

ПОЛУАВТОМАТИЧЕСКАЯ МОДЕЛЬ ВНИМАНИЯ И ЕЕ ПРИМЕНЕНИЕ ДЛЯ СЖАТИЯ ВИДЕО

Гитман Юрий Александрович

Студент

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

E-mail: ygitman@graphics.cs.msu.ru

Моделирование визуального внимания — это перспективный подход для того, чтобы улучшить качество многих существующих приложений по обработке изображений и видео, таких как, сжатие [1], сопоставление [2], оценка качества [3], и перепланирование [4]. Но каждое из этих приложений требует высокого качества предсказаний, куда будут смотреть пользователи.

К сожалению, недавно проведенное сравнение [5], показало, что большая часть из разработанных моделей визуального внимания не работает даже на уровне простой модели выбирающей центр изображения. Хотя такая модель полностью независима от содержания исходного видео. На самом деле, даже оптимальным образом выбранная линейная комбинация лучшей модели и модели выбирающей центр показывает улучшение только на 0.037 AuROC (area under receiver operating characteristic) по сравнению с последней. Для сравнения, модель выбирающая центр показывает улучшение на 0,28 AuROC против шума «соль и перец».

Намного лучшее качество может быть получено с использованием аппаратуры записывающей движения глаз, но стоимость такого решения и необходимое время оказываются неприемлемо высоки для коммерческого применения.

Мы предлагаем компромисс между этими двумя подходами: полуавтоматическую модель визуального внимания. Предложенное решение использует точки фиксации взгляда, собранные по записям о движении глаз только одного наблюдателя, после чего применяется автоматический алгоритм. Такая постобработка позволяет нам увеличить надежность первоначальных точек фиксации в пространстве и времени. Предложенный метод опирается на способность кратковременной памяти сохранять информацию о структуре окружающего пространства во времени. Мозг использует эту информацию для того, чтобы планировать взаимодействие с внешним миром, в частности для того, чтобы планировать движения глаз [6]. Предложенный компромисс остается разумным в терминах потраченных времени и денег, так как требует запись движений глаз только одного челове-



Рис. 1: Карты внимания, предсказанные различными методами.

ка (не десятков, как обычно), и, благодаря предложенному методу постобработки, предсказательная способность этих данных может быть значительно улучшена.

Результаты объективного сравнения показали, что предложенный алгоритм превосходит существующие автоматические модели визуального внимания и увеличивает коэффициент схожести [5] между траекторией одного наблюдателя и траекторией 50-и на 0,012.

Высокое качество предсказаний предложенного метода делает возможным использование его в приложениях; В частности, мы описали, как разработанный алгоритм может быть применен для того, чтобы сжимать видео с учетом распределения внимания зрителей.

Новизна работы

- База данных с записями о траекториях движений глаз (full-HD стерео видео, 50 человек, 19.760 кадров)
- Концепция полуавтоматического подхода
- Простая, но эффективная модель визуального внимания, основанная на полуавтоматической концепции и кусочной непрерывности визуального внимания во времени
- Методология и результаты объективного сравнения предложенного алгоритма и 13-и других распространенных подходов
- Разработка алгоритма сжатия с использованием карт внимания на основе распространенного кодека x264 [7]

Иллюстрации

Литература

1. C. Guo and L. Zhang, “A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression,” *IEEE Transactions on Image Processing (TIP)*, vol. 19, no. 1, p. 185–198, 2010. doi: 10.1109/TIP.2009.2030969



Рис. 2: Результаты предложенного подхода к сжатию видео и кодека x264 с одинаковым битрейтом (1500 kbps).

2. A. Shrivastava, T. Malisiewicz, A. Gupta, and A. A. Efros, “Data-driven visual similarity for cross-domain image matching,” *ACM Transactions on Graphics (TOG)*, vol. 30, no. 6, p. 154:1–154:10, 2011. doi: 10.1145/2070781.2024188
3. D. Culibrk, M. Mirkovic, V. Zlokolica, M. Pokric, V. Crnojevic, and D. Kukolj, “Salient motion features for video quality assessment,” *IEEE Transactions on Image Processing (TIP)*, vol. 20, no. 4, p. 948–958, 2011. doi: 10.1109/TIP.2010.2080279
4. S. Goferman, L. Zelnik-Manor, and A. Tal, “Context-aware saliency detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 34, no. 10, p. 1915–1926, 2012. doi: 10.1109/TPAMI.2011.272
5. T. Judd, F. Durand, and A. Torralba, “A benchmark of computational models of saliency to predict human fixations,” *Computer Science and Artificial Intelligence Lab, Massachusetts Institute of Technology, Tech. Rep.*, 2012.
6. M. P. Aivar, M. M. Hayhoe, C. L. Chizk, and R. E. B. Mruzek, “Spatial memory and saccadic targeting in a natural task,” *Journal of Vision*, vol. 5, no. 3, 2005.
7. x264 software video encoder, <http://www.videolan.org/developers/x264.html>.