УДК 004.627

СЖАТИЕ ИЗОБРАЖЕНИЯ ЛИЦА С ИСПОЛЬЗОВАНИЕМ ПЕРЕНОСА ГЕОМЕТРИИ НА ОСНОВЕ КЛЮЧЕВЫХ ТОЧЕК ДЛЯ СИСТЕМ ВИДЕОКОНФЕРЕНЦСВЯЗИ

Шутов В.В.

(Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики)

Научный руководитель — к.т.н., вед. н.с. Беляев Е.А.

(Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики)

В последние годы объем сетевого трафика, используемого в видеоконференцсвязи, растет из года в год. Например, согласно отчетам Cisco, ежегодный рост трафика IP-видео составляет порядка 23%. Это приводит к тому, что сервисы видеотрансляций сталкиваются с проблемой нехватки пропускной способности. Традиционно, для сжатия видео в видеоконференциях используются такие кодеки как VP9 (например, в приложениях VK, Zoom) или H.264/AVC (например, в приложении Skype). Данные кодеки для сжатия используют межкадровую и внутрикадровую избыточность видеоданных. При этом, модель источника видеоинформации, характерная для видеоконференций, является частным случаем, в котором на относительно неподвижном фоне отображается лицо человека (как правило, одного). Эту особенность можно использовать для повышения коэффициента сжатия передаваемых данных.

В настоящей работе рассматривается подход, который позволяет в режиме реального времени аппроксимировать мимику лица используя ключевые точки. В этом подходе первый кадр кодируется традиционным кодеком видеоинформации, например, H.264/AVC. Для второго и следующего кадров вычисляется и передается относительно небольшое количество ключевых точек. Декодер, используя принятый ключевой кадр и ключевые точки, генерирует новый кадр при помощи нейронной сети. В случае смены сцены вместо ключевых точек снова передаётся ключевой кадр.

Известные аппроксимирующие мимику подходы, такие как First Order Motion Model или FSGAN, основаны на генеративно-состязательных нейронных сетях и поиске ключевых точек при помощи библиотеки Dlib. Однако, данные подходы не могут работать в режиме реального времени на большом кадровом разрешении и требуют мощной видеокарты, которая отсутствует у большинства пользователей. Упомянутые проблемы были решены следующим образом:

1. Алгоритм детекции лица и поиска ключевых точек был заменен на связку BlazeFace и Google FaceMesh, которые могут работать в режиме реального времени на большинстве мобильных устройств.
2. Топология нейронной сети для переноса геометрии была модифицирована следующим образом:

a) Были убраны BatchNorm слои. В ходе экспериментов было выявлено, что это так же улучшает качество работы сети.

b) Использован алгоритм Variational Dropout для регуляризации параметров сети. Так же, посредством данного алгоритма получилось уменьшить число параметров сети, что также привело к снижению вычислительной сложности.

c) Для увеличения глубины сети и снижения вычислительной сложности были добавлены SE/Dense блоки.

В результате проделанной работы был получен алгоритм, позволяющая аппроксимировать мимику по ключевым точкам в режиме реального времени на процессорах Apple A13 Bionic и ему эквивалентных.

Шутов В.В. (автор) Подпись

Беляев Е.А. (научный руководитель) Подпись