

Вписывание баннеров рекламного характера в видео методами компьютерного зрения и улучшение качества замены баннера

Осадчий М.Д.

Научный руководитель – к.т.н., доцент Ефимова В. А.

Университет ИТМО

maxosad@mail.ru

Введение.

В современном видеопроизводстве, включая рекламную индустрию и постпродакшн, возникает потребность в реалистичной интеграции визуальных элементов из различных источников в единый видеоряд. Существующие решения часто используют фрагментированный подход, где геометрическое выравнивание, цветокоррекция и обработка окклюзий выполняются изолированно, что приводит к накоплению ошибок и снижению качества итогового продукта. Разработанная система предлагает интегрированный фреймворк для видеокомпози́тинга, объединяющий нейронную гармонизацию на основе Vision Transformer с адаптивным механизмом обработки окклюзий. В рамках алгоритма применяется метод гармонизации изображения, целью которой является оптимизация вставленной области для ее визуально гармоничного сочетания с фоном. Алгоритм выполняет геометрическое выравнивание с помощью особых точек, нейросетевую гармонизацию вставленного изображения и интеллектуальную обработку перекрытий с объектами переднего плана. В результате формируется видеопоток, в котором целевой объект выглядит реалистично и органично вписывается в общий контекст сцены.

Основная часть.

С помощью математических моделей и алгоритмов решаются задачи геометрической регистрации, гармонизации и обработки окклюзий. Для геометрической регистрации и композитинга кадров положение вставляемого баннера на неразмеченных кадрах определяется путём сопоставления текущего кадра с ближайшими размеченными ключевыми кадрами с использованием алгоритма поиска особых точек SIFT [1] и их совмещения алгоритмом RANSAC [2]; на основе полученных соответствий вычисляются матрицы гомографии, обеспечивающие перспективно-корректную вставку изображения. Задача гармонизации с обеспечением временной согласованности решается путём однократной гармонизации на левом ключевом кадре временного фрагмента с помощью модели PCTNet-ViT [3] на базе Vision Transformer с последующим геометрическим преобразованием гармонизированного изображения для всех остальных кадров фрагмента, что устраняет мерцание и гарантирует временную стабильность. Адаптивная обработка окклюзий выполняется с помощью алгоритма GrabCut [5] с временной пропагацией: маска окклюзии с предыдущего кадра проецируется на текущий с помощью гомографии и используется для инициализации сегментации, обеспечивая естественную упорядоченность по глубине и плавность переходов при перекрытиях. В рамках исследования проведено сравнение трёх моделей гармонизации (PCTNet-ViT [3], Duso [4], INR [6]) по метрикам SSIM и PSNR в полнокадровом режиме. Результаты (см. таблицу) показывают, что PCTNet-ViT даёт средний относительный прирост PSNR на +0.968 дБ и SSIM на +0.00118 по сравнению с усреднённым значением по всем моделям, что подтверждает эффективность выбранного подхода. Для вычислительной эффективности используется архитектура параллельной обработки, разделяющая видео

на независимые фрагменты по ключевым кадрам, что позволяет задействовать пул рабочих процессов и сохранять согласованность на границах фрагментов.

Выводы.

В рамках исследования был разработан интегрированный алгоритм для замены баннеров в видеоматериалах, объединяющий геометрическую регистрацию, нейронную гармонизацию и обработку окклюзий. В результате сравнительного анализа трех моделей гармонизации была обоснована эффективность выбора PCTNet-ViT. Проведены экспериментальные исследования, подтверждающие высокое визуальное качество результатов (средние значения SSIM 0.79 и PSNR 20.9 дБ на полных кадрах). Данная работа может быть полезна организациям, занимающимся видеопроизводством, маркетингом, рекламой, а также в областях дополненной реальности и виртуального продакшна.

Литература.

1. Lowe D. G. Object recognition from local scale-invariant features // Proceedings of the International Conference on Computer Vision. – 1999. – Т. 2. – С. 1150–1157.
2. Fischler M. A., Bolles R. C. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography // Communications of the ACM. – 1981. – Vol. 24, No. 6. – P. 381–395.
3. Cong W. et al. PCTNet-ViT: Vision Transformer for Video Harmonization // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. – 2022.
4. Tan L. et al. Deep image harmonization in dual color spaces // Proceedings of the 31st ACM International Conference on Multimedia. – 2023. – С. 2159–2167.
5. Rother C., Kolmogorov V., Blake A. GrabCut: Interactive foreground extraction using iterated graph cuts // ACM Transactions on Graphics. – 2004. – Т. 23, № 3. – С. 309–314.
6. Chen J. et al. Dense pixel-to-pixel harmonization via continuous image representation // IEEE Transactions on Circuits and Systems for Video Technology. – 2023.

Автор _____/Осадчий М. Д.
Научный руководитель _____/Ефимова В.А.