

# Некоторые методы сегментации видео



---

Сергей Матюнин

*Video Group*  
*CS MSU Graphics & Media Lab*



# Содержание

---

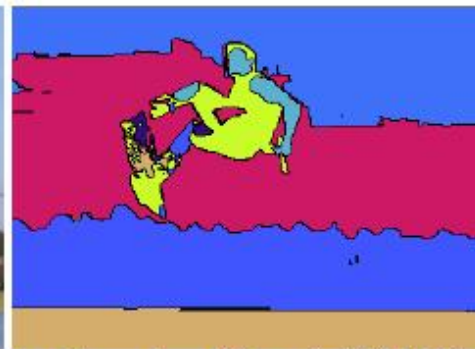
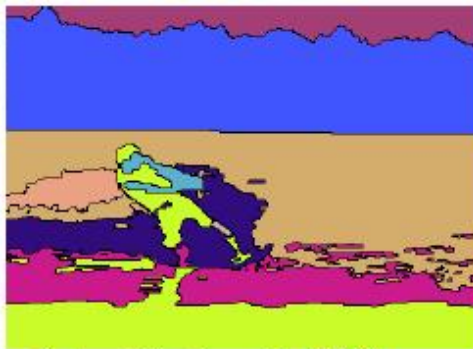
- **Введение**
- Общие подходы
- Сегментация на основе траекторий
- Online Dense Segmentation
- Заключение

# Сегментация изображений

- Сегментация – задача разбиения изображения на несколько множеств пикселей (суперпикселей) так, чтобы разбиение удовлетворяло заданным условиям
- Кластеризация – задача разбиения заданной выборки объектов на подмножества (кластеры) так, чтобы
  - каждое подмножество состояло из схожих объектов
  - объекты разных кластеров существенно отличались

# Сегментация изображений

## Примеры (1/2)



Исходное  
изображение

Результат  
сегментации

Исходное  
изображение

Результат  
сегментации

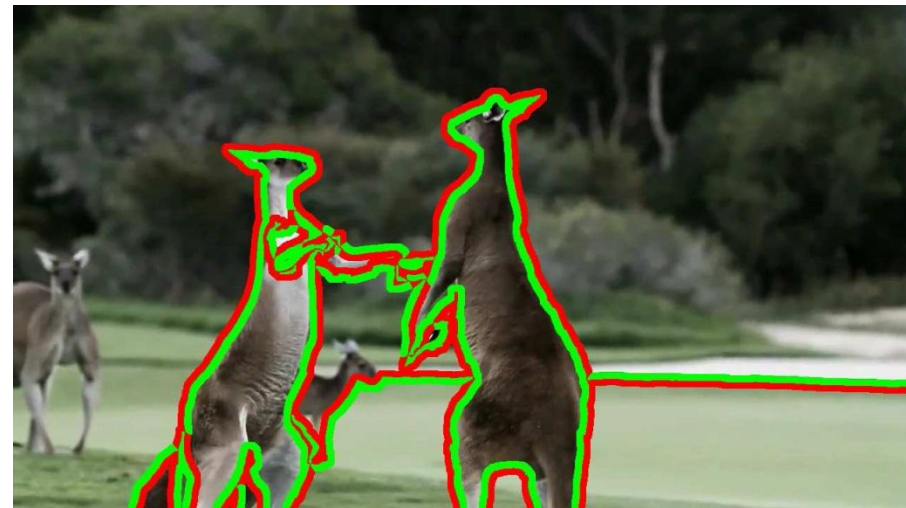
Разными цветами отмечены различные кластеры

# Сегментация изображений

## Примеры (2/2)



Исходный кадр



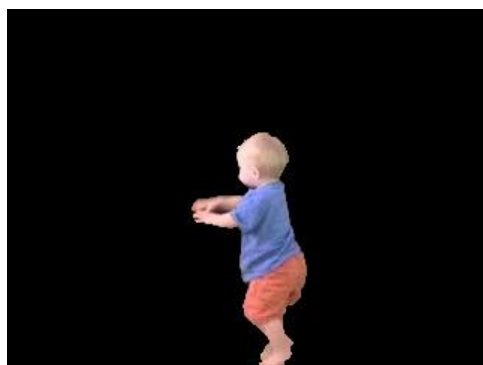
Результат сегментации

# Сегментация видео

## Пример

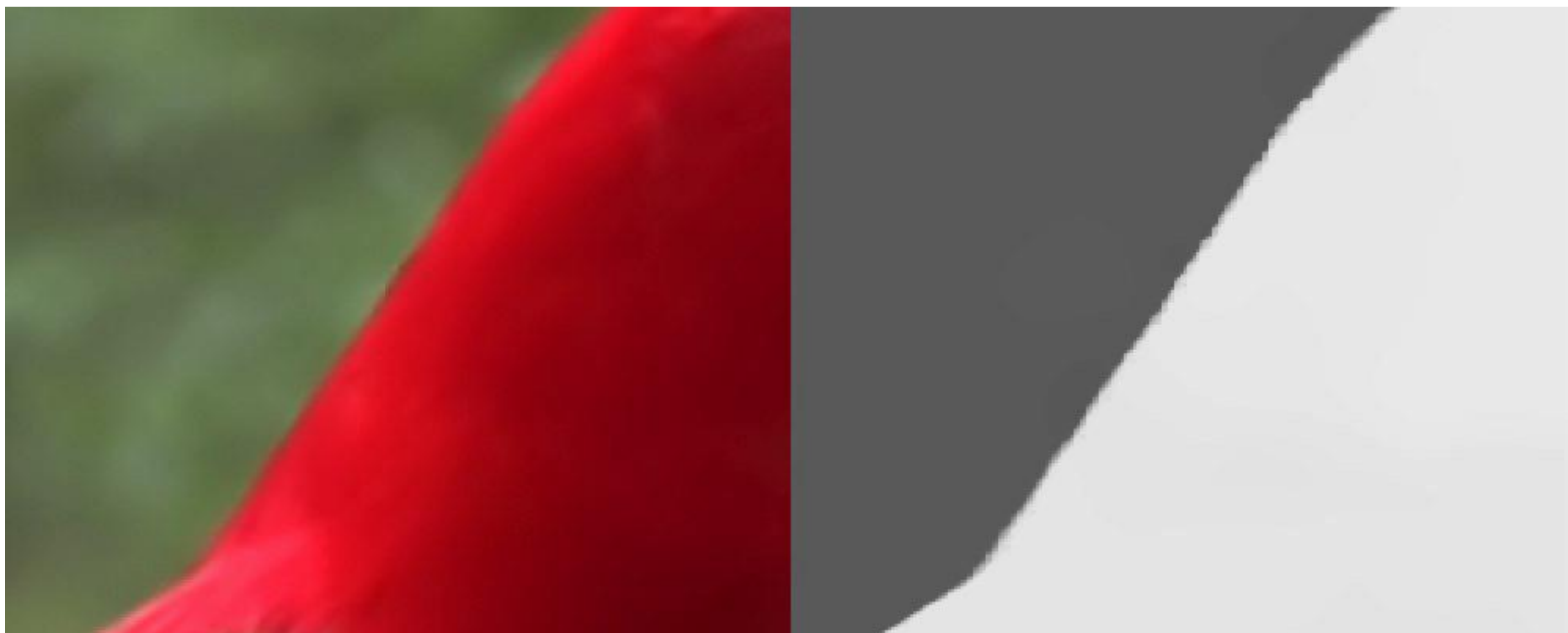


Кадры последовательности



Результаты сегментации

# Применение Depth Propagation



Исходное видео

Распространенная  
глубина

# Применение Depth from Motion







# Содержание

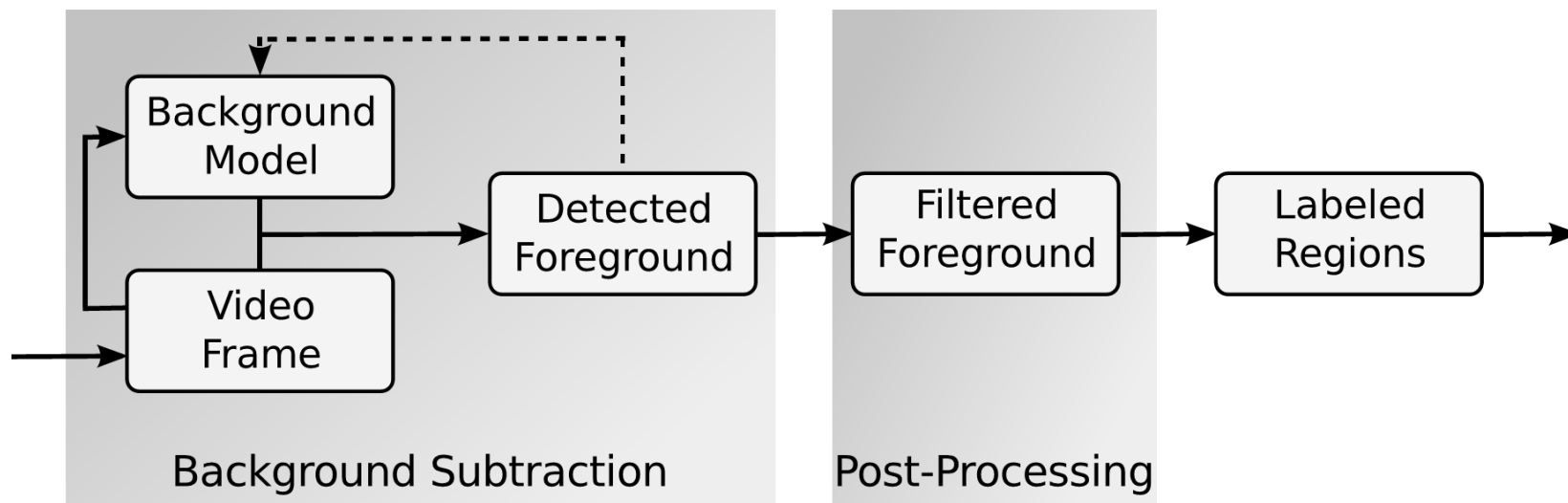
---

- Введение
- **Общие подходы**
- Сегментация на основе траекторий
- Online Dense Segmentation
- Заключение

# Сегментация объектов

## Применение в видеонаблюдении (1/2)

### Типичный процесс выделения объектов



# Сегментация объектов

## Применение в видеонаблюдении (2/2)

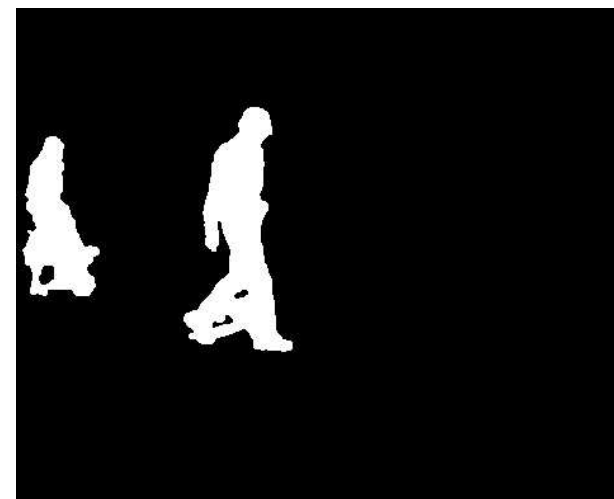
- Камера неподвижна
- Объекты малы
- Точность часто не критична



Кадр без объектов



Кадр с объектами



Маска объектов

# Общие подходы

## Вычитание фона

### Метод для неподвижной камеры

1. Находим фон усреднением или медианой большого количества кадров
2. Вычитаем из исходного кадра
3. То, что осталось – объекты

# Общие подходы

## Глобальное движение

- Оцениваем глобальное движение
- Считаем объектами области, где движение не совпадает с глобальным
- Остальные области – фон

# Общие подходы

## RANdom SAmple Consensus (RANSAC)

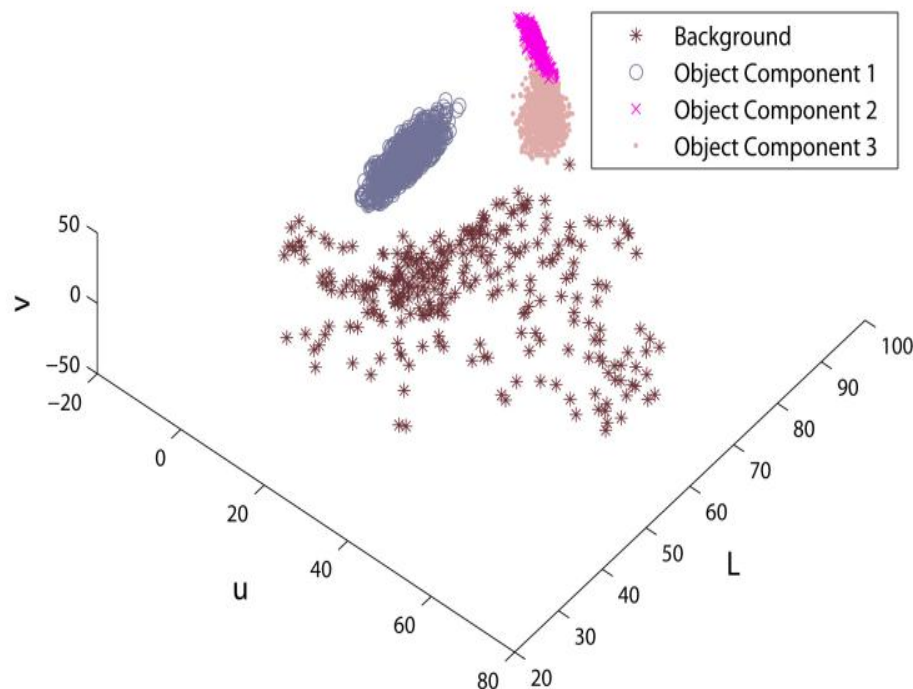
Сегментируем пространство координат точек на нескольких последовательных кадрах

1. Применяем RANSAC к исходному множеству. Находим подпространство из inliers. Остальные точки – outliers
2. Применяем шаг 1 к множеству outliers, пока не найдем все подпространства
3. Для каждого подпространства inliers применяем метод главных компонент, чтобы найти наилучший базис для каждого подпространства
4. Выбираем для каждой точки ближайшее подпространство

# Общие подходы

## Смесь гауссиан

- Пытаемся представить множество точек видео как смесь гауссиан в некотором пространстве
- Выделяем компоненты, которые соответствуют объекту или фону



# Общие подходы

## Спектральная кластеризация

- Составляем матрицу схожести точек  $\mathbf{W}$
- Ищем декомпозицию нормализованной матрицы Кирхгофа (Laplacian matrix)

$$\Gamma^T \Lambda \Gamma = \mathbf{D}^{-\frac{1}{2}} (\mathbf{D} - \mathbf{W}) \mathbf{D}^{-\frac{1}{2}}$$

$$D_{ii} = \sum_j W_{ij}$$

- $m + 1$  собственный вектор  $v_0, \dots, v_m$ , соответствующий наименьшим собственным значениям, определяет пространство меньшей размерности
- Далее можно использовать, например, метод k-средних

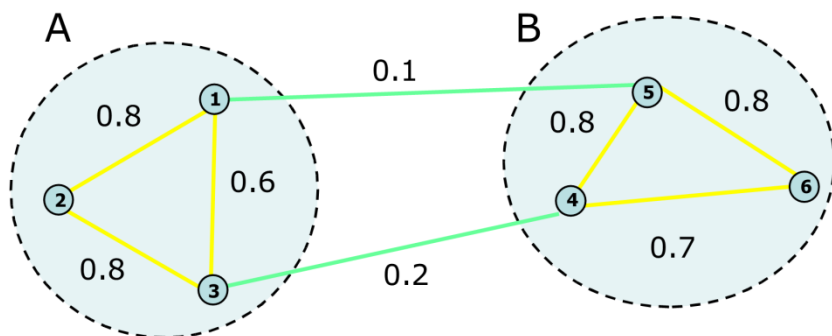
Матрица схожести точек

	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$
$x_1$	0	0.8	0.6	0	0.1	0
$x_2$	0.8	0	0.8	0	0	0
$x_3$	0.6	0.8	0	0.2	0	0
$x_4$	0.8	0	0.2	0	0.8	0.7
$x_5$	0.1	0	0	0.8	0	0.8
$x_6$	0	0	0	0.7	0.8	0



# Спектральная кластеризация

## Пример



Матрица Кирхгофа

$x_1$	1.5	-0.8	-0.6	0	-0.1	0
$x_2$	-0.8	1.6	-0.8	0	0	0
$x_3$	-0.6	-0.8	1.6	-0.2	0	0
$x_4$	-0.8	0	-0.2	2.5	-0.8	-0.7
$x_5$	-0.1	0	0	0.8	1.7	-0.8
$x_6$	0	0	0	-0.7	-0.8	1.5

$A =$

0.0
0.3
2.2
2.3
2.5
3.0

$X =$

0.4	0.2	0.1	0.4	-0.2	-0.9
0.4	0.2	0.1	-0.	0.4	0.3
0.4	0.2	-0.2	0.0	-0.2	0.6
0.4	-0.4	0.9	0.2	-0.4	-0.6
0.4	-0.7	-0.4	-0.8	-0.6	-0.2
0.4	-0.7	-0.2	0.5	0.8	0.9

Матрица расстояний в графе

	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$
$x_1$	0	0.8	0.6	0	0.1	0
$x_2$	0.8	0	0.8	0	0	0
$x_3$	0.6	0.8	0	0.2	0	0
$x_4$	0.8	0	0.2	0	0.8	0.7
$x_5$	0.1	0	0	0.8	0	0.8
$x_6$	0	0	0	0.7	0.8	0



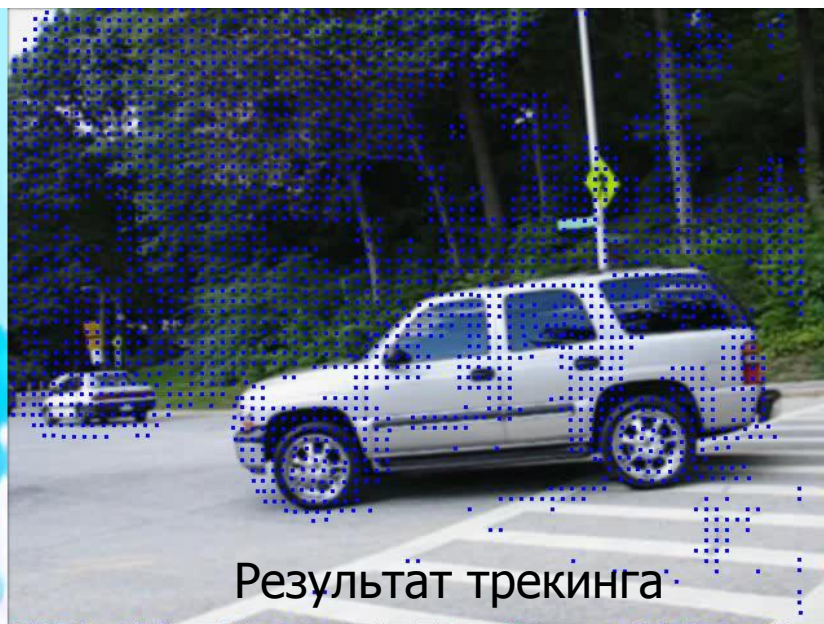
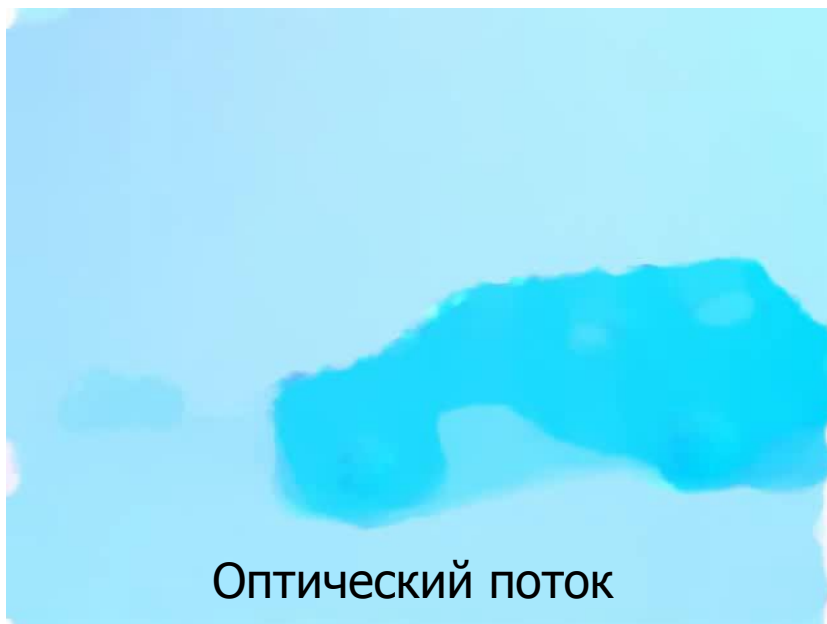
# Содержание

---

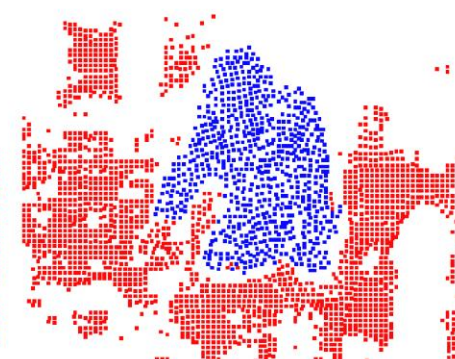
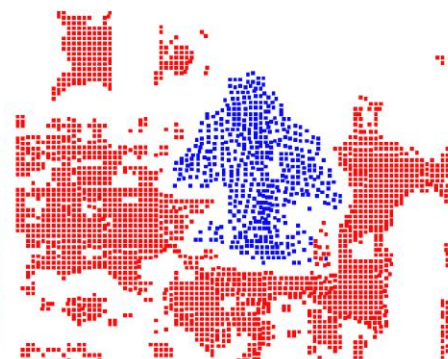
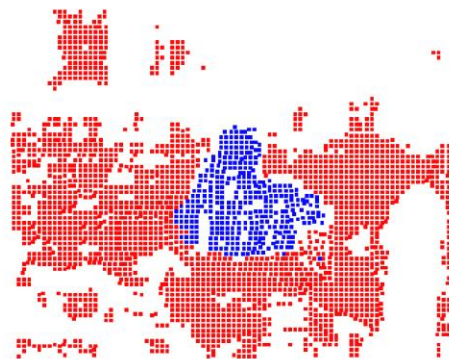
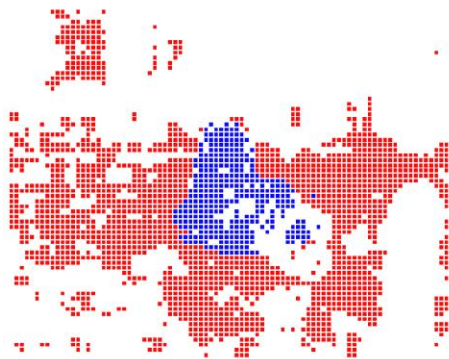
- Введение
- Общие подходы
- **Сегментация на основе траекторий**
- Online Dense Segmentation
- Заключение

# Траектории точек

- Вместо покадрового OF используем длинные траектории
- Траектории строим на основе трекера [Sundaram et al., ECCV 2010]. Трекер основан на OF



# Преимущества использования длинных траекторий



Кадр 0

Кадр 30

Кадр 50

Кадр 80

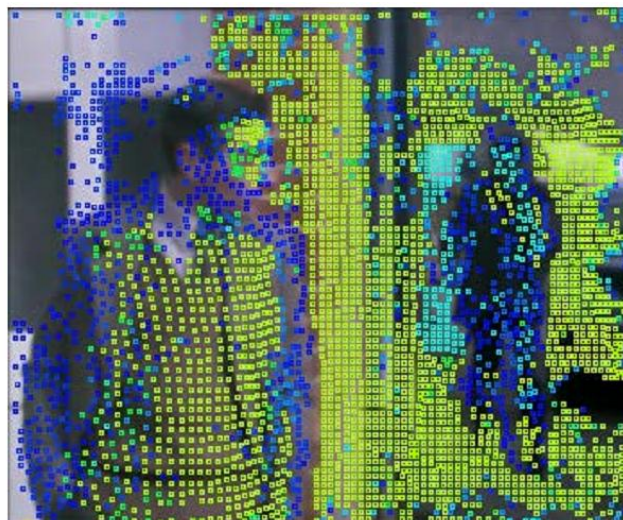
Кадры из фильма «Miss Marple: Murder at the vicarage»

# Пример данных трекера



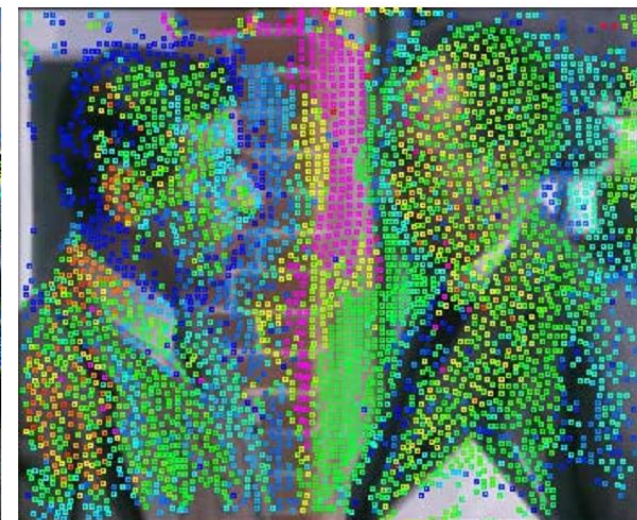
Кадр 0

Выбранные точки



Кадр 211

Результат трекинга



Кадр 400

Результат трекинга

Точки распределяются равномерно в контрастных областях  
При необходимости добавляются новые в области открытия  
Цвета от новых к старым: синий, зеленый, желтый, красный, розовый

Кадры из фильма «Miss Marple: Murder at the vicarage»

# Траектории точек

## Мера расстояния между траекториями

- Определяем меру расстояния между траекториями:

$$d^2(A, B) = \max_t d_t^2(A, B)$$

- Максимум считаем по тем кадрам, где присутствуют обе траектории
- Можем разделять сегменты даже на тех кадрах, где они движутся одинаково

# Траектории точек

## Мера расстояния между точками

Расстояние между точками в момент времени  $t$ :

$$d_t^2(A, B) = d_{\text{sp}}(A, B) \frac{(u_t^A - u_t^B)^2 + (v_t^A - v_t^B)^2}{5\sigma_t^2}$$

$d_{\text{sp}}(A, B)$  – среднее евклидово расстояние  
в окрестности  $t$

$v_t := y_{t+5} - y_t$  и  $u_t := x_{t+5} - x_t$  – компоненты  
вектора движения через 5 кадров

$$\sigma_t = \min_{a \in \{A, B\}} \sum_{t'=1}^5 \sigma(x_{t+t'}^a, y_{t+t'}^a, t + t')$$

$\sigma : \mathbb{R}^3 \rightarrow \mathbb{R}$  – локальная дисперсия векторов

# Траектории точек

## Схожесть траекторий

- От расстояния переходим к мере схожести:  
$$w(A, B) = \exp(-\lambda d^2(A, B))$$
$$\lambda = 0.1$$
- Составляем квадратную матрицу схожести  $W$   
Близость траекторий объекта, которые не пересекаются по времени, учитывается за счет транзитивности расстояния
- Можем кластеризовать траектории, используя разложение по собственным векторам



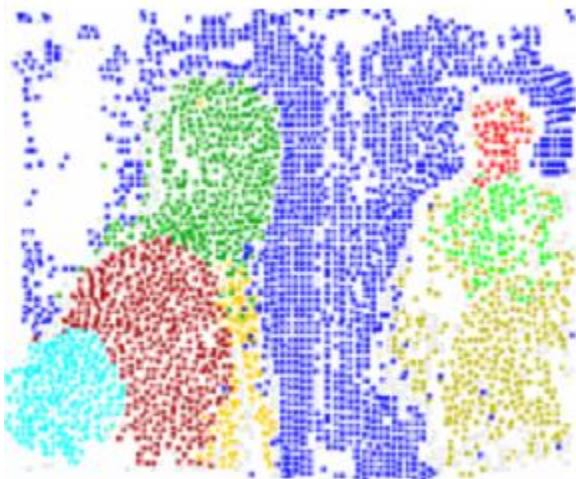
# Траектории точек

## Кластеризация

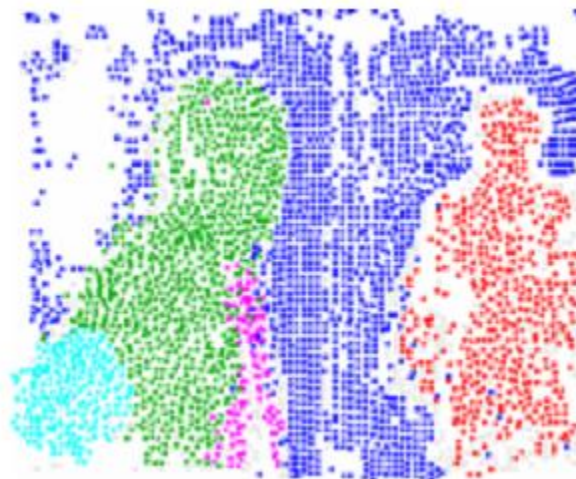
- Число кластеров неизвестно заранее
- Спектральное разложение не обеспечивает четкого разделения
- Предлагается эвристика:
  - Выбираем  $K_0$ , чтобы  $K_0$ -е собственное значение было  $< 0.2$
  - Для всех  $K < K_0$  запускаем  $K$ -means 10 раз
  - Запускаем иерархичный 2-means и берем 20 кластеров с минимальной энергией
  - Из набора кандидатов объединяем те, которые лучше всего «сглаживают» разметку и оставляют небольшой энергию
  - Объединяем кластеры с похожим движением

# Траектории точек

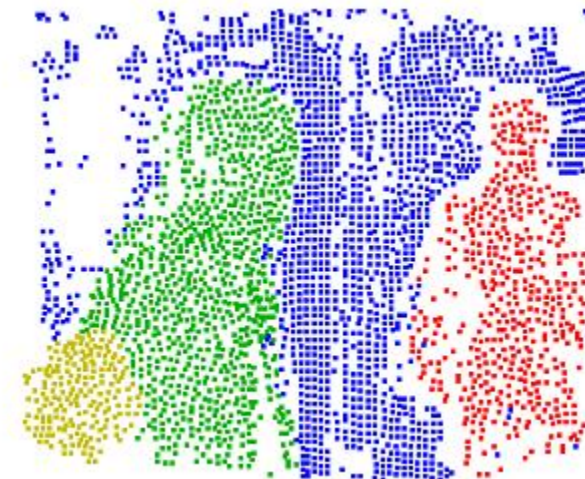
## Пример кластеризации



Чрезмерная  
сегментация



После слияния



После учета  
похожести движения

# Траектории точек

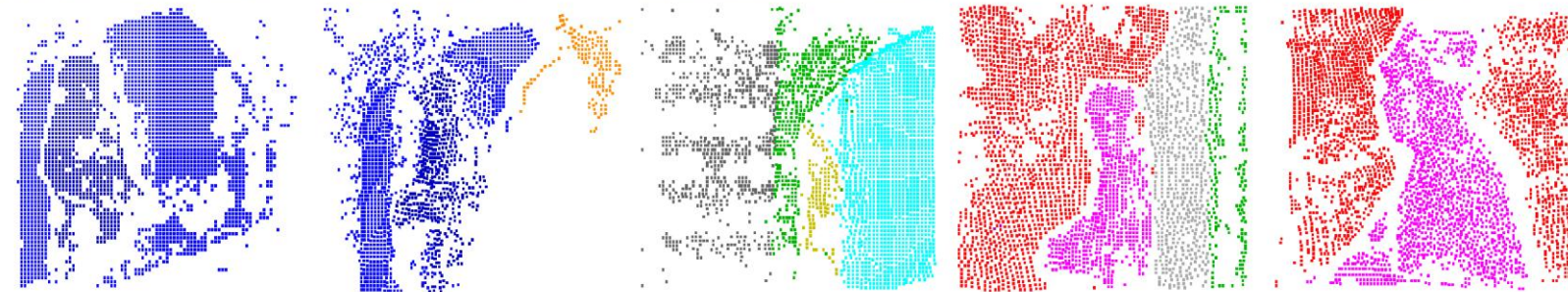
## Пример работы



Видео



Сегментация



Эталон



Кадр 1

Кадр 110

Кадр 135

Кадр 170

Кадр 250

Кадры из фильма «Miss Marple: Murder at the vicarage»

T. Brox and J. Malik, "Object Segmentation by Long Term Analysis of Point Trajectories," ECCV, 2010

# Траектории точек

## Сравнение с другими методами (1/2)

### Скорость работы

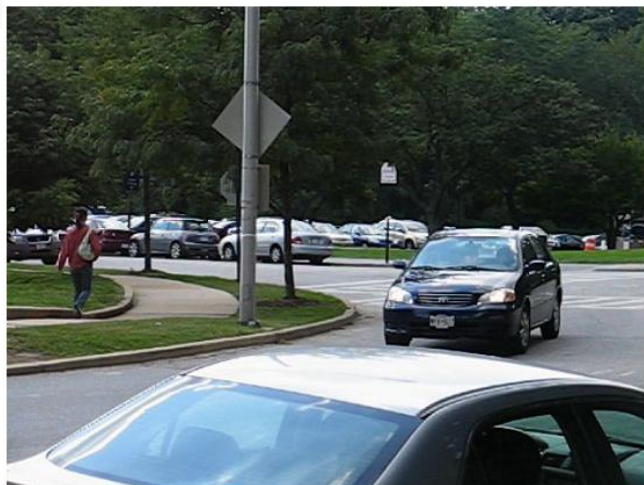
	tracks	time
our method	15486	497s
GPCA	12060	2963s
LSA	12060	38614s
RANSAC	12060	15s
ALC	957	22837s

### Качество сегментации

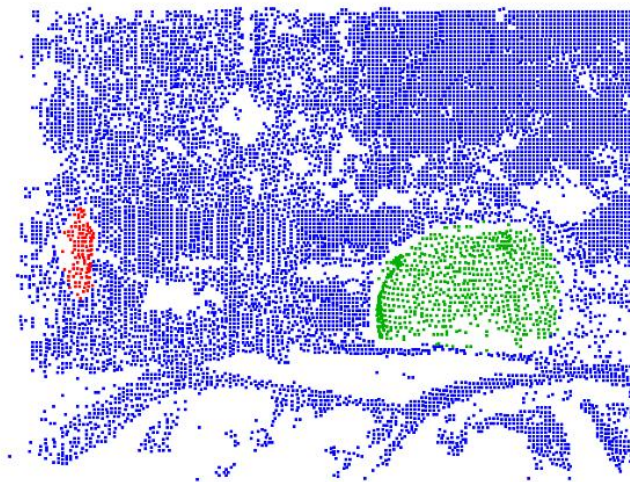
	Density	overall error	average error	over-segmentation	extracted objects
First 10 frames (26 sequences)					
our method	3.34%	7.75%	25.01%	0.54	24
GPCA	2.98%	14.28%	29.44%	0.65	12
LSA	2.98%	19.69%	39.76%	0.92	6
RANSAC	2.98%	13.39%	26.11%	0.50	15
ALC corrupted	2.98%	7.88%	24.05%	0.15	26
ALC incomplete	3.34%	11.20%	26.73%	0.54	19
First 50 frames (15 sequences)					
our method	3.27%	7.13%	34.76%	0.53	9
ALC corrupted	1.53%	7.91%	42.13%	0.36	8
ALC incomplete	3.27%	16.42%	49.05%	6.07	2
First 200 frames (7 sequences)					
our method	3.43%	7.64%	31.14%	3.14	7
ALC corrupted	0.20%	0.00%	74.52%	0.40	1
ALC incomplete	3.43%	19.33%	50.98%	54.57	0
All available frames (26 sequences)					
our method	3.31%	6.82%	27.34%	1.77	27
ALC corrupted	0.99%	5.32%	52.76%	0.10	15
ALC incomplete*	3.29%	14.93%	43.14%	18.80	5

# Траектории точек

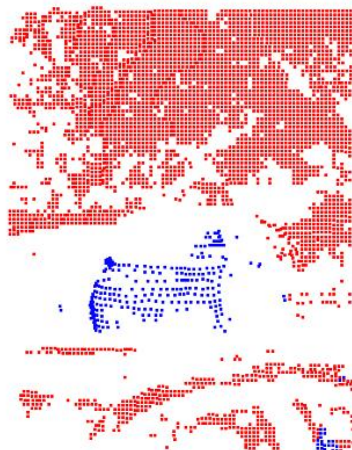
## Сравнение с другими методами (2/2)



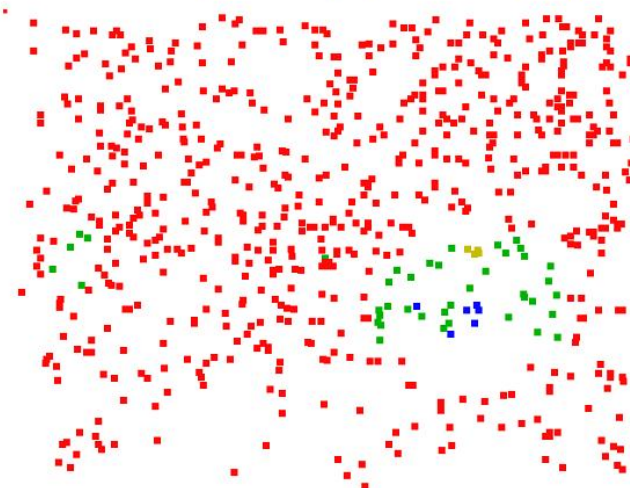
Исходный  
кадр



Предложенный  
метод



RANSAC



Rao et al., 2009

# Траектории точек

## Плотная разметка

- Разреженные траектории берутся из-за ограничений по производительности
- Предлагается распространить метки на все точки изображения после сегментации  
От тех же авторов есть статья, но метод медленный



# Содержание

---

- Введение
- Общие подходы
- Сегментация на основе траекторий
- **Online Dense Segmentation**
- Заключение

# Online Dense Segmentation

## Характеристика



Цель: разделить объект и фон

- Работает для случая нестационарной камеры
- Может обрабатывать длинные видео «online»
- Использует долгосрочные траектории (long term trajectories)
- Подходит для движения нежестких объектов
- Обрабатывает частичные перекрытия



# Online Dense Segmentation

## Краткое описание алгоритма



1. Инициализация для первого кадра  
Для каждого последующего кадра:
2. Трекинг точек с предыдущего кадра
3. Продолжение траектории с предыдущего кадра
4. Инкрементная кластеризация
5. Обновление меток (BG/FG) для кластеров
6. Оценка движения для всех пикселей объектов и фона
7. Построение восстановленного фона

# Online Dense Segmentation

## Кластеризация траекторий



- Мера расстояния между траекториями  $T_A$  и  $T_B$

$$d^M(T_A, T_B) = \max_{t \in (A \cap B)} d_{t-\Delta:t}^M(T_A, T_B)$$

$$d^S(T_A, T_B) = \max_{t \in (A \cap B)} d_t^S(T_A, T_B)$$

$$d_{t-\Delta:t}^M(T_A, T_B) = (u_A^{t-\Delta:t} - u_B^{t-\Delta:t})^2 + (v_A^{t-\Delta:t} - v_B^{t-\Delta:t})^2$$

$$d_t^S(T_A, T_B) = \| q_A^t - q_B^t \| \quad q - \text{координата точки}$$

- Матрицы расстояний

$$\mathbf{D}^M = [D_{ij} = d_{1:t}^M(T_i, T_j)] \quad \mathbf{D}^S = [D_{ij} = d_{1:t}^S(T_i, T_j)]$$

- Матрица схожести (affinity matrix)

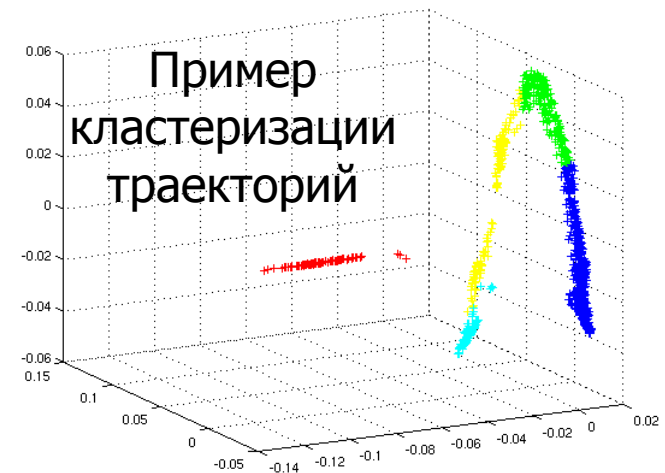
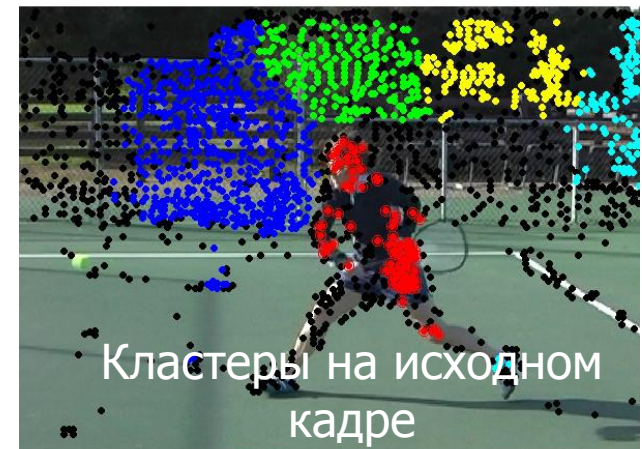
$$\mathbf{W}_{1:t} = \exp(-\mathbf{D}_{1:t}^M / \lambda_M^2) \cdot \exp(-\mathbf{D}_{1:t}^S / \lambda_S^2)$$

# Online Dense Segmentation

## Понижение размерности



- При помощи статистических методов понижаем размерность данных
- В подпространстве меньшей размерности траектории моделируем смесью гауссиан (GMM)
- Получаем кластеризованные данные



# Online Dense Segmentation

## Разметка объект/фон



- Помечаем кластеры как объект или фон в зависимости от значения нескольких признаков
- Вводим функционал энергии

$$E_l(L) = \alpha_C \sum_i \phi_C(l_i) + \alpha_A \sum_{(i,j)} \phi_A(l_i, l_j) + \alpha_B \sum_{(i,j)} \phi_B(l_i, l_j) + \alpha_S \phi_S(L)$$

Компактность (Compactness)      Схожесть по движению (Affinity)      Границы в подпространстве (Borders)      Вложенность (Surroundedness)

- Ищем разметку

$$L = \{l_1, \dots, l_R\}$$

$$l_i \in \{0, 1\} \quad (0 \equiv \text{foreground}, 1 \equiv \text{background})$$

# Online Dense Segmentation

## Оценка компактности



$$\phi_C(l_i) = (1 - l_i) \cdot \left( \max\left(\frac{\text{var}(x)}{\text{var}(y)}, \frac{\text{var}(y)}{\text{var}(x)}\right) - 1.5 \right)$$

$\text{var}(x)$  – дисперсия траекторий  
по направлению  $x$

Предполагается, что кластеры,  
принадлежащие объекту, более вытянутые  
Поощряем вытянутые кластеры для объектов

# Online Dense Segmentation

## Оценка схожести движения



- Поощряем одинаковое движение для фона
- Штрафуем одинаковое движение у объекта и у фона

$$\phi_A(l_i, l_j) = -l_i l_j \xi_{\text{Affine}} + (l_i(1-l_j) + l_j(1-l_i)) \xi_{\text{Affine}}$$

$\xi_{\text{Affine}}$  – характеристика меры близости кластеров  
в пространстве уменьшенной размерности

# Online Dense Segmentation

## Оценка границ



Штрафуем:

- Отдаленные кластеры, помеченные как фон
- Близкие кластеры, помеченные как объект и фон

$$\phi_B(l_i, l_j) = -l_i l_j \xi_{\text{Embed}} + (l_i(1 - l_j) + l_j(1 - l_i)) \xi_{\text{Embed}}$$

$$\xi_{\text{Embed}} = \exp(-\min_{\forall x_i, x_j} \|x_i - x_j\|)$$

↑  
Расстояние между кластерами

# Online Dense Segmentation

## Оценка вложенности



$$\phi_S(L) = 1 - \frac{|F \in \text{ConvexHull}(B)|}{|F|}$$

$F$  – множество точек объекта

$B$  – множество точек фона

ConvexHull – выпуклая оболочка



# Online Dense Segmentation

## Распространение разметки

- Распространяем метки на все точки кадра
- Используем байесовский подход и марковские цепи

$$p(L_t | I_t) \propto \prod_{(i,j) \in \mathcal{E}} \Phi(l_t^i, l_t^j) \prod_i \Psi(l_t^i)$$

Метки FG/BG      Яркость кадра      Произведение по всем парам соседних точек      Произведение по всем точкам

Гауссиан      Поощряем одинаковые метки

$$\Phi(l_t^i, l_t^j) = \mathcal{N}(I_t^i - I_t^j | 0, \Sigma_l) (l^i l^j + (1 - l^i)(1 - l^j))$$

$$\Psi(l_t^i) = l^i \cdot p(l_t^i | I_t^i, P_t) + (1 - l^i)(1 - p(l_t^i | I_t^i, P_t))$$

Описание Feature points

# Online Dense Segmentation

## Оценка вероятности метки

$$p(l_t^i = k | I_t^i, P_t) \propto \int_{a_k^i} p(I_t^i | l_t^i, a_k^i) p(a_k^i | P_t) \mathbf{d}a_k^i p(l_t^i | P_t)$$

↑  
Описание особых точек

← Цвет (яркость) пикселя

$$p(a_k^i | P_t) \approx p(a_{k,t-1}^{j(i, \mu_{k,t}^i)})$$

Скомпенсированный пиксель

← Средний вектор движения

Известно с предыдущего кадра

$$p(l_t^i = f | P_t) = \sum_{m_{f,t}^i \in \mathbb{N}^2} p(m_{f,t}^i | P_t) p(l_{t-1}^{j(i, m_{f,t}^i)} = f)$$

↑  
Оценивается по движению особых точек

# Online Dense Segmentation

Пример работы



Sheikh et al., 2009

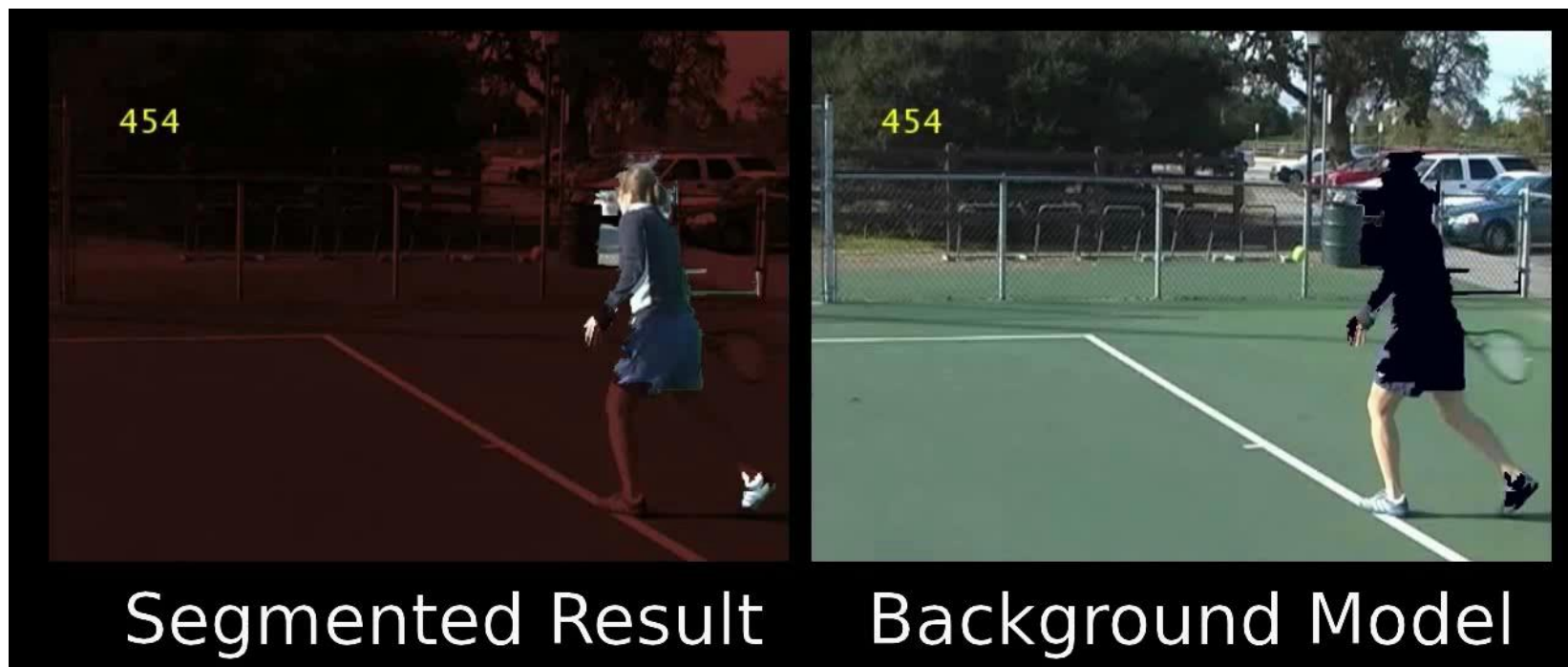


Proposed



# Восстановление фона

## Последовательность Tennis



# Сравнение с конкурентом

## Последовательность Tennis



# Сравнение с конкурентом

## Последовательность Drive1



Our result



Sheikh et al

# Сравнение с конкурентом

## Последовательность Marple





# Содержание

---

- Введение
- Общие подходы
- Сегментация на основе траекторий
- Online Dense Segmentation
- **Заключение**



# Заключение

- Очень популярны статистические и вероятностные методы
  - Что привело к успеху: сама модель или правильно подобранные картинки и параметры?
  - Часто очень медленные
- Стоит использовать информацию о движении с нескольких кадров (траектории)
- Нужно объединять сразу много признаков
- Возможно распространение разметки с надежных точек на ненадежные
- Полезно понижать размерность данных
- Если задача позволяет, можно использовать глобальную сегментацию

# Литература

1. R. Tron and R. Vidal, "A Benchmark for the Comparison of 3-D Motion Segmentation Algorithms," in *IEEE International Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–8.
2. T. Brox and J. Malik, "Object Segmentation by Long Term Analysis of Point Trajectories," in *European Conference on Computer Vision*, 2010, pp. 282–295.
3. A. Griesser, S. Roeck, A. Neubeck, and L. V. Gool, "GPU-Based Foreground-Background Segmentation Using an Extended Colinearity Criterion," in *Vision, Modeling, and Visualization*, 2005, pp. 319–326.
4. S. Brutzer, B. Hoferlin, and G. Heidemann, "Evaluation of Background Subtraction Techniques for Video Surveillance," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 1937–1944.
5. Y. Sheikh, O. Javed, and T. Kanade, "Background Subtraction for Freely Moving Cameras," in *IEEE International Conference on Computer Vision*, 2009, pp. 1219–1225.
6. A. Elqursh and A.M. Elgammal, "Online Moving Camera Background Subtraction," in *European Conference on Computer Vision (ECCV)*, 2012, pp. 228–241.
7. D. Hamad, P. Biela, "Introduction to spectral clustering," in *IEEE Information and Communication Technologies: From Theory to Applications*, 2008, pp. 1–6.

# Лаборатория компьютерной графики и мультимедиа



Видеогруппа — это:

- Выпускники в аспирантурах Англии, Франции, Швейцарии (в России в МГУ и ИПМ им. Келдыша)
- Выпускниками защищены 5 диссертаций
- Наиболее популярные в мире сравнения видеокодеков
- Более 3 миллионов скачанных фильтров обработки видео