

Метод создания связанных сцен из художественного видео

Студент: Домнин Егор Олегович, ИУ7-43М

Руководитель: Рудаков Игорь Владимирович

Актуальность работы

Создание связных сцен является промежуточным этапом анализа видео, его результаты могут использоваться для:

- упрощения поиска отдельных частей видео, например, вступительных заставок и титров
- упрощения навигации по длительным видео
- созданию оглавления

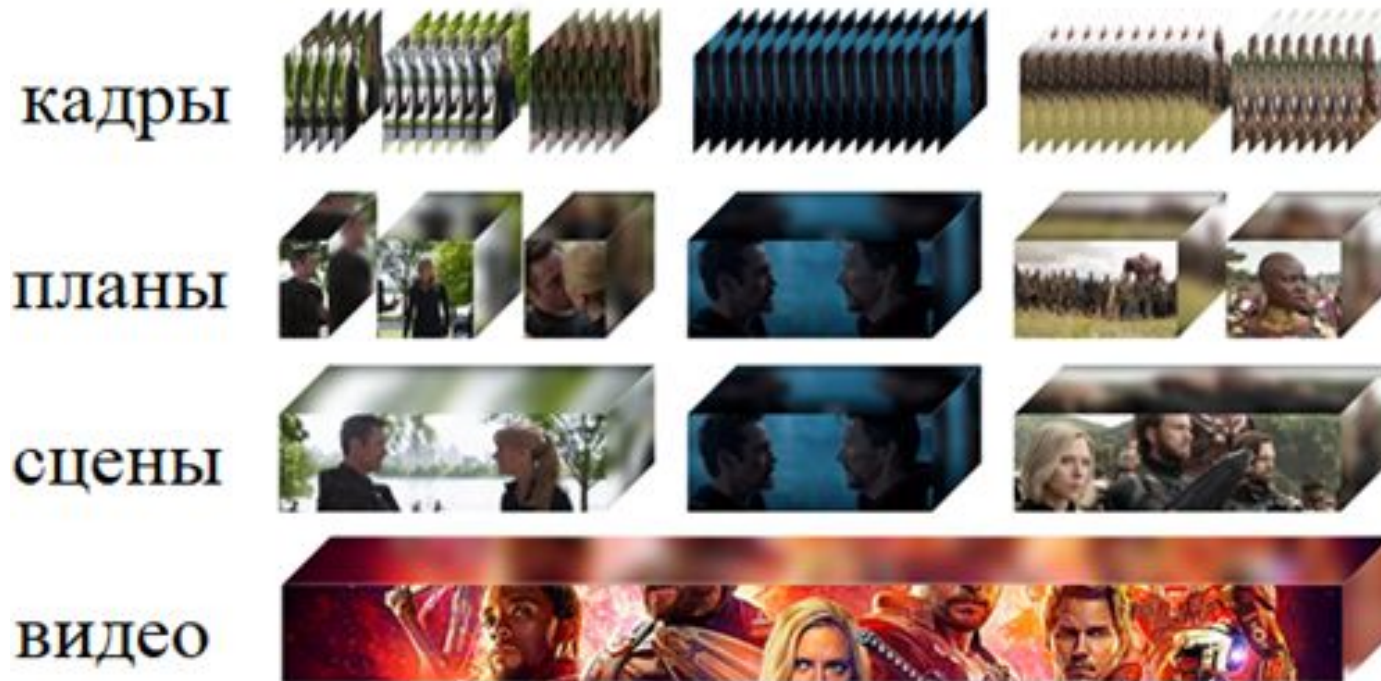
Цель и задачи работы

Цель работы – разработка метода создания связных сцен из художественного видео.

Задачи работы:

- анализ предметной области, выделение основных подходов к созданию связных сцен;
- создание алгоритма для решения поставленной задачи;
- описание специфических особенностей предложенного метода;
- исследование наиболее важных аспектов разработанного программного продукта.

Предметная область

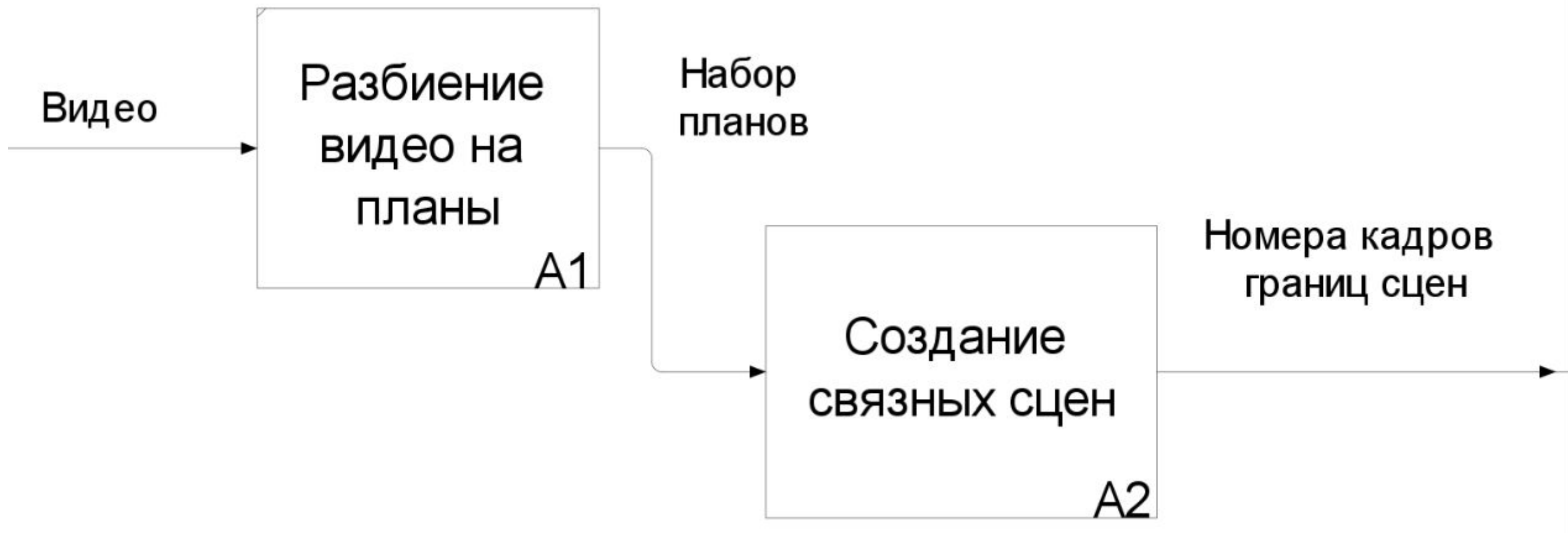


План – это набор кадров, непрерывно снятых с одной камеры без монтажных склеек

Сцена - это набор последовательно идущих планов, которые характеризуются похожим аудиовизуальным составяющим

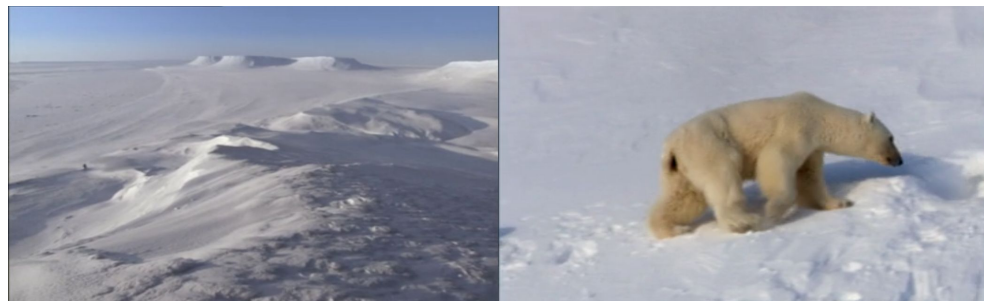
Иерархическая структура видео

Постановка задачи



Разбиение видео на планы

- Определять границы планов можно находя монтажные склейки, они бывают двух видов:
- Склейка встык
- Плавный переход



С помощью современных нейронных сетей возможно определять монтажные склейки достаточно достоверно для дальнейшего анализа. Будем использовать для этого сеть TransNet.



Первый план первая сцена



Первый план первая сцена

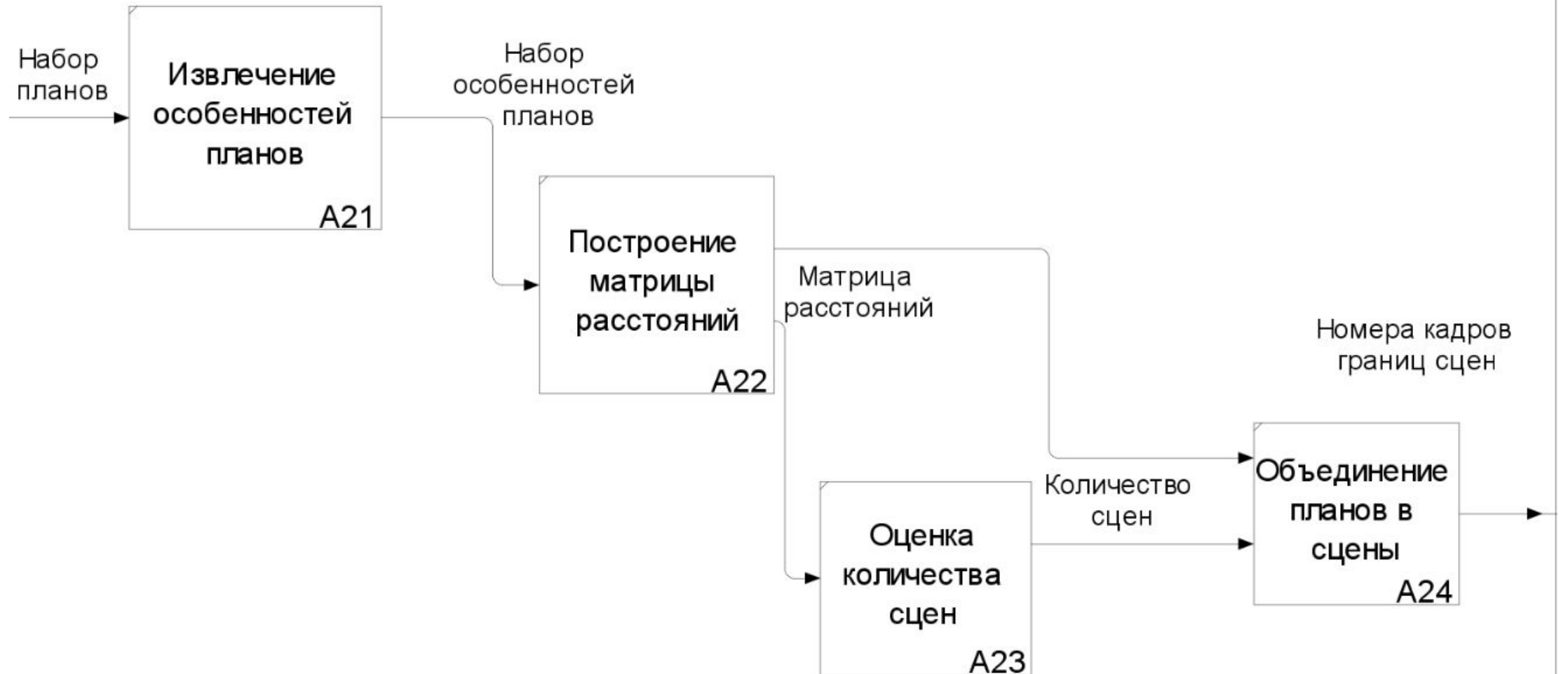


Второй план первая сцена

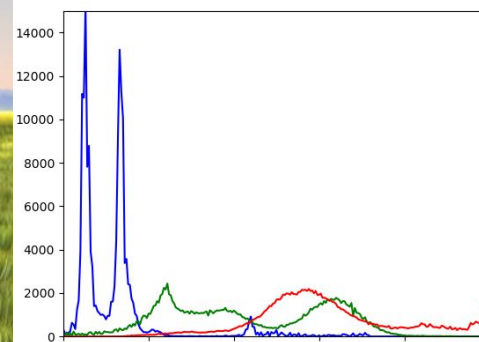
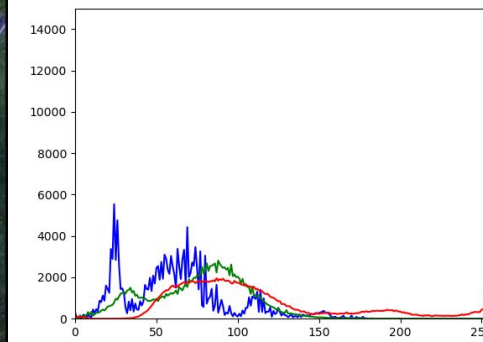
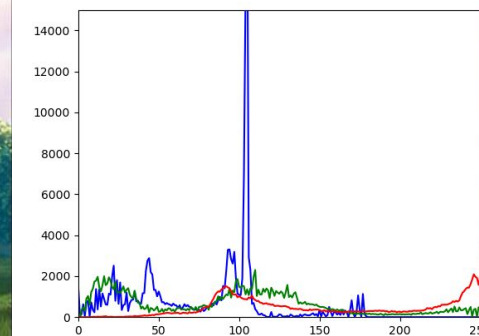
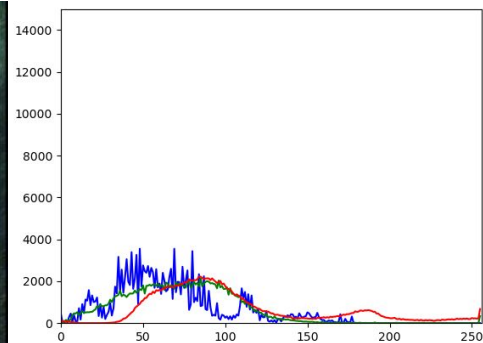


Третий план вторая сцена

Метод динамической оптимизации

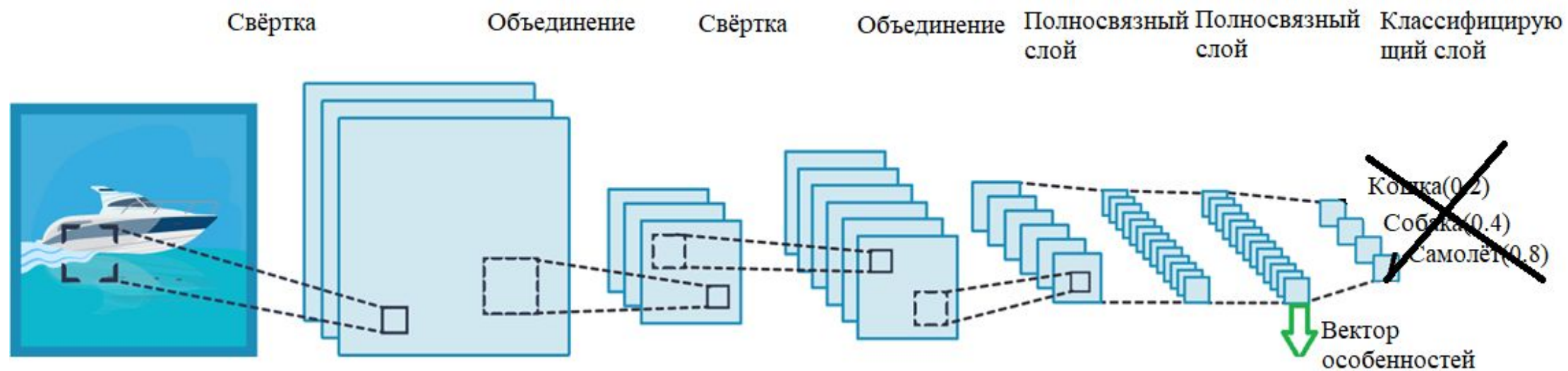


Особенности планов



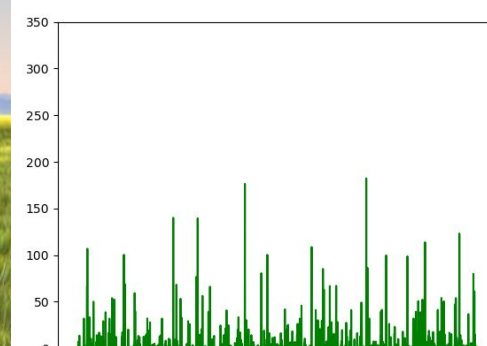
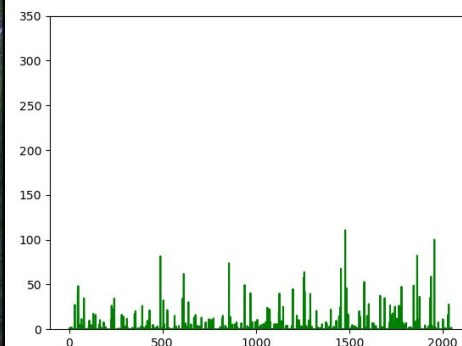
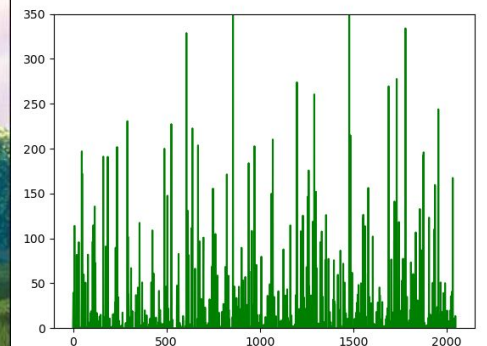
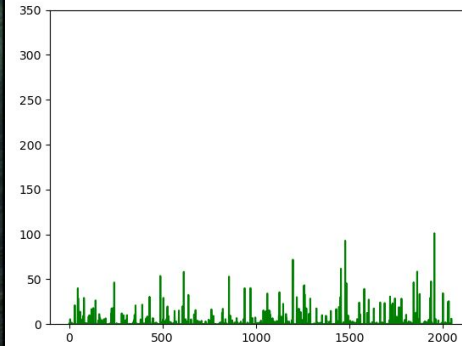
HSV гистограммы кадров как особенности: синий – цветовой тон, зелёный – насыщенность, красный - яркость

Извлечение вектора особенностей из классифицирующей нейросети



Для извлечения визуальных особенностей будем использовать свёрточные нейронные сети обученные на наборе данных ImageNet

Неявные собенности планов



Визуализация вектора особенностей извлеченного нейросетью Xception обученной на наборе изображений ImageNet

Матрица расстояний планов

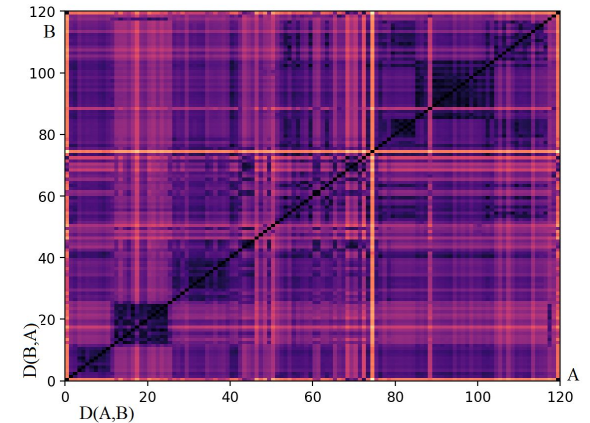
Сравнивая каждый план с каждым, используя функцию расстояния векторов $D(A,B)$, строим матрицу расстояний D . Матрица симметрична относительно побочной диагонали т.к. $D(A,B)=D(B,A)$.

В качестве функции расстояния будем использовать косинусное расстояние векторов

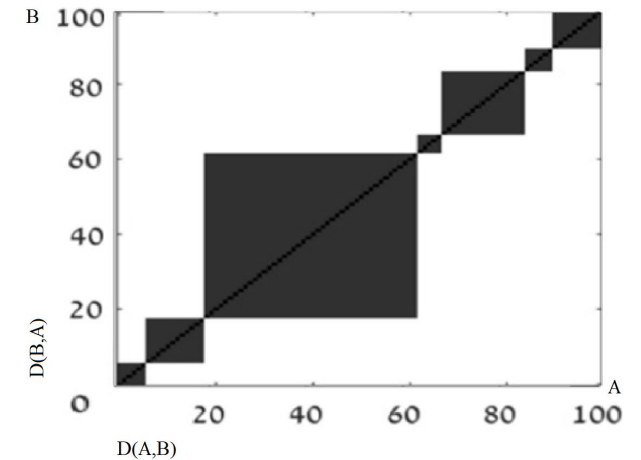
$$D(A, B) = \cos(A, B) = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}$$

,где A и B вектора длины n

Пример матрицы расстояний планов



Пример идеальной матрицы расстояний планов



Оценка количества сцен

- Идеальная матрица расстояний имеет линейно зависимые строки и столбцы, так что её ранг равен числу сцен
- Для неидеальной матрицы необходимо отфильтровать шумы что возможно сделать с помощью сингулярного разложения матрицы
- Среди сингулярных значений, отсортированных по убыванию, находим локтевую точку — ту, с которой уменьшение значений резко замедляется. Индекс локтевой точки — это примерное количество сцен.

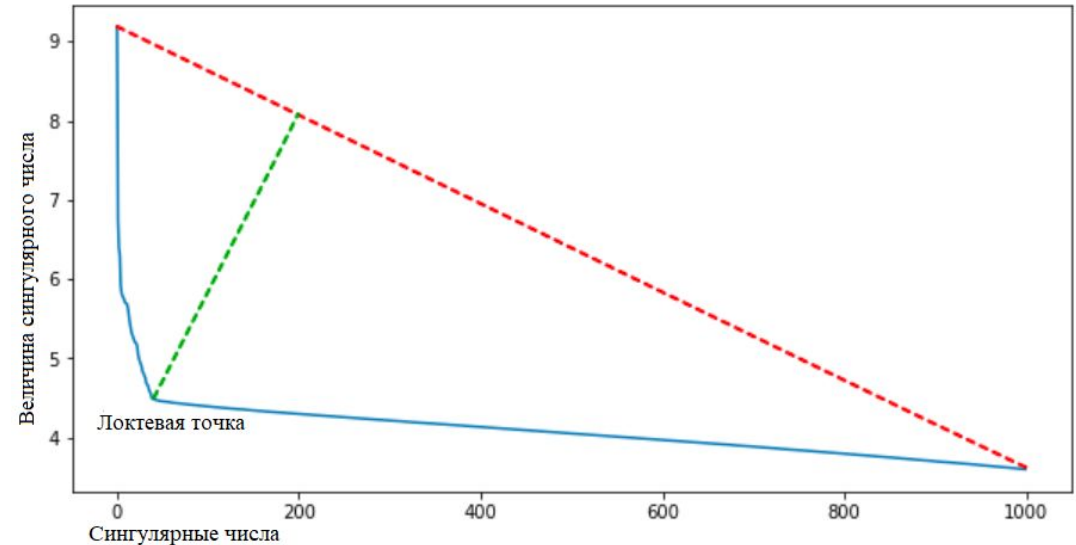
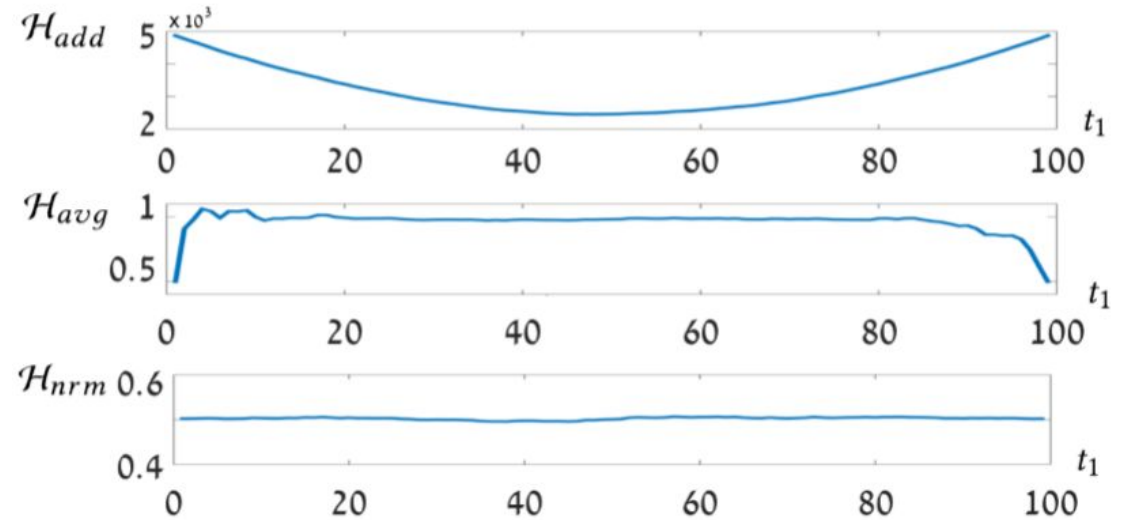


График отображающий сортированные по убыванию сингулярные числа матрицы расстояний из 1000 планов.

Минимизации функции стоимости

Можно определить как задачу построения связанных сцен как задачу минимизации функции $H_D^{K,N}$, где K – количество сцен, N – количество планов, D – матрица расстояний планов.

Функция			
Сложность	$O(NK)$	$O(NK)$	



Значения функций для матрицы равноудалённых планов при $K=2$, $N=100$. Ось X – индекс плана, Ось Y значение функции

Минимизация аддитивной функции

$$\dot{H}_{add_D}^{K,N}(t) = \sum_{i=1}^K \left\{ \sum_{j=t_{i-1}+1, j'=t_{i-1}+1}^{t_i} D(x_j, x_{j'}) \right\}$$

,где t – вектор границ сцен

Найти оптимальное t возможно с помощью построения двух матриц $C(n,k)$ и $J(n,k)$, но мы также будем использовать матрицу $E(n,n)$ для хранения промежуточных значений. n принимает значения от 1 до N , k от 2 до K .

Инициализируем матрицы значениями

$$C(n, 1) = \sum_{j=n}^N \sum_{j'=n}^N D(x_j, x_{j'})$$

$$J(n, 1) = N$$

$$E(i, i') = \sum_{j=i}^{i'} \sum_{i'=i}^{i'} D(x_j, x_j)$$

Метод динамической оптимизации аддитивной функции

E может быть рассчитана рекурсивно при запросе как

$$E(i, i') = E(i - 1, i') + E(i, i' - 1) - E(i - 1, i' - 1) + D(x_i, x_{i'}) + D(x_{i'}, x_i)$$

А матрицы C и J вычисляются изменениями значения n от 1 до N для каждого k от 2 до K

$$C(n, k) = \min_{i=n}^N \{E(i, i) + C(i + 1, k - 1)\}$$

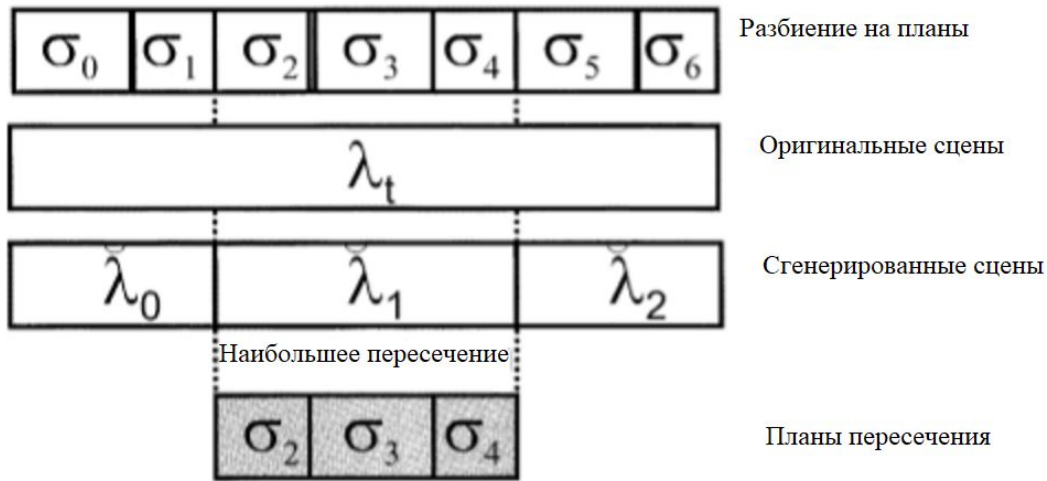
$$J(n, k) = \operatorname{argmin}_{i=n}^N \{E(i, i) + C(i + 1, k - 1)\}$$

После чего границы сцен можно вычислить с помощью формулы

$$t_i^* = J(t_{i-1} + 1, K - i + 1)$$

для i от 1 до K

Метрика Coverage



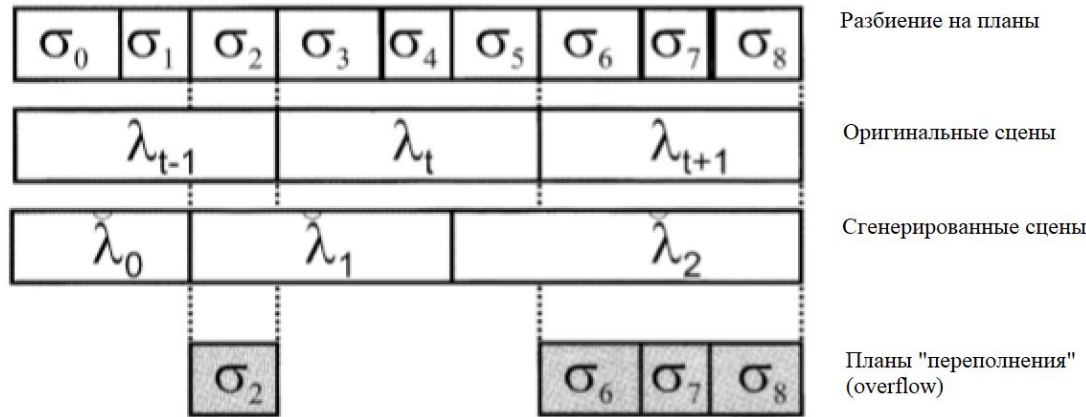
$$C(L_{new}, L_{orig}) = \frac{\sum^n \max_{intersec}(L_{new}, L_{orig_i})}{\sum^n len(L_{orig_i})}$$

Где L_{new} созданные связные сцены, L_{orig} оригинальные сцены, $len(L)$ – длина сцены в планах, а $\max_{intersec}(L_{new}, L_{orig_i})$ длина в планах наибольшего пересечения сцены L_{orig_i} со сценой из набора L_{new}



Зелёная полоска означает границы оригинальной сцены, красная - созданной, белая - пересечение

Метрика Overflow



$$O(L_{new}, L_{orig}) = 1 - \frac{\sum^n overflow(L_{new}, L_{orig_i})}{\sum^n len(L_{orig_i})}$$

Где L_{new} созданные связные сцены, L_{orig} оригинальные сцены, $len(L)$ – длина сцены в планах, а $overflow(L_{new}, L_{orig_i})$ количество лишних планов в сценах соответствующих L_{orig_i} в наборе L_{new}



Зелёная полоска означает границы оригинальной сцены, красная - созданной, белая – пересечение, фиолетовая – overflow созданной

Зависимость результата работы метода от реализации модуля извлечения особенностей

Способ	Xception	ResNet	Inception	VGG	MobileNet	HSV гистограммы	DenseNet
F-мера Coverage Overflow	0,753	0,732	0,706	0,696	0,695	0,615	0,612

В таблице представлены усреднённые значения F-меры метрик Coverage и Overflow для результатов работы метода на наборе данных OVSD с использованием разных способов извлечения особенностей

Результаты создания СВЯЗНЫХ СЦЕН

Применим разработанное ПО к видео Big Buck Bunny и получим набор кадров, являющихся границами сцен.

Для демонстрации результатов работы выберем средний кадр из каждой сцены.



Заключение

В результате выполнения работы были решены поставленные задачи:

- проведён анализ предметной области, выделены основные подходы к созданию связных сцен;
- разработан алгоритм решающий поставленную задачу;
- описаны специфические особенности предложенного метода;
- разработано программное обеспечение для создания связных сцен;
- исследованы наиболее важные аспекты разработанного программного продукта.

Направления развития

- повышение точности предсказания количества сцен;
- извлечение высокоуровневых особенностей планов;
- повышение точности сегментации;
- замена аддитивной функции стоимости на функцию менее зависящую от номера плана ;
- использование сиамских нейронных сетей для определения расстояния между планами