблица 2 - Принятые сокращения

DOI: <https://doi.org/10.60797/IRJ.2025.160.54.2>

Полное	Сокращенное	Полное	Сокращенное
Название видео	Наз.	SSIM	S
Искажение	Иск.	PSNR	P
Метрика	Мет.	DISTS	D
DMOS	DM	MOS	M

Для сравнения результатов объективных и субъективных оценок, из базы данных BVI-VFI был взят набор видео с разрешением 1920x1080 и частотой кадров 60 кадров в секунду. Результаты субъективной оценки видео, в соответствии с принятой шкалой перевода, представлены в таблице 3.

Таблица 3 - Результаты субъективной оценки видео из базы данных BVI-FVI

DOI: <https://doi.org/10.60797/IRJ.2025.160.54.3>

Наз.	Иск.				
	Average	DVF	QVI	Repeat	STM-FNET
Bubblehead (BH)	46,75	80,38	33,04	48,6	37,45
Books (Bo)	57,3	76,11	38,18	57,04	37,82
Bouncyball (BB)	51,36	79,92	38,21	54,17	37,43
Catch_track (CT)	57,01	78,52	45,01	56,39	40,27
Cyclist (Cy)	51,07	82,04	50,04	52,81	37,71
Golf_side (GS)	48,66	81,69	38,3	48,98	40,72
Hamster (Ha)	47,98	78,29	37,47	50,67	38,57
Lampspot (La)	48,52	74,1	34,53	49,67	36,49
Plasma (Pl)	47	71,89	40,95	49,59	30,76
Pond (Po)	38,18	51,22	38,6	38,19	33,49
Sparkler (Sp)	47,7	81,09	65,03	44,45	51,21
Water (Wa)	51,98	69,32	58,78	59,29	51,53

Для выполнения дальнейшего сравнения была проведена объективная оценка эталонных и искаженных видеофайлов базы данных BVI-FVI с помощью метрик (см. табл. 4).

Таблица 4 - Результаты объективной оценки видео из базы данных BVI-FVI

DOI: <https://doi.org/10.60797/IRJ.2025.160.54.4>

Иск.	Average			DVF			QVI			Repeat			STM-FNET		
Мет. Наз.	S	P	D	S	P	D	S	P	D	S	P	D	S	P	D

Иск.	Average			DVF				QVI				Repeat				STM-FNET			
BH	0,966	40,33	0,065	0,929	36,34	0,126	0,979	42,95	0,020	0,957	39,03	0,033	0,980	43,18	0,039				
Bo	0,787	33,03	0,170	0,721	30,20	0,251	0,982	43,91	0,019	0,771	32,83	0,089	0,984	44,41	0,020				
BB	0,978	41,97	0,099	0,955	38,79	0,254	0,983	42,92	0,077	0,973	40,30	0,093	0,985	43,45	0,088				
CT	0,840	35,13	0,188	0,788	33,30	0,403	0,973	43,70	0,054	0,814	34,73	0,084	0,970	43,36	0,146				
Cy	0,852	34,90	0,099	0,785	32,07	0,260	0,976	42,21	0,024	0,832	34,92	0,040	0,979	43,38	0,023				
GS	0,969	40,29	0,053	0,963	37,90	0,083	0,955	39,27	0,045	0,966	39,06	0,031	0,973	41,48	0,037				
Ha	0,934	36,99	0,060	0,917	35,31	0,189	0,961	39,19	0,033	0,961	39,19	0,043	0,964	39,53	0,065				
La	0,935	35,41	0,047	0,844	31,72	0,081	0,951	36,61	0,031	0,905	33,54	0,034	0,959	37,40	0,029				
Pl	0,957	37,34	0,082	0,915	34,91	0,254	0,972	38,54	0,066	0,950	36,23	0,082	0,975	39,07	0,069				
Po	0,983	42,67	0,011	0,982	41,83	0,012	0,975	40,13	0,017	0,982	41,89	0,012	0,984	42,79	0,012				
Sp	0,923	36,78	0,121	0,914	39,64	0,215	0,938	37,52	0,171	0,920	36,12	0,091	0,901	28,06	0,394				
Wa	0,839	35,15	0,161	0,795	34,92	0,262	0,844	35,54	0,220	0,814	34,33	0,103	0,878	36,67	0,442				

Данные, представленные в таблицах 3, 4 принадлежат разным оценочным диапазонам, что затрудняет их сравнение. Для приведения данных к единому диапазону значений необходимо обратиться к таблице 1.

Приведенные данные субъективной и объективной оценок показаны в таблице 5.

Таблица 5 - Приведенные значения субъективной и объективной оценок

DOI: <https://doi.org/10.60797/IRJ.2025.160.54.5>

Наз.	Avarage				DVF				QVI				Repeat				STM-FNET			
Мет. Наз	DM	S	P	D	DM	S	P	D	DM	S	P	D	DM	S	P	D	DM	S	P	D
BH	3	5	5	5	4	4	4	5	2	5	5	5	2	5	5	5	2	5	5	5
Bo	3	2	3	5	4	1	2	4	2	5	5	5	3	1	2	5	2	5	5	5
BB	3	5	5	5	4	5	5	4	2	5	5	5	3	5	5	5	2	5	5	5
CT	3	2	4	5	4	2	3	2	2	5	5	5	3	2	3	5	2	5	5	5
Cy	3	3	3	5	4	2	2	4	3	5	5	5	3	2	3	5	2	5	5	5
GS	2	5	5	5	4	5	4	5	2	5	5	5	2	5	5	5	2	5	5	5
Ha	2	5	4	5	4	4	4	5	2	5	5	5	3	5	5	5	2	5	5	5
La	2	5	4	5	3	3	2	5	2	5	4	5	2	4	3	5	2	5	4	5
Pl	2	5	4	5	3	4	3	4	2	5	5	5	2	5	4	5	2	5	5	5
Po	2	5	5	5	3	5	5	5	2	5	5	5	2	5	5	5	2	5	5	5
Sp	2	4	4	5	4	4	5	4	3	5	4	5	2	4	4	5	3	4	1	2
Wa	3	2	4	5	3	2	3	4	3	3	4	4	3	2	3	5	3	3	4	1

На рисунке 1 представлены графики, построенные по полученным субъективным и объективным оценкам базы данных BVI — FVI.

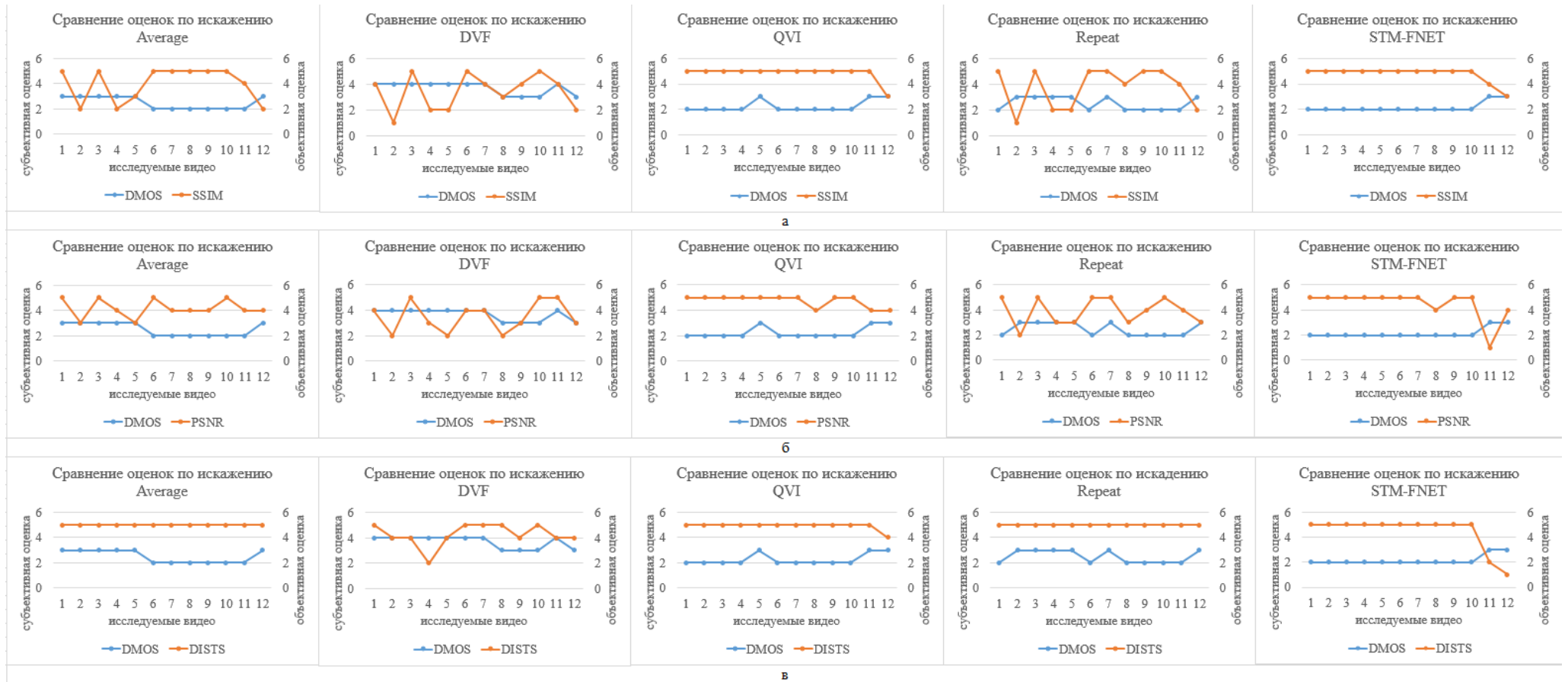


Рисунок 1 - Графики сравнения оценок:  
 а – SSIM; б – PSNR; в – DISTS  
 DOI: <https://doi.org/10.60797/IRJ.2025.160.54.6>

На полученных графиках видно, что результаты объективной и субъективной оценок видео заметно расходятся.

Для определения количественных показателей расхождения двух типов оценок были рассчитаны их относительные отклонения. Данные расчетов представлены в таблице 6.

Таблица 6 - Результаты относительного отклонения (среднее по метрике) в процентах

DOI: <https://doi.org/10.60797/IRJ.2025.160.54.7>

	Average			DVF			QVI			Repeat			STM-FNET		
Мет. Наз.	S от DM	P от DM	D от DM	S от DM	P от DM	D от DM	S от DM	P от DM	D от DM	S от DM	P от DM	D от DM	S от DM	P от DM	D от DM
BH	67	67	67	0	0	25	150	150	150	150	150	150	150	150	150
Bo	33	0	67	75	50	0	150	150	150	67	33	67	150	150	150
BB	67	67	67	25	25	0	150	150	150	67	67	67	150	150	150
CT	33	33	67	50	25	50	150	150	150	33	0	67	150	150	150
Cy	0	0	67	50	50	0	67	67	67	33	0	67	150	150	150
GS	150	150	150	25	0	25	150	150	150	150	150	150	150	150	150
Ha	150	100	150	0	0	25	150	150	150	67	67	67	150	150	150
La	150	100	150	0	33	67	150	100	150	100	50	150	150	100	150
Pl	150	100	150	33	0	33	150	150	150	150	100	150	150	150	150
Po	150	150	150	67	67	67	150	150	150	150	150	150	150	150	150
Sp	100	100	150	0	25	0	67	33	67	100	100	150	33	67	33
Wa	33	33	67	33	0	33	0	33	33	33	0	67	0	33	67
Средн ее	90	75	108	30	23	27	124	119	126	92	72	108	128	129	133

Проведя изучение расчетных данных, было замечено, что среднее отклонение ожидаемых и фактических результатов на всех типах искажений составляет значительно больше 10% (О.У. — оптимальный уровень), что говорит об очень высоком уровне расхождения объективной и субъективной оценок, что визуально отображено на графиках на рисунке 2.

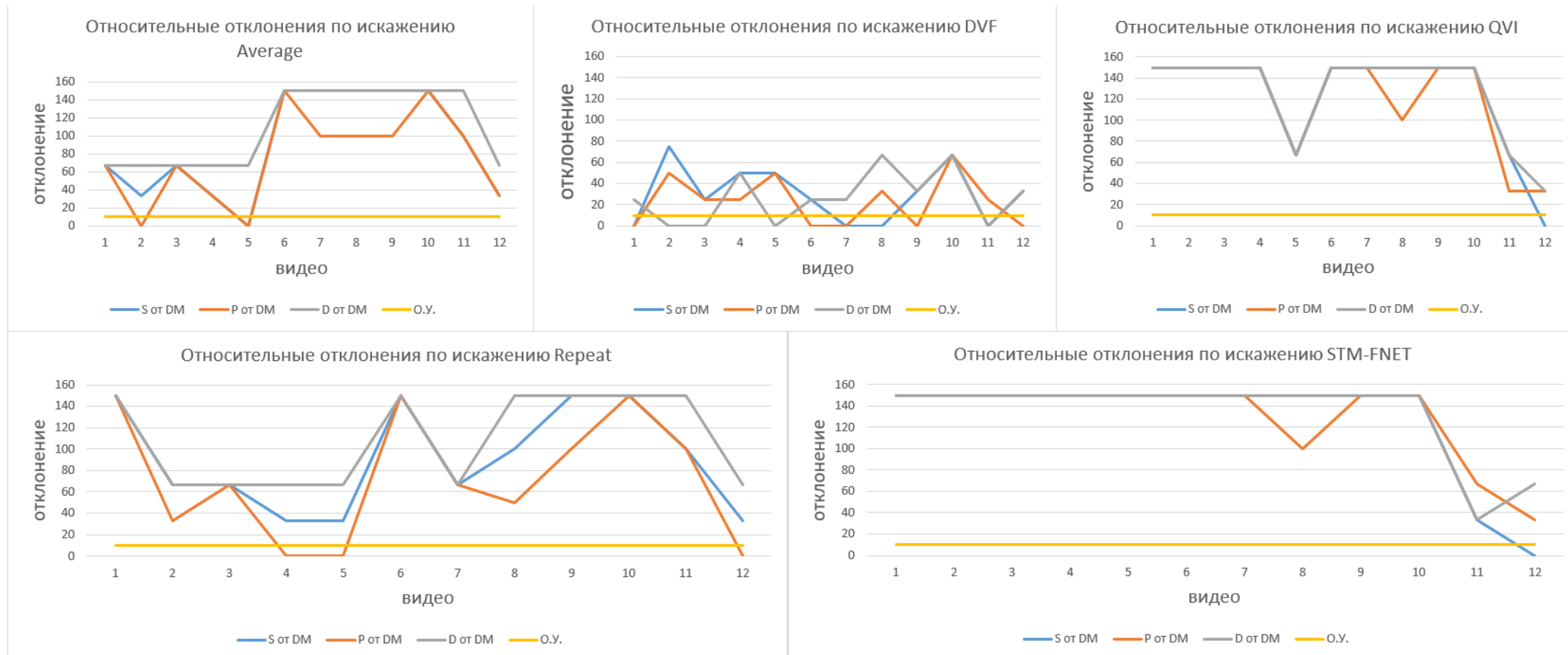


Рисунок 2 - Графики относительного отклонения по искажениям  
DOI: <https://doi.org/10.60797/IRJ.2025.160.54.8>

Проанализировав графики можно сделать вывод, что наименьшие отклонения наблюдаются по искажению DVF, наибольшие — по искажению STM-FNET. Лучше всего сходятся с субъективными оценками по всем типам искажений результаты работы метрики PSNR, в свою очередь, метрика DISTS обладает наименьшей сходимостью, из чего можно сделать вывод, что PSNR больше подходит для применения в случае интерполяционного увеличения кадров.

Следующим этапом был выбран ряд видео с разрешением 1920x1080 и соответствующий этому ряду набор субъективных оценок из базы MCL-V. Результаты субъективной оценки видео приведены в таблице 7.

Таблица 7 - Результаты субъективной оценки видео из базы данных MCL-V

DOI: <https://doi.org/10.60797/IRJ.2025.160.54.9>

Наз.	Иск.			
	264_1	264_2	264_3	264_4
Big Buck Bunny (BB)	0,759	2,603	4,931	6,672
BQ Terrace (BQ)	0,897	3,155	5,362	6,966
Dance Kiss (DK)	0,352	2,593	4,611	6,630
El Fuente B (EB)	0,704	2,852	4,500	6,426
Kimono (KM)	0,796	2,611	4,852	6,926
Seeking (SK)	0,865	3,000	4,904	6,808
Birds in Cage (BC)	1,103	3,431	5,345	7,069
Crowd Run (CR)	1,052	3,155	5,379	6,914
El Fuente A (EA)	0,296	2,870	4,667	6,815
Fox Bird (FB)	1,615	2,885	4,346	6,577
Old Town Cross (OT)	0,923	3,231	5,577	7,077
Tennis (TN)	0,346	2,654	4,519	6,923

Для дальнейшего сравнения была проведена объективная оценка эталонных и искаженных видеофайлов базы данных MCL-V с помощью выбранных ранее метрик (см. табл. 8).

Таблица 8 - Результаты объективной оценки видео из базы данных MCL-V

DOI: <https://doi.org/10.60797/IRJ.2025.160.54.10>

Иск.	264_1			264_2			264_3			264_4		
Мет. Наз.	S	P	D	S	P	D	S	P	D	S	P	D
BB	0,857	33,87	0,170	0,888	35,18	0,142	0,924	37,10	0,107	0,951	39,17	0,107
BQ	0,793	31,69	0,215	0,837	33,40	0,170	0,875	35,30	0,127	0,893	36,53	0,105
DK	0,917	37,54	0,250	0,930	39,14	0,211	0,940	40,50	0,175	0,946	41,57	0,141
EB	0,796	33,15	0,222	0,821	34,21	0,201	0,841	35,00	0,183	0,862	35,95	0,167
KM	0,885	36,57	0,185	0,920	38,76	0,151	0,934	39,85	0,133	0,956	40,99	0,119
SK	0,679	29,96	0,232	0,734	31,09	0,212	0,777	32,09	0,189	0,834	33,76	0,148
BC	0,902	36,66	0,202	0,925	38,82	0,163	0,944	41,08	0,129	0,954	42,48	0,111
CR	0,733	29,25	0,221	0,775	30,26	0,196	0,810	31,18	0,181	0,859	32,76	0,145
EA	0,825	32,97	0,243	0,867	34,91	0,204	0,889	36,05	0,180	0,918	37,79	0,150
FB	0,894	34,33	0,193	0,901	34,87	0,162	0,827	36,79	0,130	0,950	38,91	0,100
OT	0,800	33,45	0,220	0,838	34,81	0,180	0,877	36,77	0,124	0,888	37,57	0,096
TN	0,916	37,89	0,286	0,936	39,80	0,262	0,948	41,27	0,221	0,964	43,53	0,156

Данные, представленные в таблицах 7, 8 принадлежат разным оценочным диапазонам, что затрудняет их сравнение. Для приведения данных к единому диапазону значений необходимо вновь обратиться к таблице 1.

Приведенные данные субъективной и объективной оценки показаны в таблице 9.



Таблица 9 - Приведенные данные субъективной и объективной оценок

DOI: <https://doi.org/10.60797/IRJ.2025.160.54.11>

Иск.	264_1				264_2				264_3				264_4			
Мет. Наз.	M	S	P	D	M	S	P	D	M	S	P	D	M	S	P	D
BB	1	3	3	5	2	4	4	5	3	4	4	5	4	5	5	5
BQ	1	2	2	4	2	2	3	5	3	3	4	5	4	4	4	5
DK	1	4	4	4	2	5	5	4	3	5	5	5	4	5	5	5
EA	1	2	3	4	2	2	3	5	3	3	3	5	4	3	4	5
KM	1	4	4	5	2	4	5	5	3	5	5	5	4	5	5	5
SK	1	1	1	4	2	1	2	4	3	1	2	5	4	2	3	5
BC	1	4	4	5	2	4	5	5	3	5	5	5	4	5	5	5
CR	1	1	1	4	2	1	2	5	3	2	2	5	4	3	2	5
EA	1	2	2	4	2	3	3	5	3	4	4	5	4	4	4	5
FB	1	4	3	5	2	4	3	5	3	2	4	5	4	5	5	5
OT	1	2	3	4	2	3	3	5	3	3	4	5	4	4	4	5
TN	1	4	4	3	2	5	5	4	3	5	5	4	4	5	5	5

На рисунке 3 представлены графики, построенные по результатам полученных субъективных и объективных оценок базы данных MCL-V.

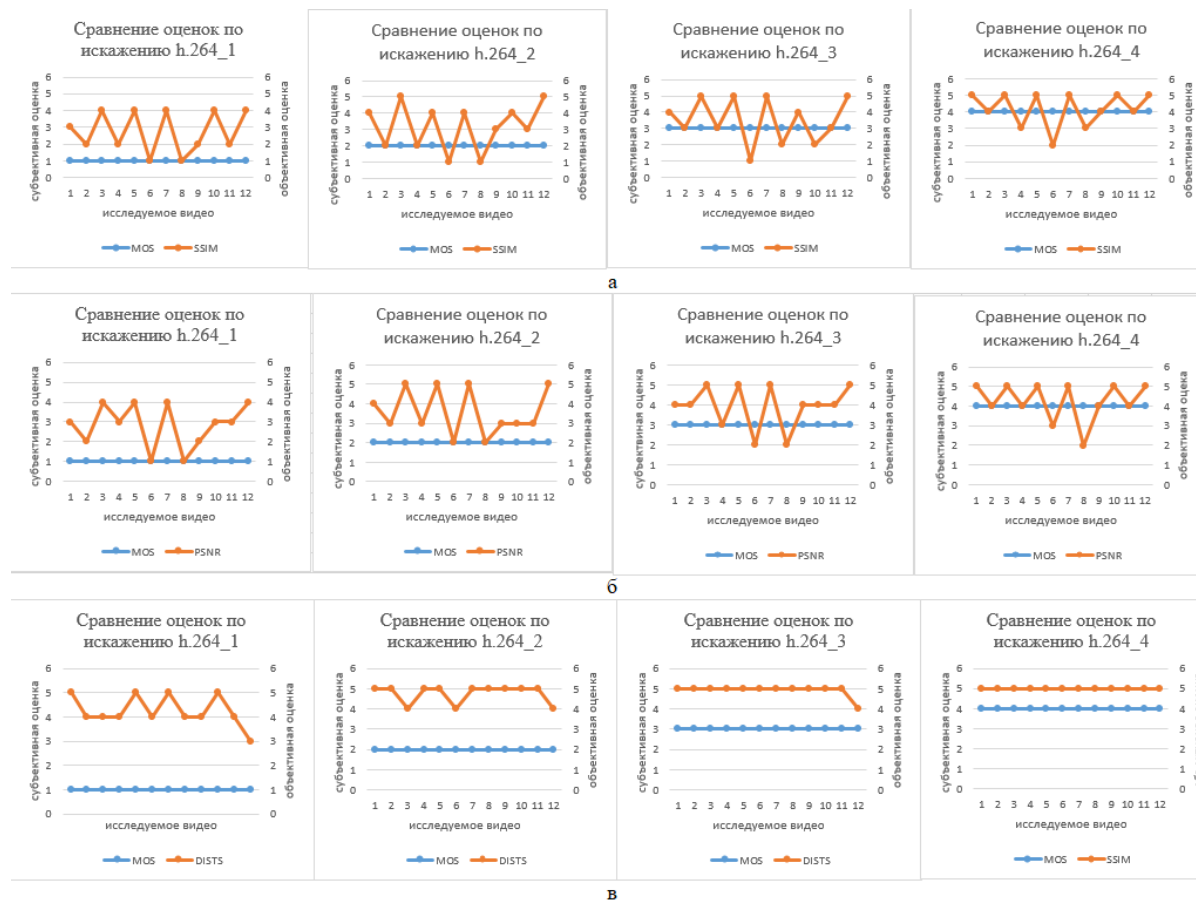


Рисунок 3 - Графики сравнения оценок:  
 а – SSIM; б – PSNR; в – DISTs  
 DOI: <https://doi.org/10.60797/IRJ.2025.160.54.12>

Как и в случае с базой данных BVI – FVI, полученные результаты проведенной объективной оценки качества значительно расходятся с субъективными оценками.

Относительные отклонения расхождения между ожидаемым (субъективная оценка) и фактическим (объективная оценка) результатами представлены в таблице 10.

Таблица 10 - Результаты относительного отклонения (среднее по метрике)

DOI: <https://doi.org/10.60797/IRJ.2025.160.54.13>

Иск.	264_1			264_2			264_3			264_4		
Мет. Наз.	S от M, %	P от M, %	D от M, %	S от M, %	P от M, %	D от M, %	S от M, %	P от M, %	D от M, %	S от M, %	P от M, %	D от M, %
BB	200	200	400	100	100	150	33	33	67	25	25	25
BQ	100	100	300	0	50	150	0	33	67	0	0	25
DK	300	300	300	150	150	100	67	67	67	25	25	25
EA	100	200	300	0	50	150	0	0	67	25	0	25
KM	300	300	400	100	150	150	67	67	67	25	25	25
SK	0	0	300	50	0	100	67	33	67	50	25	25
BC	300	300	400	100	150	150	67	67	67	25	25	25
CR	0	0	300	50	0	150	33	33	67	25	50	25
EA	100	100	300	50	50	150	33	33	67	0	0	25
FB	300	200	400	100	50	150	33	33	67	25	25	25
OT	100	200	300	50	50	150	0	33	67	0	0	25
TN	300	300	200	150	150	100	67	67	33	25	25	25
Среднее	175	183	325	75	79	138	39	42	64	21	19	25

Проведенный анализ расчетных данных, показал, что среднее отклонение ожидаемых и фактических результатов на всех типах искажений составляет более 10% (О.У. — оптимальный уровень), из чего следует, что результаты объективной и субъективной оценки не сходятся. Полученные результаты для наглядности представлены в виде графиков на рисунке 4.

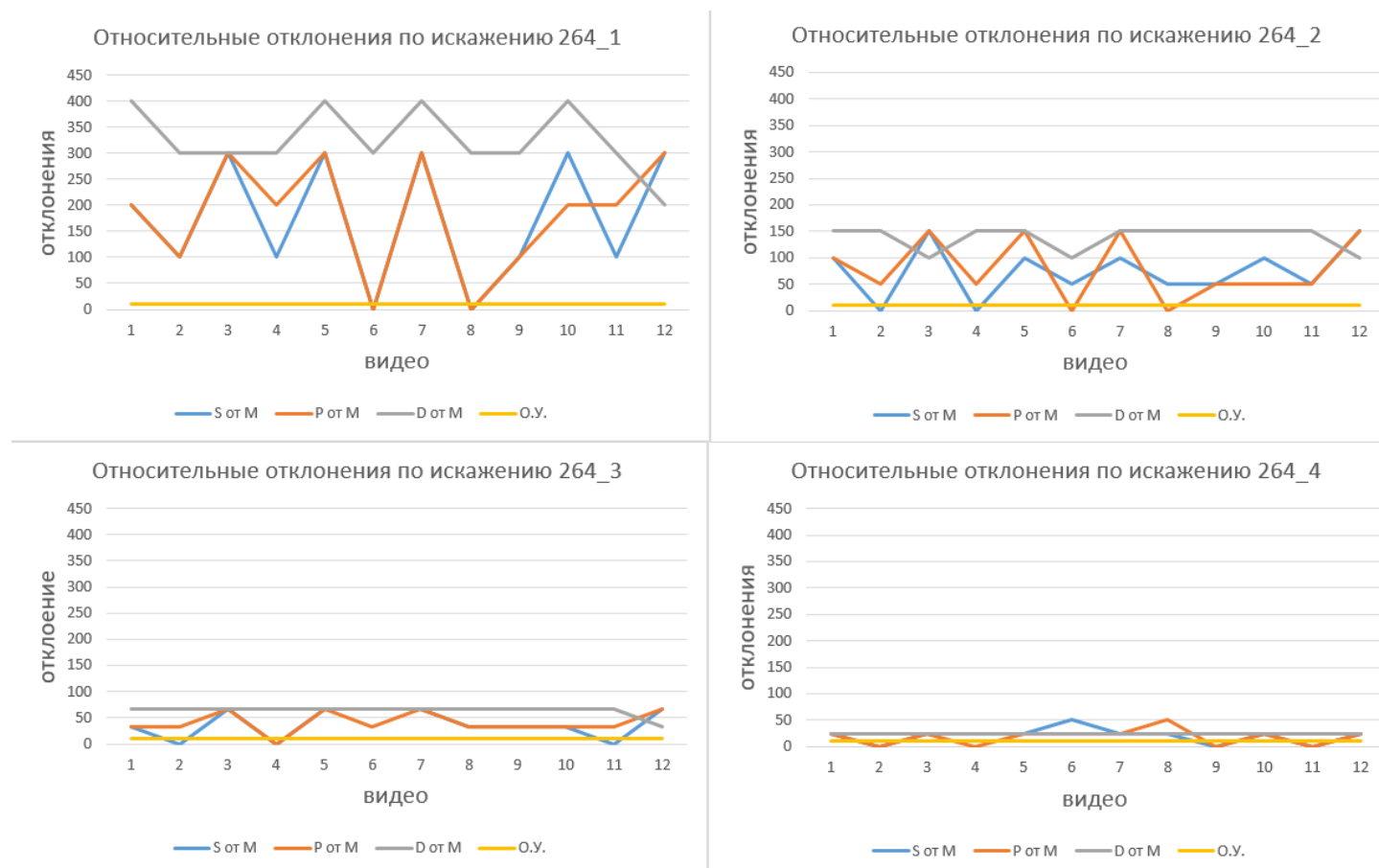


Рисунок 4 - Графики относительного отклонения по искажениям  
DOI: <https://doi.org/10.60797/IRJ.2025.160.54.14>

Проанализировав данный график можно сделать вывод, что наименьшие отклонения наблюдаются по искажению 264\_4, наибольшие отклонения наблюдаются по искажению 264\_1. Также следует отметить, что результаты метрики SSIM наиболее сходятся с субъективными по всем типам искажений, в свою очередь метрика DISTS обладает наименьшей сходимостью, из чего можно сделать вывод, что SSIM наиболее применима в случае оценки сжатого видеофайла.

### Заключение

Оценка качества играет значимую роль при поддержании требований к качеству видео, оценке производительности устройств сбора и отображения видео и оценке систем обработки видео на предмет сжатия. Субъективная оценка является единственным надежным методом оценки качества видео, так как конечным потребителем видео является пользователь и, именно, его оценка является определяющей при оценке качества. Однако из-за сложности и трудоемкости реализации субъективной оценки качества, все чаще применяются объективные методы и очень важно соответствие результатов, полученных при применении разных оценок для одного и того же набора кадров.

Экспериментальное сравнение результатов субъективных и объективных оценок качества видео из баз данных BVI-FVI и MCL-V показало значительное расхождение при различных типах искажений.

При анализе и сравнении работы метрик были получены следующие результаты:

1) Для базы данных BVI-FVI:

– лучшую сходимость показала метрика PSNR на всех типах искажений (самая большая сходимость на искажении DVF).

– худшую сходимость показала метрика DISTS на всех типах искажений (самая плохая сходимость на искажении STM-FNET).

Данный результат можно объяснить тем, при интерполяции появляются такие артефакты как: геометрические искажения, искажения контуров, размытия. Данные типы искажений плохо воспринимаются метрикой DISTS, но хорошо метрикой PSNR.

Расхождения субъективных и объективных оценок при интерполяционных искажениях можно объяснить тем, что человеческий глаз лучше реагирует на данный тип искажений, также следует обратить внимание на то, что с внесением дополнительных кадров происходят временные искажения, которые указанные метрики не видят.

2) Для базы данных MCL-V:

– лучшую сходимость показала метрика SSIM на всех типах искажений (самая большая сходимость на искажении 264\_4).

– худшую сходимость показала метрика DISTS на всех типах искажений (самая плохая сходимость на искажении 264\_1).

Данный результат можно объяснить тем, что при сжатии появляются артефакты сжатия, такие как: блоковые артефакты, потеря градиентов, искажение границ изображений. Данные типы искажений плохо воспринимаются метрикой DISTS, но хорошо метрикой SSIM.

Расхождения субъективных и объективных оценок при искажениях сжатия можно объяснить тем, что человеческий глаз более чувствителен к данным типам искажений, также следует обратить внимание на то, что сжатие может вносить несущественные искажения, которые могут игнорироваться некоторыми метриками, но они видимы для человека.

Получены результаты о сложности получения адекватной оценки качества видео с помощью одной метрики. Для объективной оценки необходим целый набор эталонных метрик, способный оценить изменения в качестве видео, под воздействием различных искажений. Или необходимо решение, альтернативное эталонной оценке качества. Таким решением может стать безэталонная нейросетевая модель.

### Конфликт интересов

Не указан.

### Рецензия

Все статьи проходят рецензирование. Но рецензент или автор статьи предпочли не публиковать рецензию к этой статье в открытом доступе. Рецензия может быть предоставлена компетентным органам по запросу.

### Conflict of Interest

None declared.

### Review

All articles are peer-reviewed. But the reviewer or the author of the article chose not to publish a review of this article in the public domain. The review can be provided to the competent authorities upon request.

### Список литературы / References

1. Min X. Perceptual video quality assessment: a survey / X. Min, H. Duan, W. Sun [et al.] // Sci. China Inf. Sci. — 2024. — № 67. — 211301. — DOI: 10.1007/s11432-024-4133-3
2. Грачева М.А. Субъективная оценка качества статических и видеоизображений: методологический обзор / М.А. Грачева, А.А. Божкова, А.А. Казакова [и др.] // Сенсорные системы. — 2019. — Т. 33, № 4. — С. 287–304
3. Zerman E. Subjective and Objective Quality Assessment for Volumetric Video Compression. Burlingame / E. Zerman, P. Gao, C. Ozcinar [et al.] // Electronic Imaging. — 2019. — Vol. 31. — P. 323.1–323.6 — DOI: 10.2352/ISSN.2470-1173.2019.10.IQSP-323
4. Danier D. A Subjective Quality Study for Video Frame Interpolation / D. Danier, F. Zhang, D. Bull // IEEE International Conference on Image Processing (ICIP). — 2022. — P. 1361–1365. — DOI: 10.48550/arXiv.2202.07727.

5. Токтасынов С.Э. Набор объективных метрик качества, ориентированных на особенности восприятия видеосигнала человеком / С.Э. Токтасынов, Е.А. Калиберда // Международный научно-исследовательский журнал. — 2025. — № 4(154). — DOI 10.60797/IRJ.2025.154.119. — EDN DAPGOW
6. PSNR и SSIM или как работать с изображениями под C // Habr. — 2011. — URL: <https://habr.com/ru/articles/126848/> (дата обращения: 20.05.2025)
7. Nilsson J. Understanding SSIM / J. Nilsson, T. Akenine-Möller. — 2020. — DOI: 10.48550/arXiv.2006.13846
8. Deep Image Structure and Texture Similarity (DISTS) Metric // Github. — URL: <https://github.com/dingkeyan93/DISTS> (accessed: 20.05.2025)
9. Duolikun D. BVI Database / D. Duolikun. — URL: <https://danier97.github.io/BVI-VFI> (accessed: 20.05.2025)
10. MCL-V Database // USC Media Communications Lab. — URL: <https://mcl.usc.edu/mcl-v-database> (accessed: 20.05.2025)
11. Danier D. BVI-VFI: A video quality database for video frame interpolation / D. Danier, Fan Zhang, D. Bull // IEEE Transactions on Image Processing. — 2023. — Vol. 32. — P. 6004-6019. — DOI: 10.1109/TIP.2023.3327912.
12. Hipp D. The Digital Media Overuse Scale (dMOS): A modular and extendible questionnaire for indexing digital media overuse / D. Hipp, E. Blakkley, N. Nipp [et al.] // Technology, Mind, and Behavior. — 2023. — №4. — DOI: 10.1037/tmb0000117
13. Video Compression // Stream. — URL: <https://getstream.io/glossary/video-compression> (accessed: 20.05.2025)
14. Video Scaling // Cloudinary. — URL: <https://cloudinary.com/glossary/video-scaling> (accessed: 20.05.2025)
15. What is a Mean Opinion Score (MOS) // Twilio. — URL: <https://www.twilio.com/docs/glossary/what-is-mean-opinion-score-mos> (accessed: 20.05.2025)

### Список литературы на английском языке / References in English

1. Min X. Perceptual video quality assessment: a survey / X. Min, H. Duan, W. Sun [et al.] // Sci. China Inf. Sci. — 2024. — № 67. — 211301. — DOI: 10.1007/s11432-024-4133-3
2. Gracheva M.A. Sub'ektivnaya ocenka kachestva staticheskikh i videoizobrazhenij: metodologicheskij obzor [Subjective assessment of the quality of static and video images: a methodological review] / M.A. Gracheva, A.A. Bozhkova, A.A. Kazakova [et al.] // Sensornye sistemy [Sensory Systems]. — 2019. — Vol. 33, № 4. — P. 287–304 [in Russian]
3. Zerman E. Subjective and Objective Quality Assessment for Volumetric Video Compression. Burlingame / E. Zerman, P. Gao, C. Ozcinar [et al.] // Electronic Imaging. — 2019. — Vol. 31. — P. 323.1–323.6 — DOI: 10.2352/ISSN.2470-1173.2019.10.IQSP-323
4. Danier D. A Subjective Quality Study for Video Frame Interpolation / D. Danier, F. Zhang, D. Bull // IEEE International Conference on Image Processing (ICIP). — 2022. — P. 1361–1365. — DOI: 10.48550/arXiv.2202.07727.
5. Toktasynov S.E. Nabor ob'ektivnykh metrik kachestva, orientirovannykh na osobennosti vospriyatiya videosignala chelovekom [A set of objective quality metrics focused on the peculiarities of human perception of a video signal] / S.E. Toktasynov, E.A. Kaliberda // Mezhdunarodnyj nauchno-issledovatel'skij zhurnal [International Research Journal]. — 2025. — № 4(154). — DOI 10.60797/IRJ.2025.154.119. — EDN DAPGOW [in Russian]
6. PSNR i SSIM ili kak rabotat' s izobrazheniyami pod S [PSNR and SSIM or how to work with images from C] // Habr. — 2011. — URL: <https://habr.com/ru/articles/126848/> (accessed: 20.05.2025) [in Russian]
7. Nilsson J. Understanding SSIM / J. Nilsson, T. Akenine-Möller. — 2020. — DOI: 10.48550/arXiv.2006.13846
8. Deep Image Structure and Texture Similarity (DISTS) Metric // Github. — URL: <https://github.com/dingkeyan93/DISTS> (accessed: 20.05.2025)
9. Duolikun D. BVI Database / D. Duolikun. — URL: <https://danier97.github.io/BVI-VFI> (accessed: 20.05.2025)
10. MCL-V Database // USC Media Communications Lab. — URL: <https://mcl.usc.edu/mcl-v-database> (accessed: 20.05.2025)
11. Danier D. BVI-VFI: A video quality database for video frame interpolation / D. Danier, Fan Zhang, D. Bull // IEEE Transactions on Image Processing. — 2023. — Vol. 32. — P. 6004-6019. — DOI: 10.1109/TIP.2023.3327912.
12. Hipp D. The Digital Media Overuse Scale (dMOS): A modular and extendible questionnaire for indexing digital media overuse / D. Hipp, E. Blakkley, N. Nipp [et al.] // Technology, Mind, and Behavior. — 2023. — №4. — DOI: 10.1037/tmb0000117
13. Video Compression // Stream. — URL: <https://getstream.io/glossary/video-compression> (accessed: 20.05.2025)
14. Video Scaling // Cloudinary. — URL: <https://cloudinary.com/glossary/video-scaling> (accessed: 20.05.2025)
15. What is a Mean Opinion Score (MOS) // Twilio. — URL: <https://www.twilio.com/docs/glossary/what-is-mean-opinion-score-mos> (accessed: 20.05.2025)