

## НЕЙРОСЕТЕВОЕ ВЫЧИСЛЕНИЕ ДОВЕРИЯ ДЛЯ МЕТОДОВ ОЦЕНКИ КАРТ ДИСПАРАНТНОСТИ В ВИДЕО

*Малышев Кирилл Владимирович*

*Студент*

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

*E-mail: kirill.malyshev@graphics.cs.msu.ru*

*Научный руководитель — Ватолин Дмитрий Сергеевич*

Каждый кадр стереоскопического видео состоит из двух изображений — левого и правого ракурса. Классической задачей компьютерного зрения является восстановление информации о картах диспарантности таких кадров через сопоставление ракурсов. Фактически требуется восстановить геометрическую конфигурацию сцены, ее трехмерную структуру.

В последние годы глубокое обучение все больше вытесняет другие подходы при разработке методов построения карт доверия к методам сопоставления ракурсов [1, 2]. Большинство из них в качестве входных данных используют только карты диспарантности и соответствующие им RGB-изображения ракурсов. Однако на основе стереоскопического кадра и по результатам сопоставления можно получить и другие данные. Мы решили исследовать этот вопрос и выбрать такие входные параметры, на которых наша модель будет работать качественнее всего.

Предлагается нейросетевой метод построения карт доверия к методам сопоставления ракурсов, который не принимает на вход ни карты диспарантности, ни исходный кадр, а использует другие виды входных данных: карту согласованности диспаритета, RGB-дисперсию кадра, ошибку сопоставления и разность между левым кадром и приведенным к нему с помощью диспарантности правым кадром.

Еще одной отличительной особенностью нашего метода является то, что предсказывается непрерывное доверие в диапазоне от 0 до 1, что позволяет более точно предсказать достоверность работы метода сопоставления ракурсов.

Для создания истинных карт доверия, на которых обучалась модель, использовался синтетический набор стереоскопических данных Sintel [3]. Архитектура нейронной сети представляет из себя кодировщик-декодировщик, в котором слой из деподировщика связываются со слоями из кодировщика (рис. 1). Модель принимает на

вход данные с разрешением  $256 \times 512$ , а кадры в наборе данных Sintel имеют разрешение  $436 \times 1024$ . Во время обучения подаваемые на вход примеры вырезались случайным образом из случайной пары последовательных кадров. Таким образом достигалось увеличение количества возможных уникальных примеров для обучения, которые могли подаваться на вход сети, так как сдвиг, при котором происходила обрезка примера, выбирался случайным образом. Проверка работы модели с помощью кросс-валидации показала среднюю ошибку (MSE) между предсказанными на тестовой выборке и истинными картами доверия, равную 0.034.

### Иллюстрации

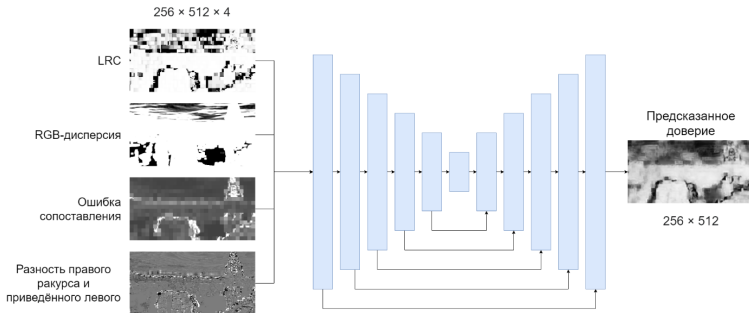


Рис. 1. Архитектура модели

### Литература

1. Fu, Z. and Ardabilian M. and Stern G. Stereo matching confidence learning based on multi-modal convolution neural networks // In Communications, Signal Processing, and Systems, 2019, С. 69–81.
2. Tosi, F. and Poggi, M. and Benincasa, A. and Mattocchia, S. Beyond local reasoning for stereo confidence estimation with deep learning // In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 2018, С. 319–334.
3. Butler D. J. and Wulff J. and Stanley G. B. and Black M. J. A naturalistic open source movie for optical flow evaluation // In Proceedings of the European Conference on Computer Vision (ECCV), Florence, Italy, 2012, С. 611–625.