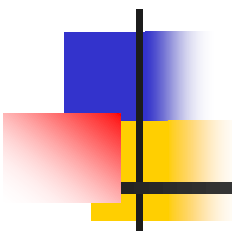


Обзор методов сегментации видео



Денис Сумин

Video Group

CS MSU Graphics & Media Lab



Содержание

- **Введение**
- Сегментация изображений
- Сегментация видео
- Video Segmentation by Tracking Regions
- Hierarchical Graph-Based Video Segmentation

Введение

Сегментация изображения — выделение регионов, соответствующих объектам





Применения

Компьютерное зрение:

- Распознавание движения
- Слежение за объектами
- Извлечение информации из изображений

Видео:

- Создание карт глубины
- Удаление объектов

Изображение → видео

Нельзя применять методы для сегментации изображений к каждому кадру видео:

- Сложно установить связь между сегментами от кадра к кадру
- Сегменты будут дрожать
- В видео информации гораздо больше, чем в одном кадре — ее надо использовать!



Содержание

- Введение
- **Сегментация изображений**
- Сегментация видео
- Video Segmentation by Tracking Regions
- Hierarchical Graph-Based Video Segmentation

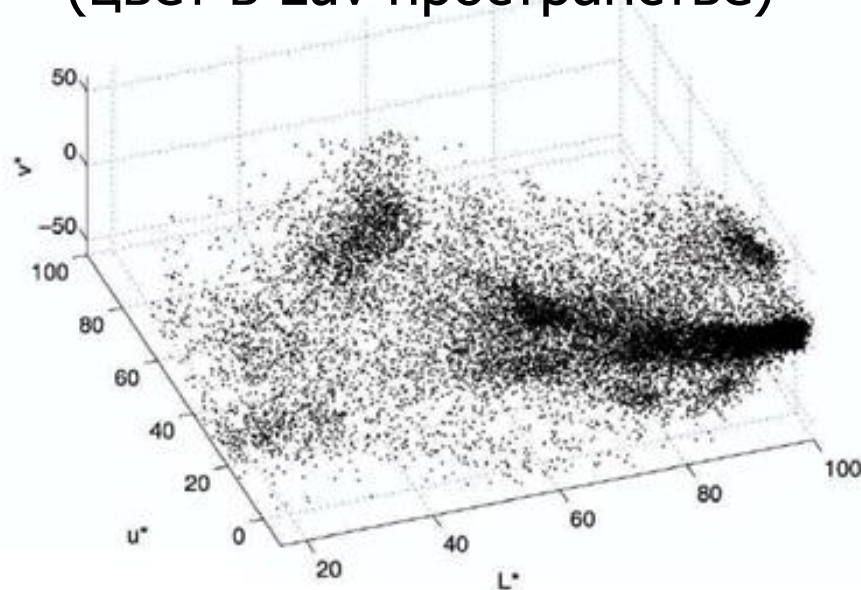
Mean Shift

Алгоритм ищет локальные максимумы плотности в поле свойств изображения

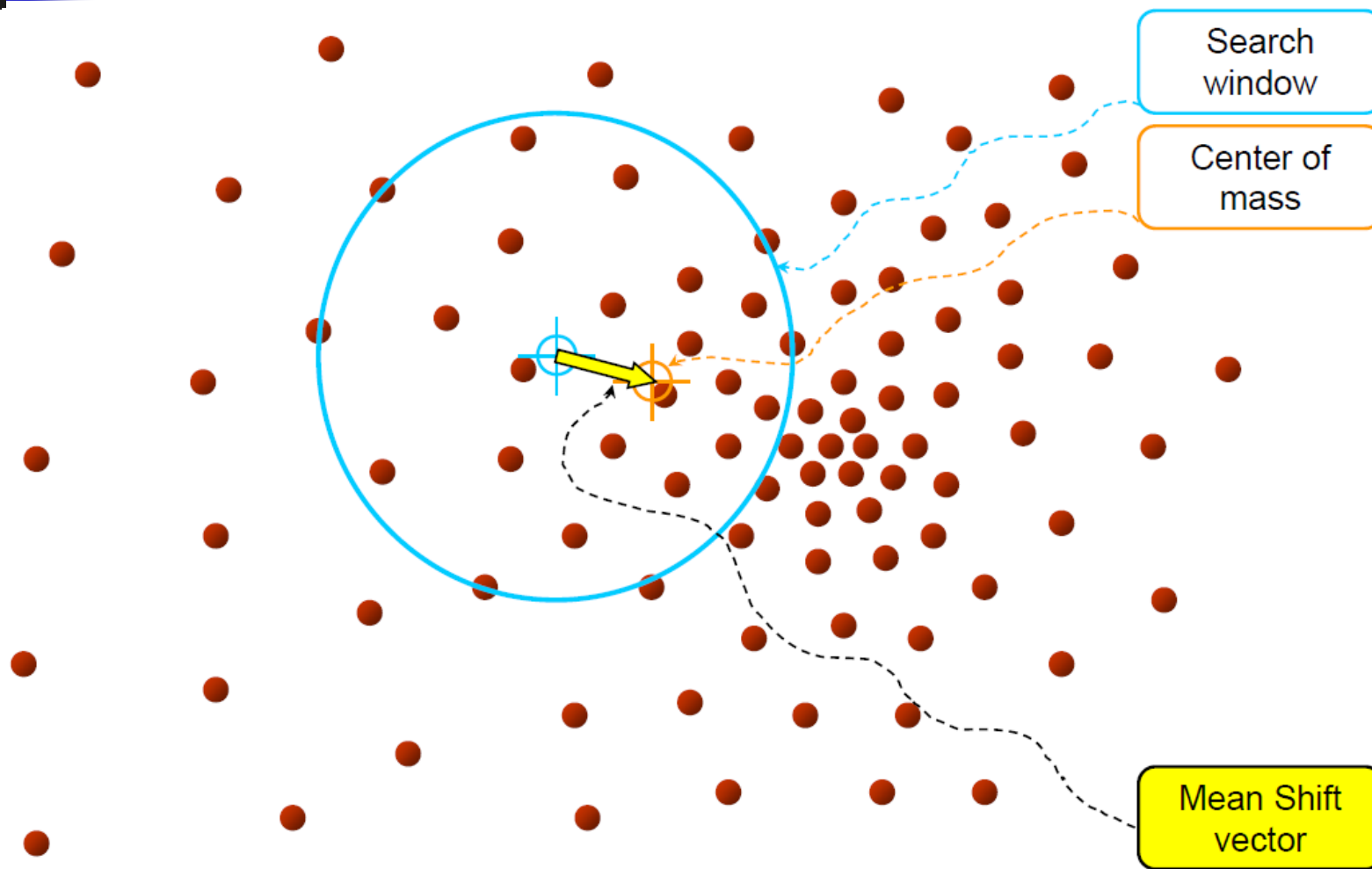
Изображение



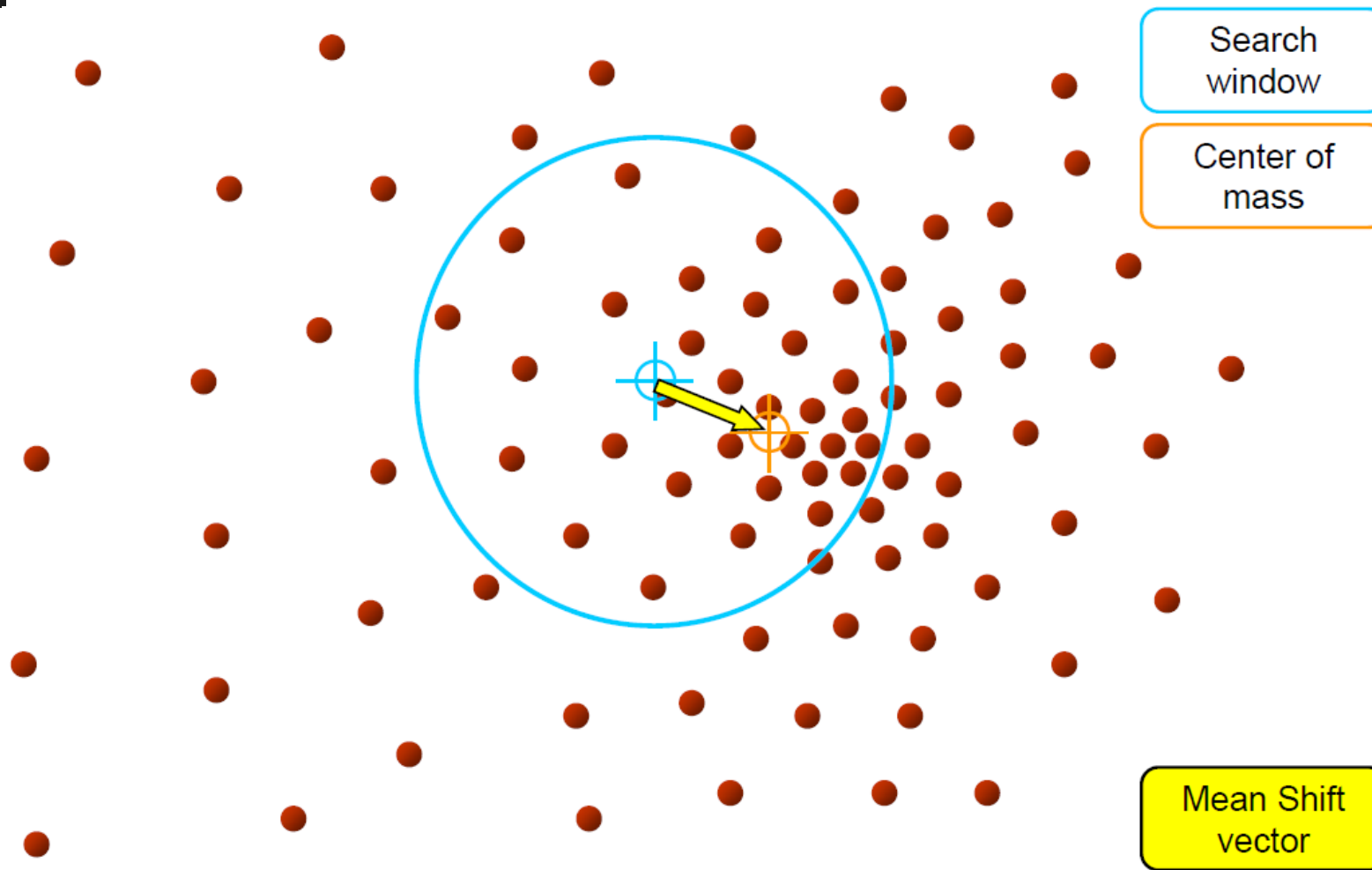
Поле свойств
(цвет в Luv пространстве)



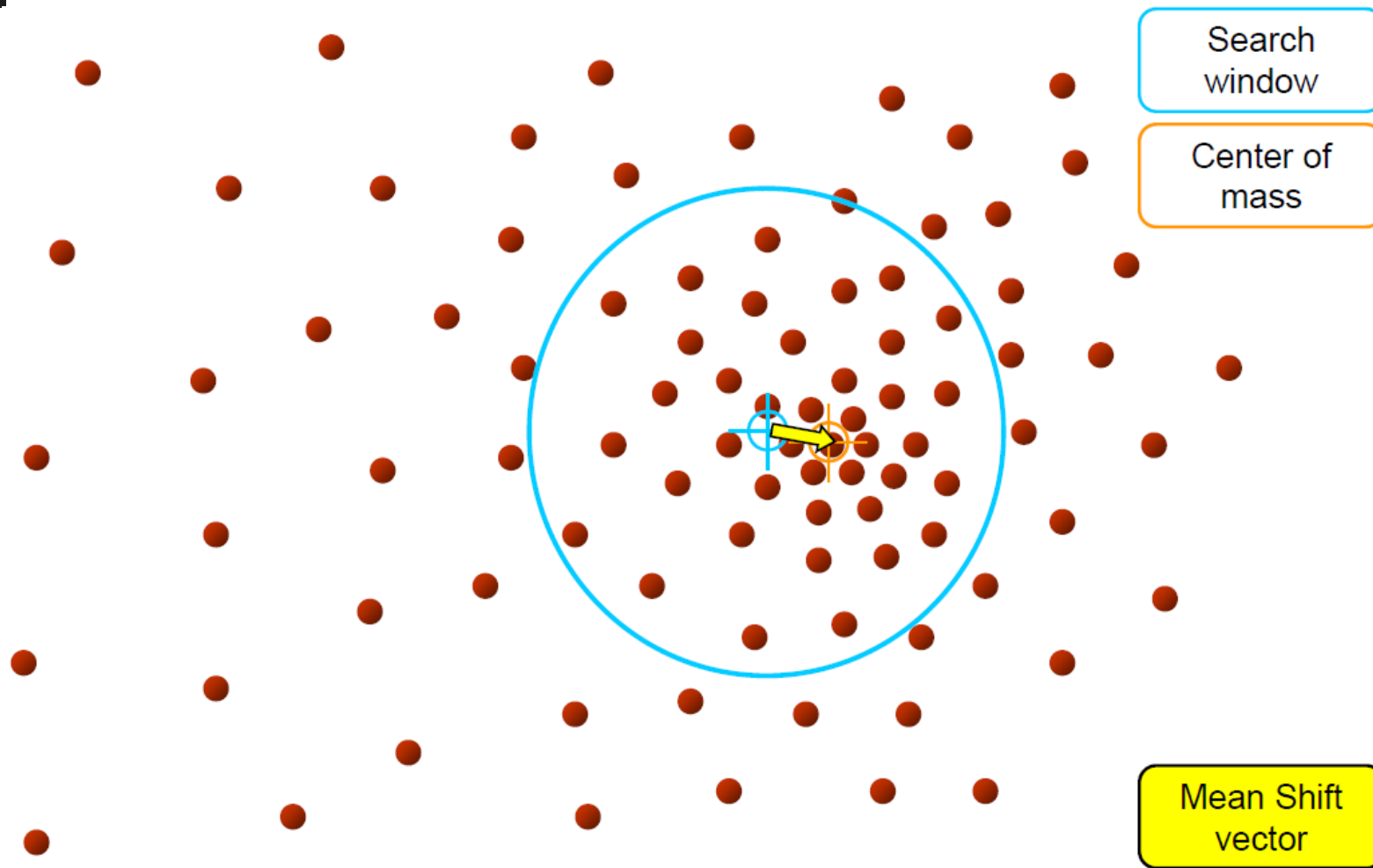
Mean Shift



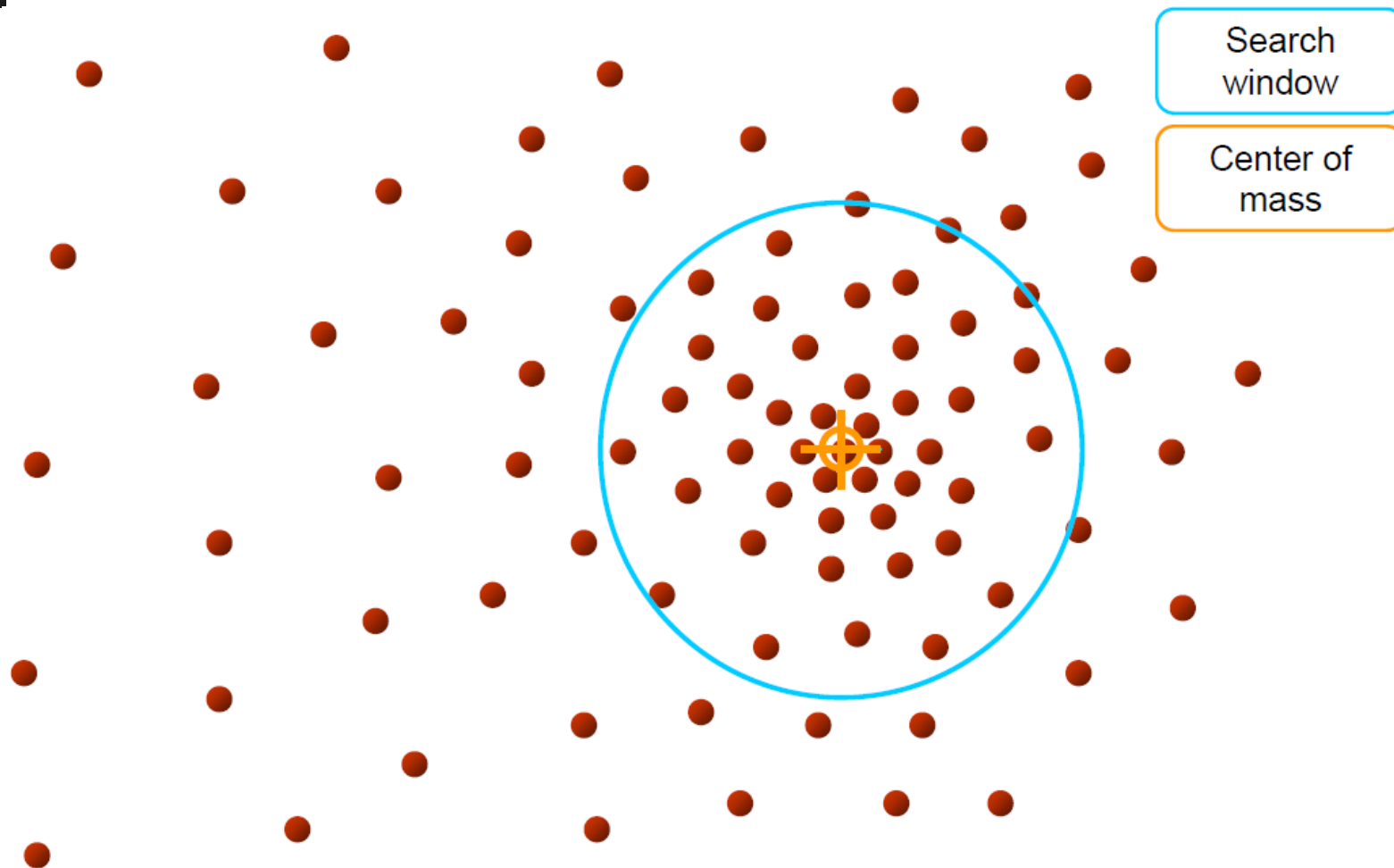
Mean Shift



Mean Shift

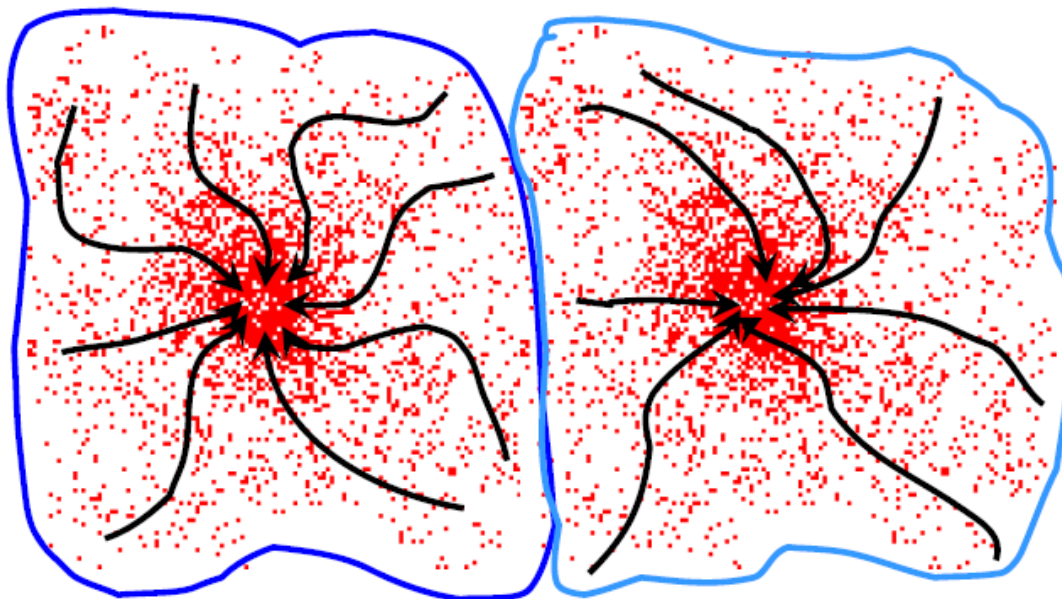


Mean Shift



Mean Shift

Кластер: точки, сдвиг среднего для которых приводит к одной и той же моде распределения



Mean Shift

Результаты



Mean Shift

Итоги

- Плюсы:
 - Не делается предположений о форме кластеров и форме распределения
 - Число кластеров определяется автоматически
- Минусы:
 - Выбор параметров ядровой функции нетривиален, и от них очень сильно зависит результат сегментации
 - Вычислительно сложный метод



Метод графов

Изображение — взвешенный
неориентированный граф

- Вершины графа — пиксели
- Ребра — связи между соседними пикселями
- Вес ребер пропорционален «похожести» пикселей

Критерии «похожести»

- Расстояние
- Цвет
- Яркость
- Текстура

Метод графов

Разрез графа $G=(V,E)$:

- Непересекающиеся подмножества вершин A и B из V
- Удаляем все ребра, связывающие A и B

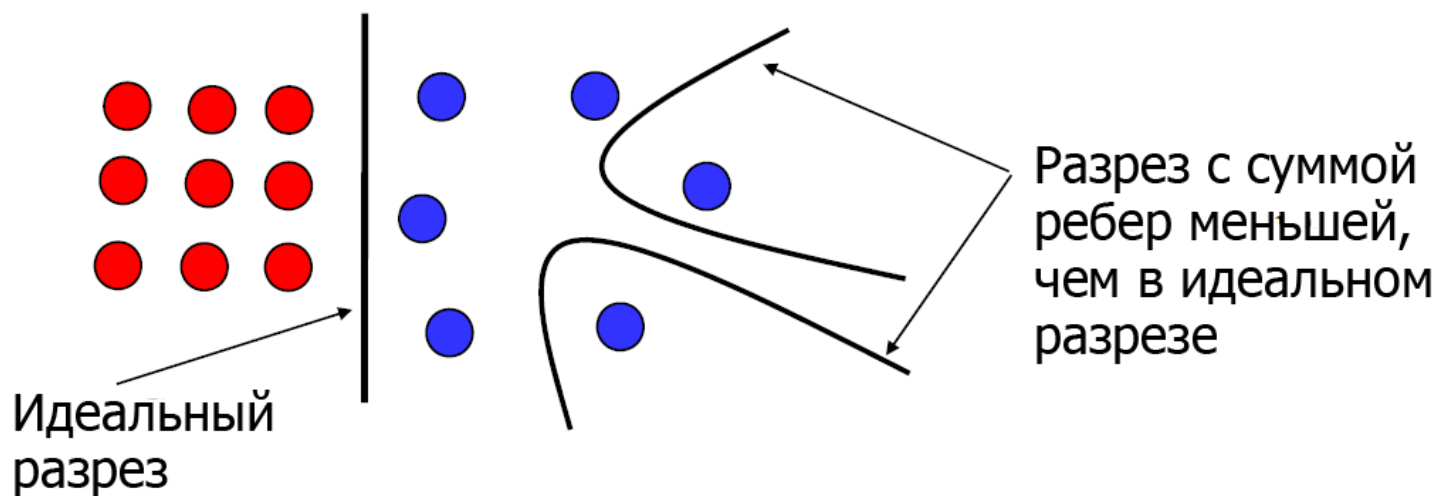
$$Cut(A, B) = \sum_{u \in A, v \in B} w(u, v)$$

- Разрез графа превращает граф в два несвязанных друг с другом подграфа

Метод графов (1 из 2)

Если множества A и B не заданы заранее, разрезать граф можно по-разному:

- Минимальный разрез: сумма весов удаленных ребер минимальна



Метод графов (2 из 2)

Если множества A и B не заданы заранее, разрезать граф можно по-разному:

- Нормализованный разрез: группы вершин нормируются на «объем», занимаемый ими в графе

$$Ncut(A, B) = 2 - \left(\frac{assoc(A, A)}{assoc(A, V)} + \frac{assoc(B, B)}{assoc(B, V)} \right) \quad assoc(A, V) = \sum_{u \in A, t \in V} w(u, t)$$

Все ребра графа \uparrow

- Минимальный нормализованный: величина $Ncut$ минимальна

Метод графов

Результат



Метод графов

Итоги

- Плюсы:
 - Хорошие результаты
 - Гибкость
- Минусы:
 - Требуется много памяти
 - Вычислительная сложность
 - В случае нормализованного разреза стремится к равным сегментам

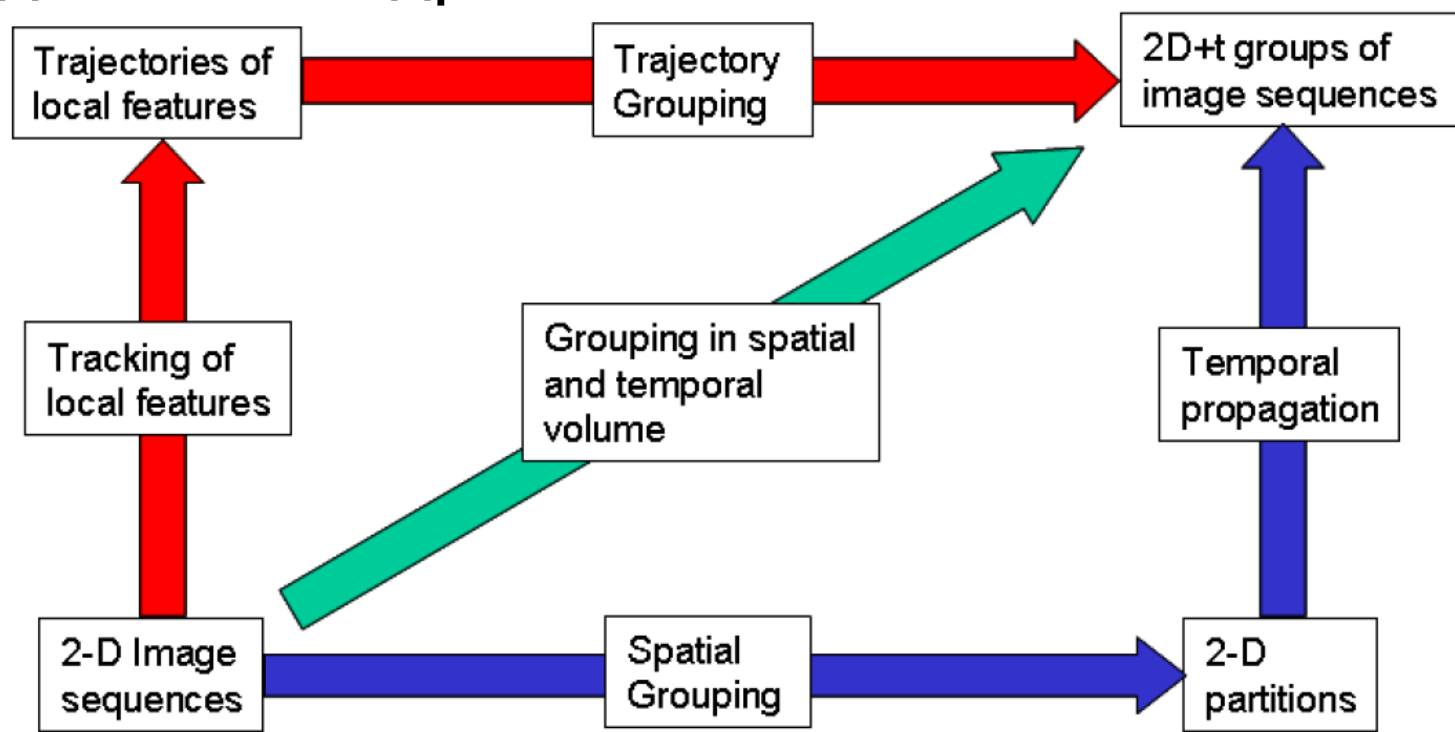


Содержание

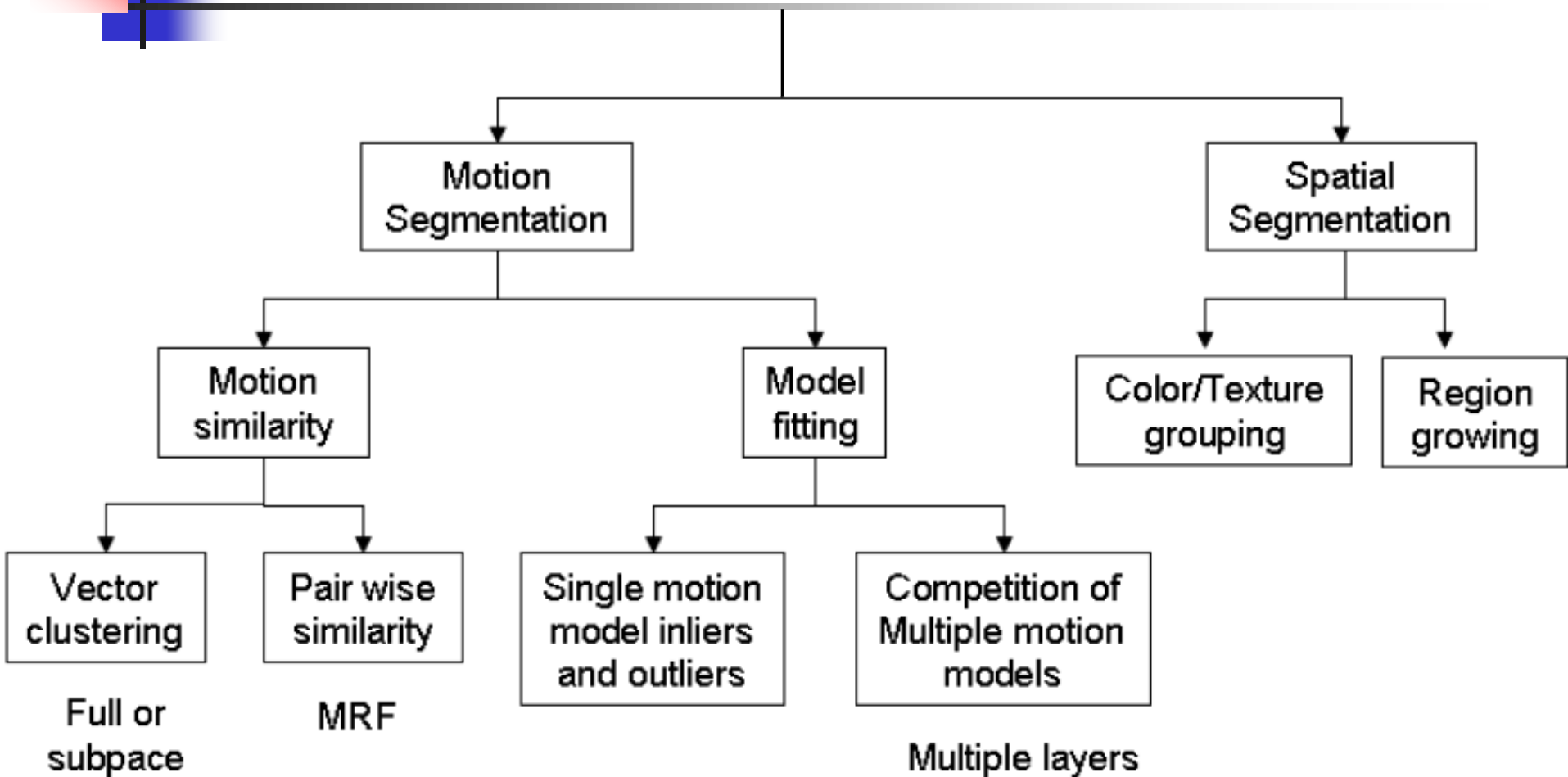
- Введение
- Сегментация изображений
- **Сегментация видео**
- Video Segmentation by Tracking Regions
- Hierarchical Graph-Based Video Segmentation

Сегментация видео

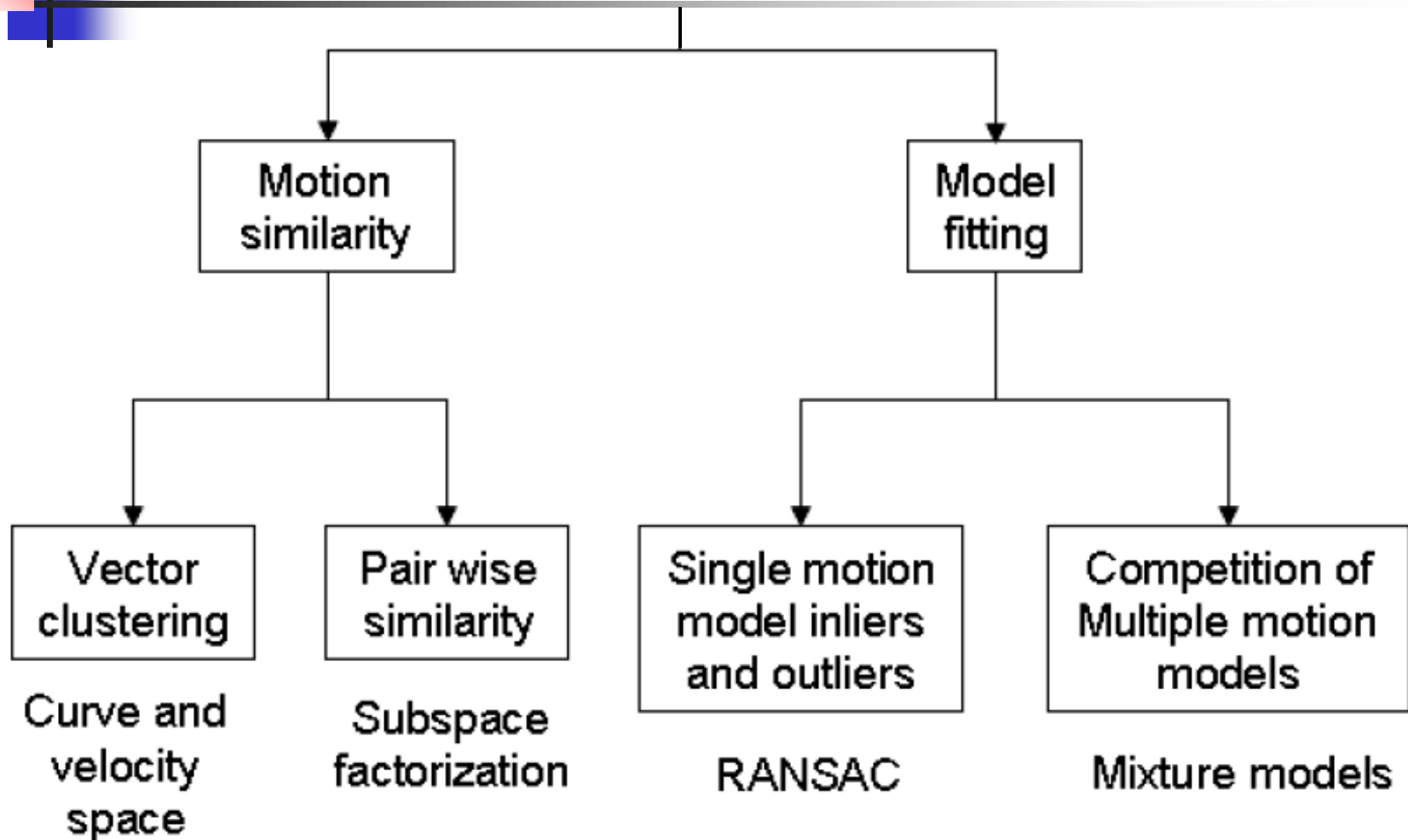
Все алгоритмы сначала сегментируют отдельные кадры



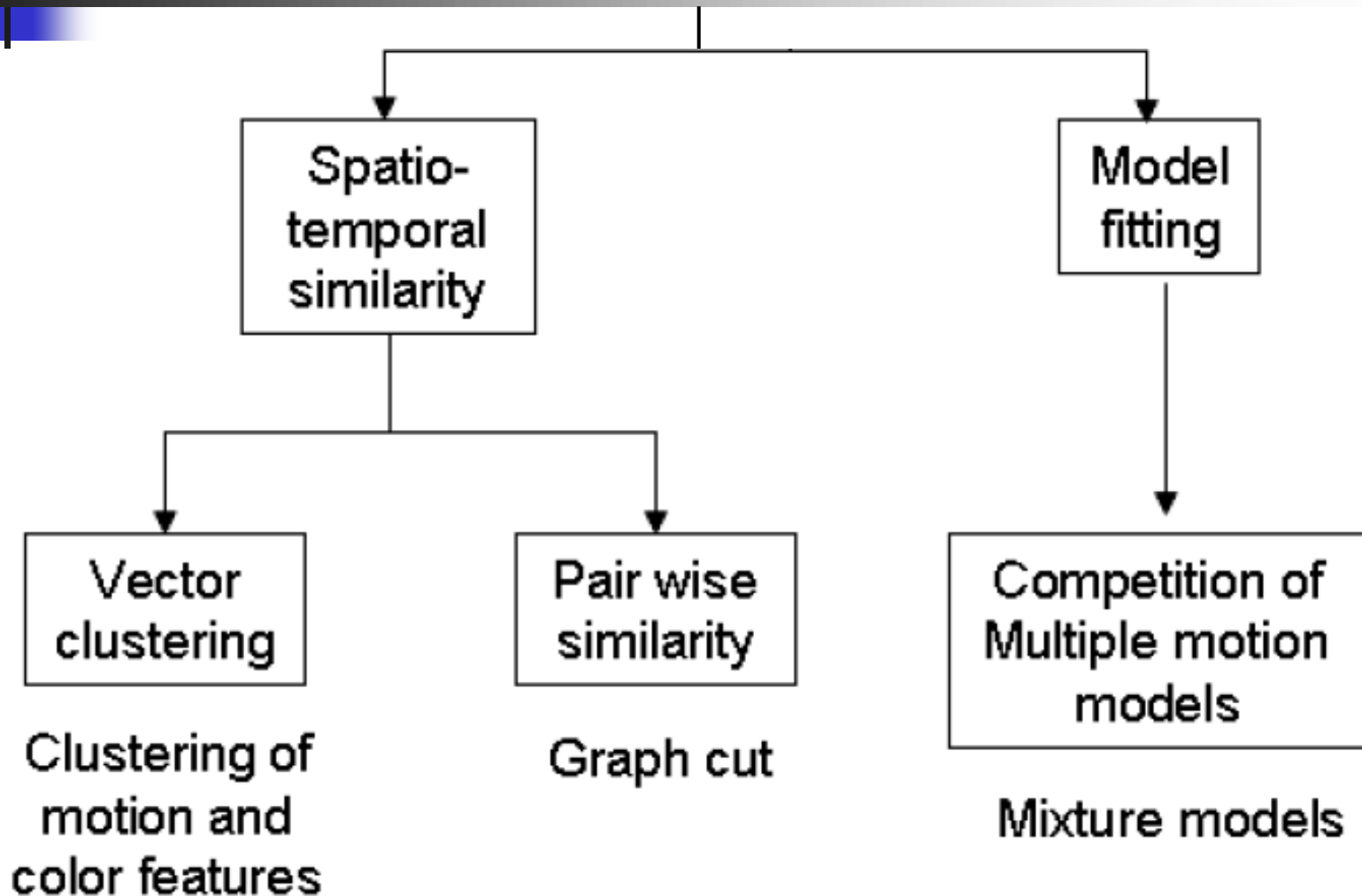
Spatial grouping approaches



Trajectory grouping approaches



Grouping in spatial and temporal volume





Содержание

- Введение
- Сегментация изображений
- Сегментация видео
- **Video Segmentation by Tracking Regions**
- Hierarchical Graph-Based Video Segmentation

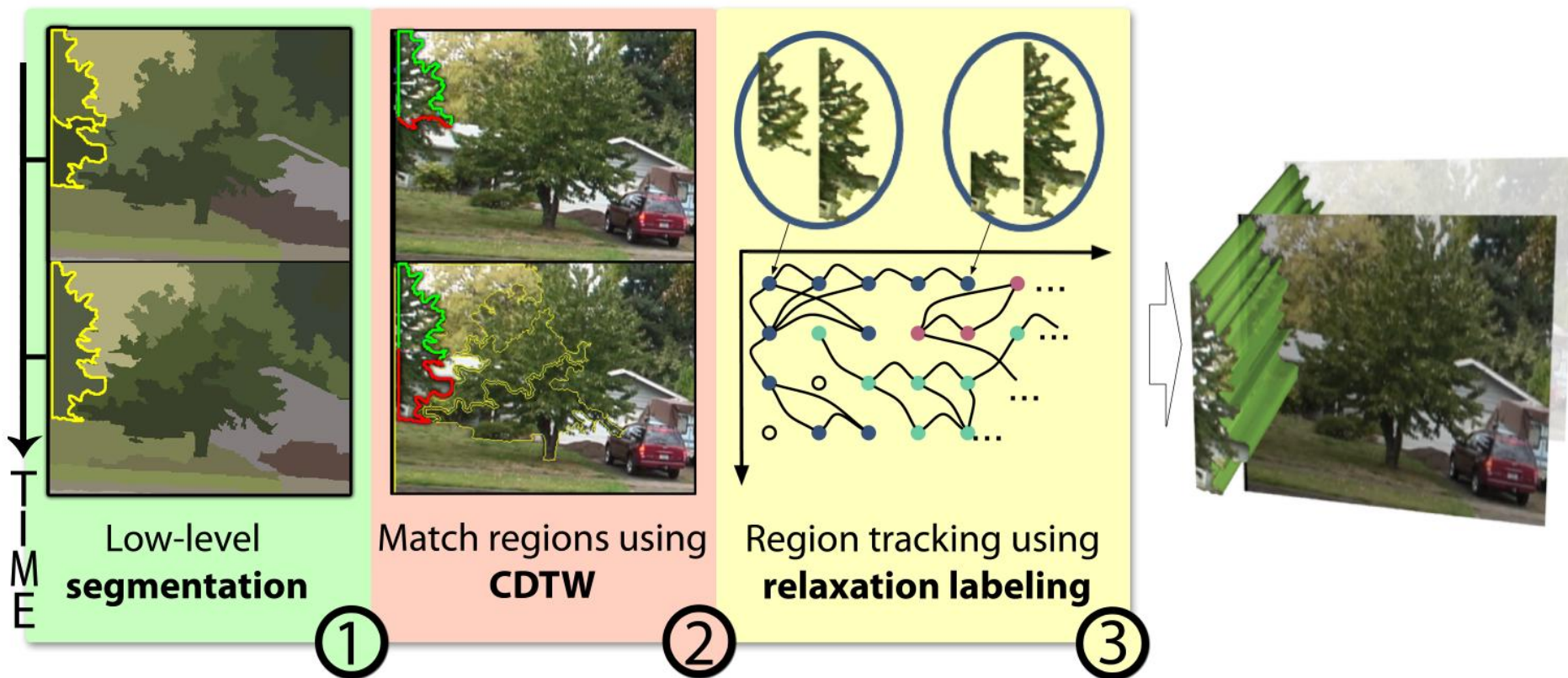
Tracking Regions

Алгоритм

1. Каждый кадр сегментируется независимо (например, mean shift)
2. Сегменты каждой пары последовательных кадров соотносятся между собой
3. Границы пар сегментов соотносятся с помощью CDTW (cyclic dynamic-time warping)
4. Полученные пары сегментов из всего видео кластеризуются

Tracking Regions

Алгоритм



Tracking Regions

DTW

- Две последовательности точек:
 $b_i = \{b_{i1}, \dots, b_{iM}\}$ и $b_j = \{b_{j1}, \dots, b_{jN}\}$
- $f = \{(b_{iu}, b_{jv}) : u=1..M, v=1..N\}$
- $c(b_{i.}, b_{j.})$ — мера различия точек
- Имея соответствия (b_{i1}, b_{j1}) и (b_{iM}, b_{jN}) ,
DTW находит оптимальное соответствие
последовательности точек

$$f^* = \arg \min_f \sum_f c(b_{iu}, b_{jv})$$

Tracking Regions

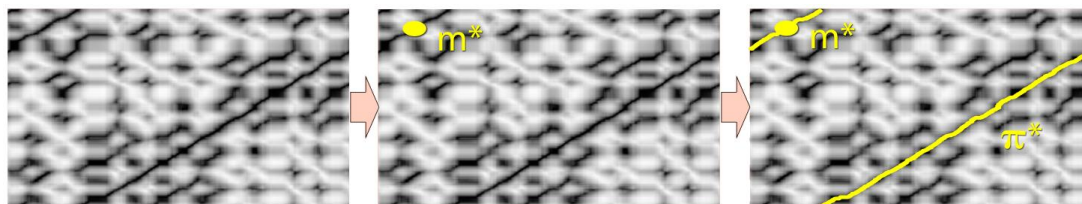
CDTW

$$\begin{aligned}
 m^* &= \max_{m \in C} P(m \in \pi^*), \\
 &= \max_{m \in C} P(\{\pi_m \sim \pi^*\}, \{\forall n \in C_m, \pi_n \sim \pi_m\}), \\
 &= \max_{m \in C} P(\{\pi_m \sim \pi^*\}) \prod_{n \in C_m} P(\{\pi_n \sim \pi_m\} | \{\pi_m \sim \pi^*\}).
 \end{aligned}$$

$$\begin{aligned}
 P(\{\pi_m \sim \pi^*\}) &\propto \exp(-\mu c(\pi_m)), \\
 P(\{\pi_n \sim \pi_m\} | \{\pi_m \sim \pi^*\}) &\propto \exp(-\lambda |c(\pi_n) - c(\pi_m)|),
 \end{aligned}$$

π^* — оптимальная граница C_m — окрестность точки m
 π_m — граница через точку m $1/\lambda$ и $1/\mu$ — ML-оценки

ЭЛЕМЕНТОВ В C



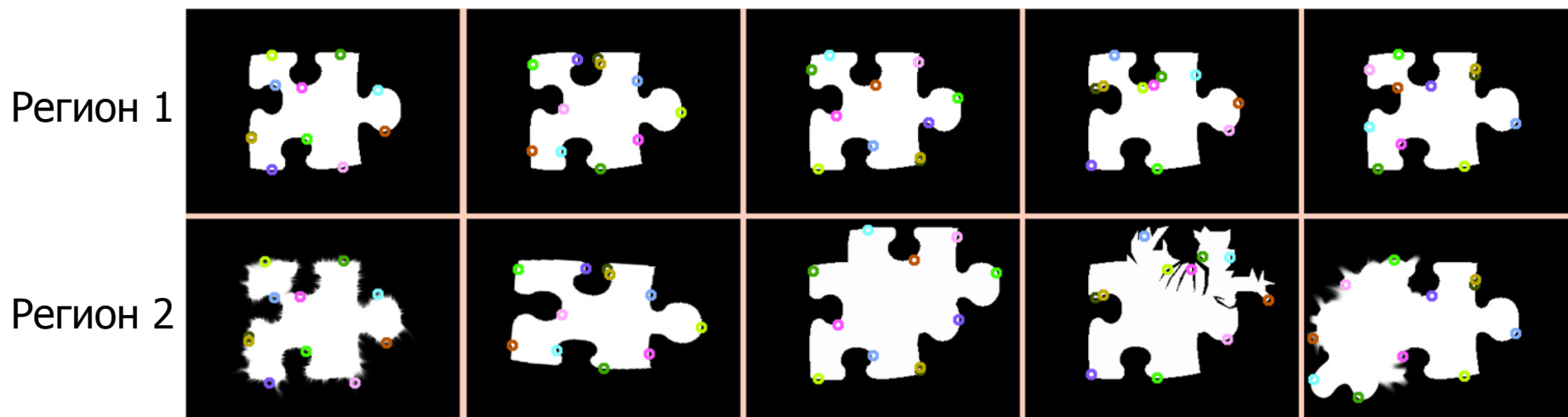
cost matrix C

find starting point m^*

DTW from m^*

Tracking Regions

Устойчивость CDTW



Tracking Regions

Region tracking

- Оцениваем правдоподобность границы на предыдущем шаге $s(i, j) = \exp(-c(i, j))$
- Все правдоподобности записываем в граф G
- Применяем метод кластеризации relaxation labeling (RL) к графу G
- Считаем вероятности правильного сегментирования для всего видео

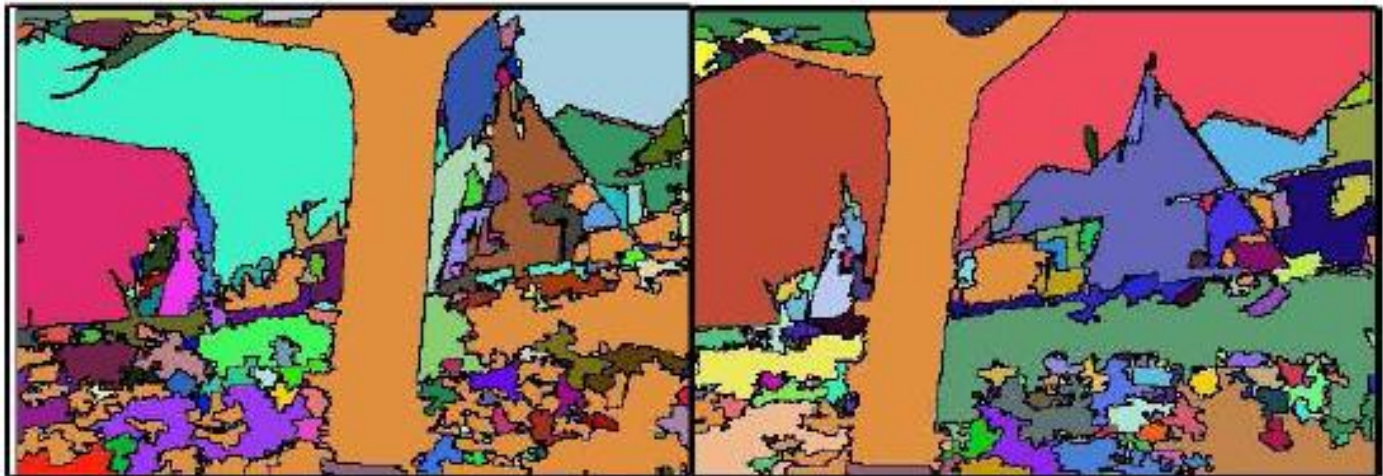
Tracking Regions

Результаты

Оригинальное
видео



Результат



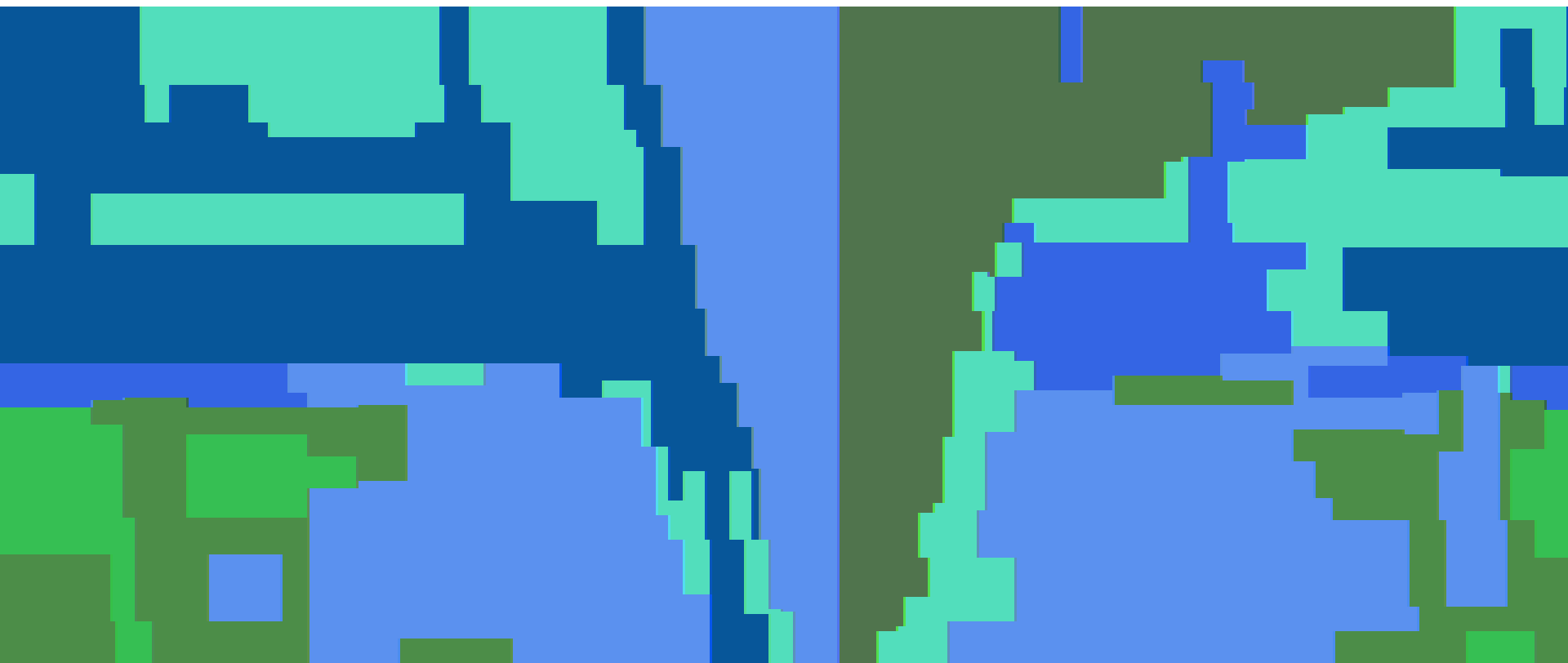
Tracking Regions

Результаты



Tracking Regions

Результаты



Tracking Regions

Итоги



- Плюсы:
 - Скорость
- Минусы:
 - Среднее качество

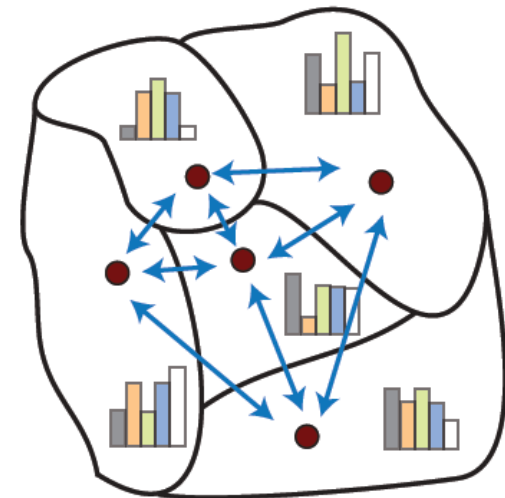


Содержание

- Введение
- Сегментация изображений
- Сегментация видео
- Video Segmentation by Tracking Regions
- **Hierarchical Graph-Based Video Segmentation**

Hierarchical Graph-Based Алгоритм

1. Обрабатываем каждый пиксель кадра, учитывая его 26 соседей. Получаем мелкие сегменты
2. Добавляем Lab гистограммы
3. Укрупняем сегменты на основе имеющейся информации. Сохраняем всю информацию в дерево



Hierarchical Graph-Based Алгоритм



Алгоритм хорошо работает, но требует слишком много памяти

- Работаем с сегментами внутри некоторого окна
- 3 типа краев: граничный, внешний, внутренний
- Постепенно увеличиваем окна

Получается тот же результат, но алгоритм получается масштабируемым

Hierarchical Graph-Based Clip-based processing



- Разбиваем все видео на клипы по 20–30 кадров
- В каждый клип добавляем треть кадров из предыдущего клипа
- Для пересечения считаем

$$S(e_{p,q}) = \begin{cases} \alpha & \text{if } R_i(p) = R_i(q), \\ 100 \times (1 - \alpha) & \text{otherwise,} \end{cases}$$

Функция R_i сопоставляет id сегментов пикселям
 $\alpha \in [0, 1]$

Hierarchical Graph-Based Результаты



Hierarchical Graph-Based Результаты



Hierarchical Graph-Based Результаты



Hierarchical Graph-Based Результаты



Hierarchical Graph-Based

Итоги



- Плюсы:
 - Очень хорошее качество
- Минусы:
 - Для действительно хорошего качества используется Optical Flow => потеря скорости

Литература

1. William Brendel, Sinisa Todorovic "Video Object Segmentation by Tracking Regions", in Proc. IEEE Int'l Conf. Computer Vision (ICCV), Kyoto, Japan, 2009
2. Matthias Grundmann, Vivek Kwatra, Mei Han, Irfan Essa, "Efficient Hierarchical Graph-Based Video Segmentation", CVPR 2010
3. Liu Siying, "Video Segmentation Temporally-constrained Graph-based Optimization", National University of Singapore, 2010
4. www.cs.unc.edu/~lazebnik
5. А. Конушин, «Методы сегментации изображений»

