

## РАЗРАБОТКА МЕТОДА ОЦЕНКИ ВИЗУАЛЬНОГО КАЧЕСТВА ВИДЕО, ПЕРЕДАВАЕМОГО ЧЕРЕЗ СИСТЕМЫ ВИДЕОКОНФЕРЕНЦСВЯЗИ

*Рахманов Михаил Алексеевич*

*Студент*

*Факультет ВМК МГУ имени М. В. Ломоносова, Москва, Россия*

*E-mail: mihail.rakhmanov@graphics.cs.msu.ru*

*Научный руководитель — Звездаков Сергей Васильевич*

Задача оценки визуального качества видео — это задача измерения и анализа видеосигнала с целью определения его визуального качества для зрителя. Эталонном при оценке визуального качества видео являются субъективные оценки контента человеком.

В последние годы количество видео, передаваемого через системы видеоконференцсвязи, сильно возросло. Существующие методы оценки качества видео не учитывают особенности контента, передаваемого через системы видеоконференцсвязи. Например, частое наличие текста в видео. Это снижает корреляцию с субъективными оценками.

В данной работе предлагается метод основанный на распознавании текста в видео. Алгоритм распознает текст в исходном и искаженном видео и предсказывает качество искаженного видео исходя из схожести распознанных текстов. Входными данными для алгоритма являются исходное видео и видео, переданное через системы видеоконференцсвязи.

На первой стадии предлагаемого подхода выделяются участки с текстом на кадрах исходного и искаженного видео. Для этого применяется нейросеть DBNet++[1]. Далее происходит распознавание текста на выделенных участках с помощью нейросети ABINet[2].

На второй стадии происходит подсчет расстояний Хемминга между текстами, получившимися после распознавания. Далее считается взвешенное среднее расстояний Хемминга для всех областей с текстом. Итоговая оценка вычисляется как среднее арифметическое оценок для каждого кадра. Общая схема функционирования приведена на Рис. 1.

Экспериментальная оценка показала превосходство предложенного подхода в сравнении с существующими аналогами в субъективном сравнении на наборе данных, состоящем из контента, передаваемого через системы видеоконференцсвязи. Таблица корреляций с

субъективными оценками для предложенного метода, PSNR, SSIM и VMAF[3] приведена в Таблице 1.

### Иллюстрации



Рис 1. Общая схема предложенного метода.

Video	Proposed method	PSNR	SSIM	VMAF
SCI1	0,74	-0,20	-0,18	-0,17
SCI2	0,84	-0,13	-0,05	-0,08
SCI3	0,73	-0,04	-0,12	0,17
SCI4	0,94	0,35	0,76	0,74
SCI5	0,77	0,79	0,81	0,78
SCI6	0,77	-0,18	-0,28	0,50
SCI7	0,83	0,44	0,45	0,40
SCI8	0,71	0,07	0,58	0,59
SCI9	0,92	0,67	0,69	0,91
Mean	0,81	0,20	0,29	0,43

Таблица 1. Корреляция Спирмана между метриками и субъективными оценками на наборе данных из 9 видео.

### Литература

1. Minghui L. Real-Time Scene Text Detection with Differentiable Binarization and Adaptive Scale Fusion // In IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, P. 919–931.
2. Shancheng F. Read Like Humans: Autonomous, Bidirectional and Iterative Language Modeling for Scene Text Recognition // In IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, P. 7094–7103.
3. Zhi L., Aaron A., Katsavounidis I., Moorthy A. and Manohara M. Toward A Practical Perceptual Video Quality Metric // In <https://netflixtechblog.com/toward-a-practical-perceptual-video-quality-metric-653f208b9652>.