

**РАЗРАБОТКА НЕЙРОСЕТЕВОГО МЕТОДА  
ПОКАДРОВОЙ СИНХРОНИЗАЦИИ ВИДЕО С  
ГЕОМЕТРИЧЕСКИМИ И ЦВЕТОВЫМИ  
ИСКАЖЕНИЯМИ**

*Дремин Михаил Витальевич*

*Студент*

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

*E-mail: mikhail.dremin@graphics.cs.msu.ru*

*Научный руководитель — Куликов Дмитрий Леонидович*

Синхронизация видеопоследовательностей является одной из важных операций предобработки видеоданных для их последующего использования в других задачах, связанных с видео, таких, как оценка качества видео, распознавание действий, выявление копий. В зависимости от накладываемых ограничений на входные последовательности результатами синхронизации могут являться: найденный кадровый сдвиг начала одного видео относительно другого; найденные сдвиги конца и начала (если видеопоследовательности имеют разную частоту кадров); карта сопоставления кадров одного видео на соответствие кадрам другого видео (если в видео присутствуют временные искажения, такие как потеря кадра, дублирование кадра). Также на видео могут быть наложены различные искажения: геометрические, временные, цветовые, которые значительно усложняют процесс детектирования соответствующих кадров.

Целью данной работы является создание алгоритма покадровой синхронизации видео устойчивого к наличию в последовательностях временных, цветовых и геометрических искажений. Также рассматриваются различные искажения, которые могут возникнуть в видео в процессе ручной обработки или передачи через сеть, и выделяются наиболее популярные из них для создания набора данных, позволяющего оценить точность синхронизации.

Для определения степени соответствия кадров друг другу используются контентные и временные представления кадров, полученные с помощью нейросетевых архитектур ResNet-50 [1] и RAFT-GMA [2] соответственно. С помощью RAFT-GMA [2] рассчитываются карты оптического потока для получения информации о движении в кадре. Затем сам кадр и его оптический поток подаются в сиамскую сеть с архитектурой ResNet-50 [1]. Конечные представления получают путем конкатенации выходов промежуточных слоев каждой ветви [3]. Смешивание данных характеристик позволяет

получить уникальные представления для каждого кадра, из которых строится матрица схожести кадров друг другу путем скалярного произведения каждой пары векторов представлений и нахождения расстояния между ними. С помощью алгоритма Дейкстры по полученной матрице находится путь с наибольшим значением — итоговый путь синхронизации.

Для оценки качества работы модели используется метрика  $F_1$  score. Проведенное тестирование на собранном наборе данных показало, что предложенная архитектура может успешно использоваться для решения поставленной задачи (метод достигает 0.84 по  $F_1$  score) [4].

### Литература

1. He K. Deep residual learning for image recognition // In Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, P. 770–778.
2. Jiang S. Learning to estimate hidden motions with global motion aggregation // In Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, P. 9772–9781.
3. Kordopatis-Zilos G. Visil: Fine-grained spatio-temporal video similarity learning // In Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, P. 6351–6360.
4. Страница проекта «MSU Video Alignment and Retrieval Benchmark Suite»: <https://videoprocessing.ai/benchmarks/aligners.html>