Распознавание динамических жестов на основе медиального представления формы изображений

Куракин А.В.

Московский Физико-Технический Институт

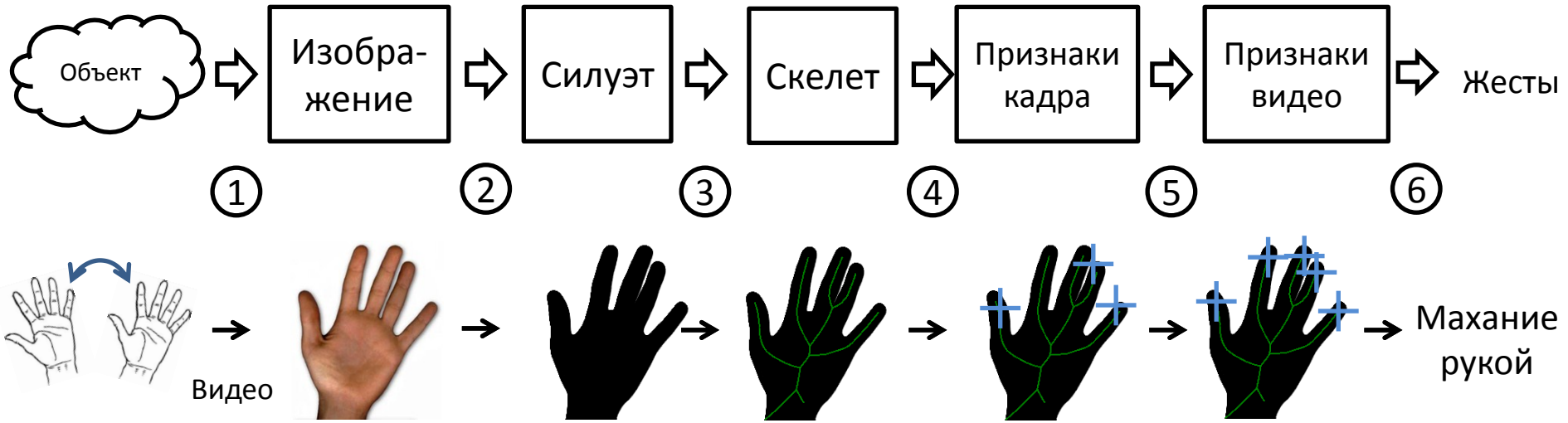
Местецкий Л.М.

МГУ им. Ломоносова

Задача

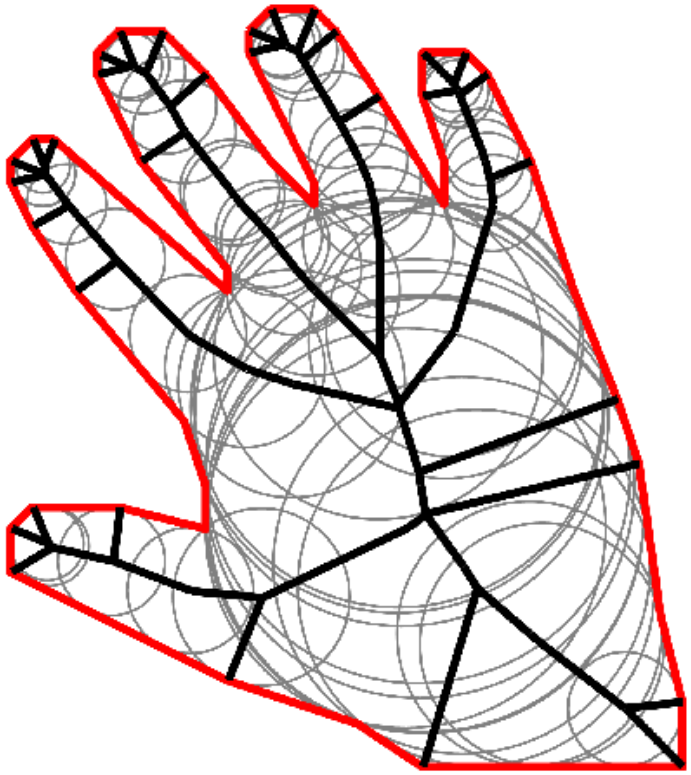
- Распознавание жестов (рук и тела)
- Динамические жесты
- Медиальное представление формы для выделения признаков

Структура предлагаемого подхода



1. Получение изображения
2. Бинаризация
3. Построение медиального представления
4. Выделение признаков для каждого отдельного кадра
5. Межкадровая обработка признаковых описаний
6. Распознавание жестов

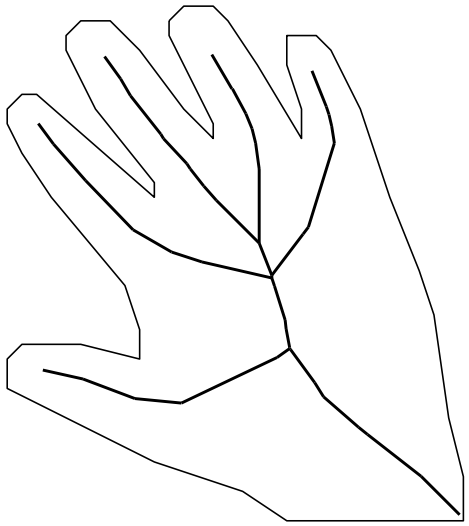
Скелет



Скелет (или **серединные оси**) фигуры – множество центров и радиусов вписанных в фигуру кругов.

Свойства скелета

- Скелет = объединение непрерывных кривых;
  Скелет = граф
- Вершины графа имеют степень 1, 2 или 3
- С каждой точкой скелета связана радиальная функция
  R(●) – расстояние до границы



Ветвь скелета

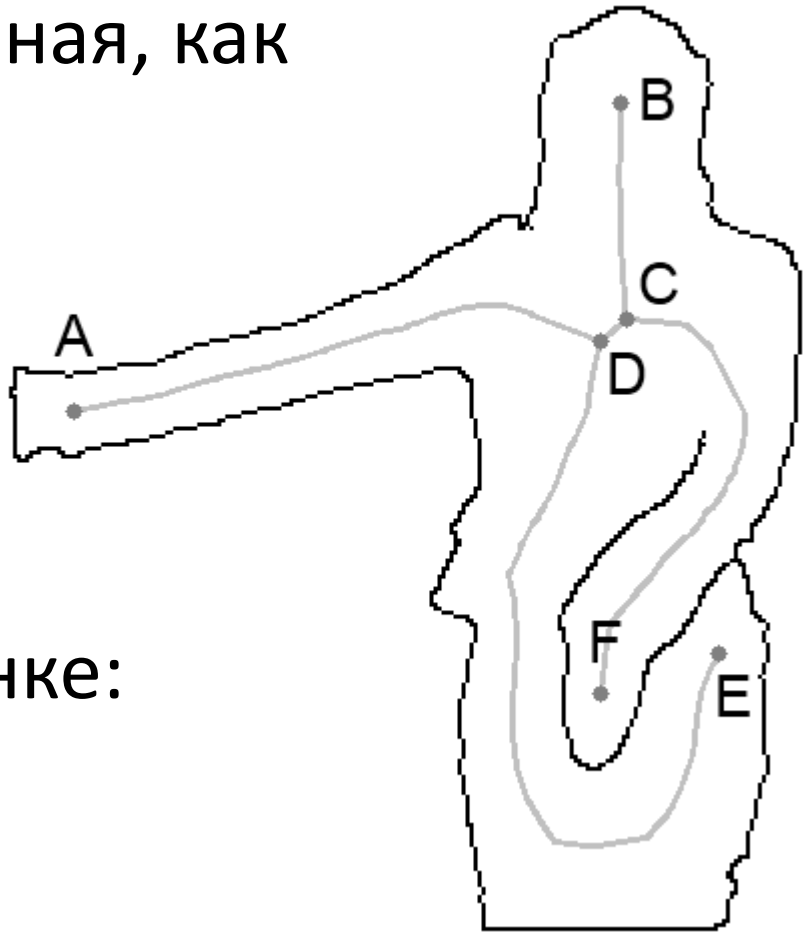
Ветвь скелета –

часть скелета рассмотренная, как непрерывная кривая,

в каждой точке ветви определена рад. ф-ция

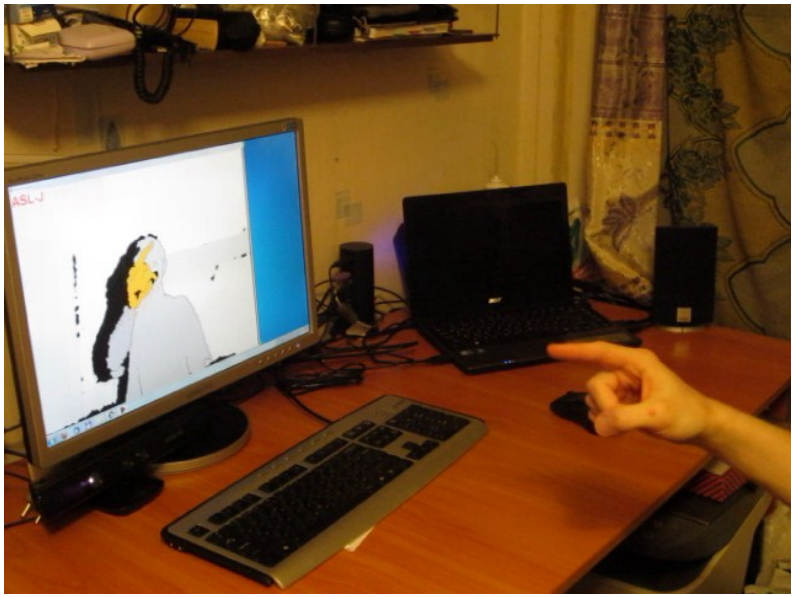
Примеры ветвей на рисунке:

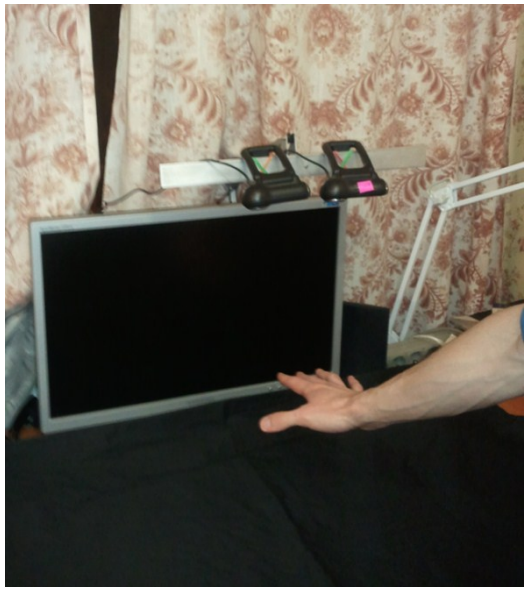
AD, BC, ADCF, DCB, …



1. Получение изображений

- Одна или две RGB камеры





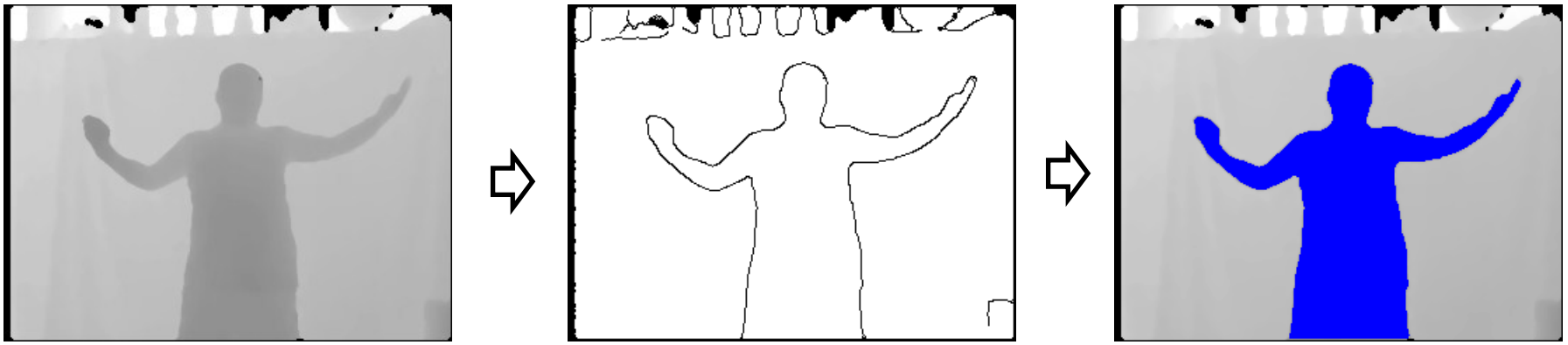
- Камера глубины (Kinect)

- База соревнования ChaLearn Gesture Challenge (видео с камеры глубины)

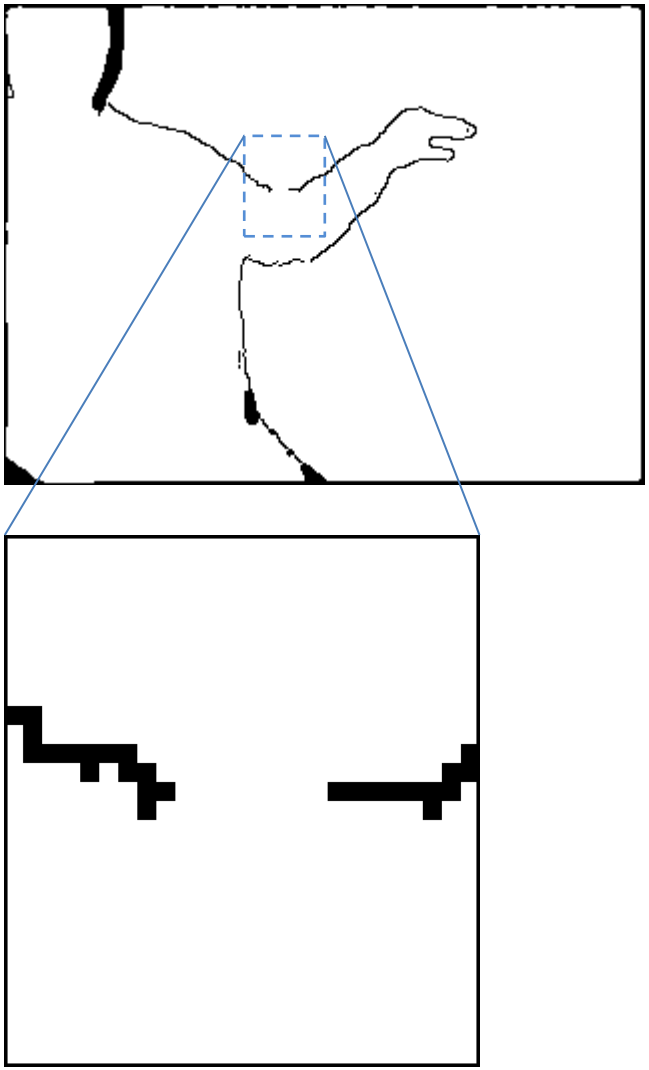
2. Выделение силуэта объекта

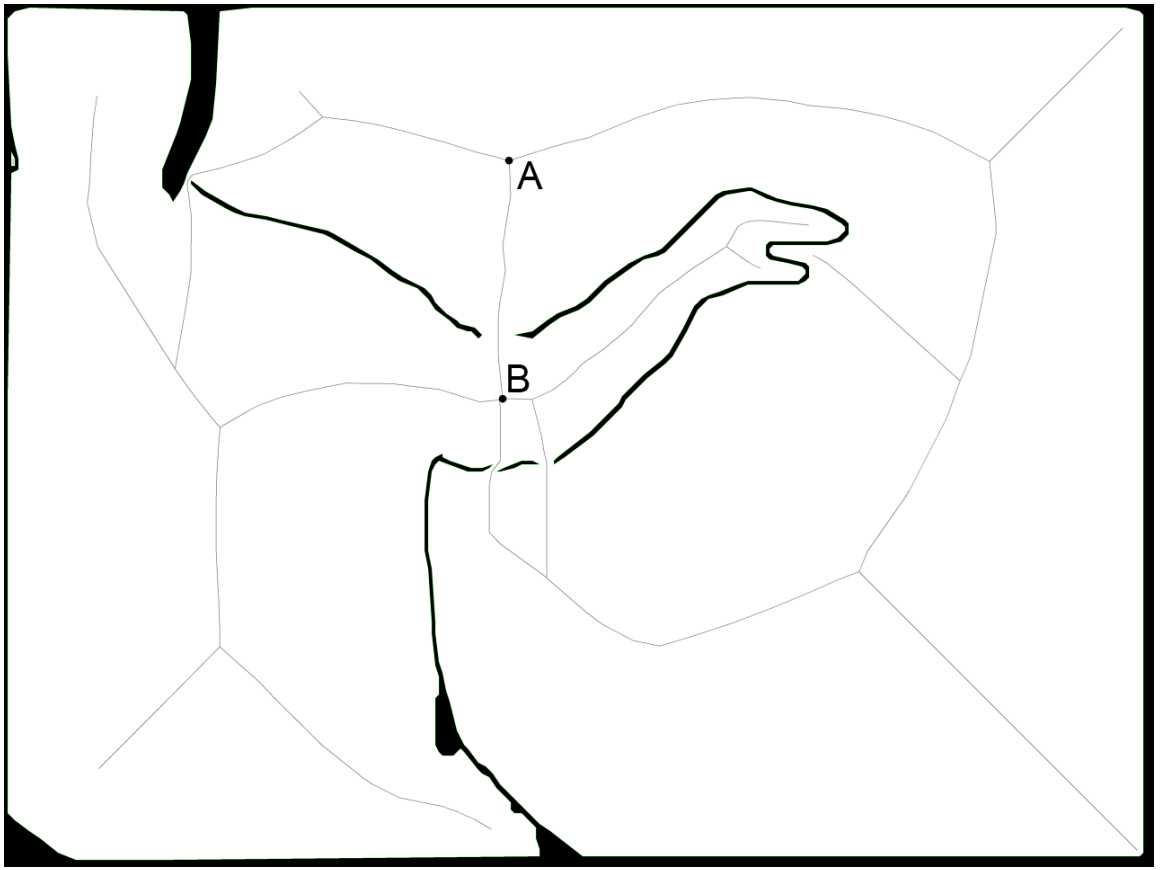
- Для RGB камер силуэт выделялся с помощью вычитания фона, для упрощения фон был однородный

- Для камеры глубины:
  - Существенные перепады глубины – границы объектов
  - Для устранения разрывов выполнялась сшивка границ

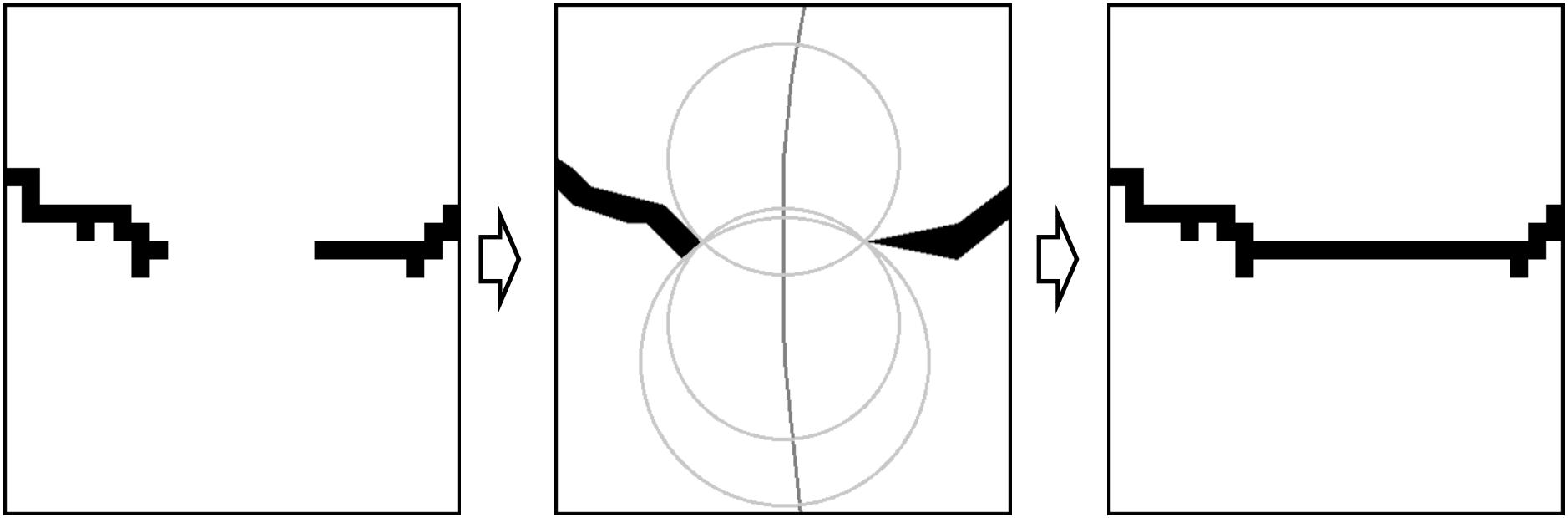


Сшивка границ

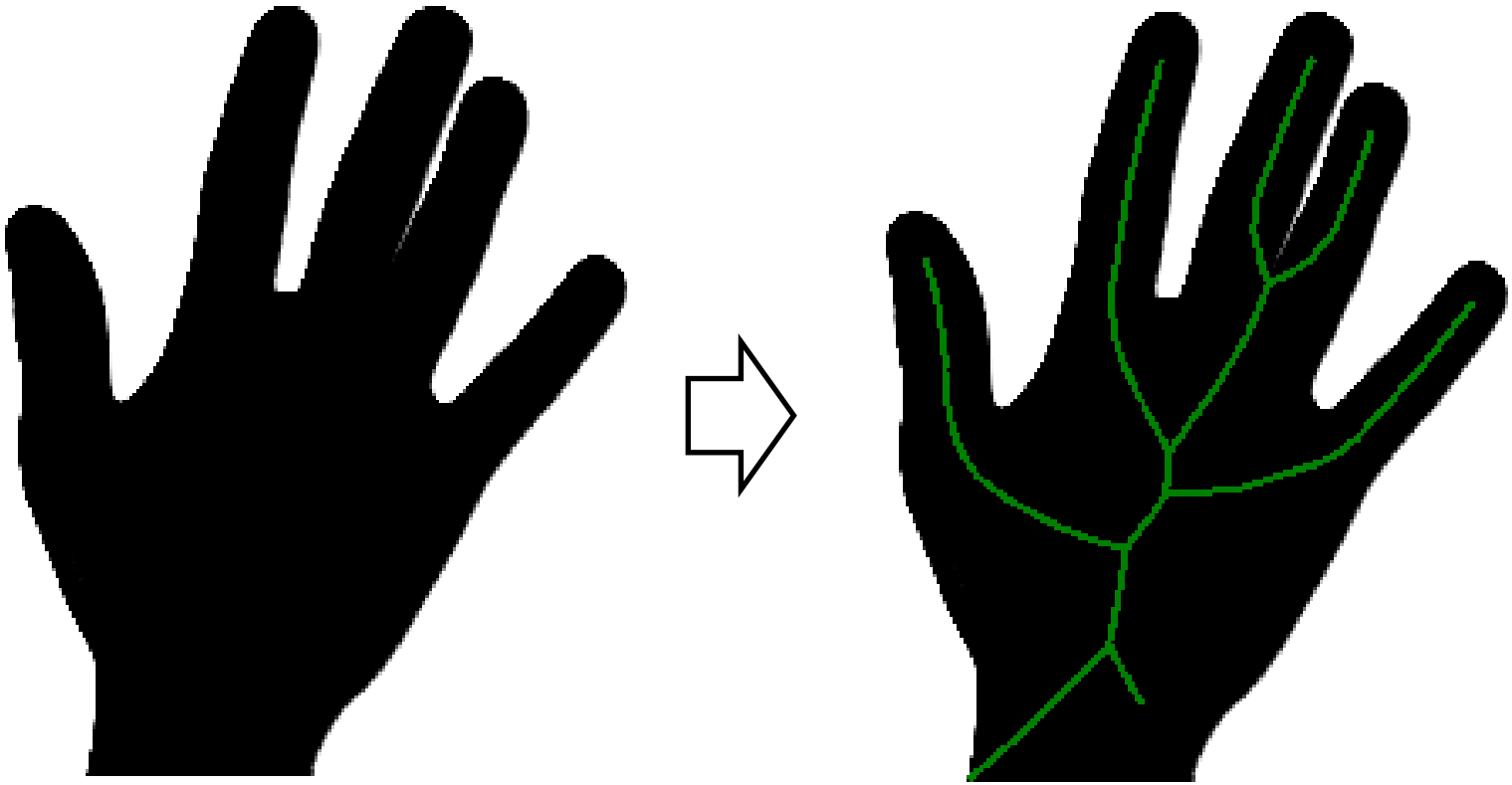




Сшивка границ

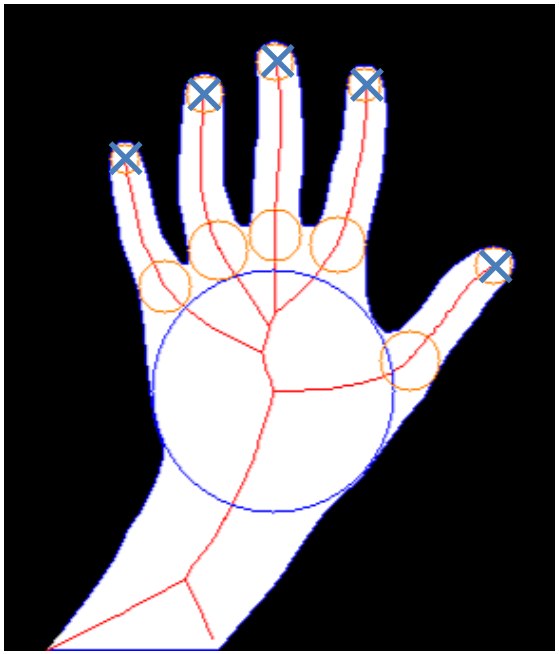


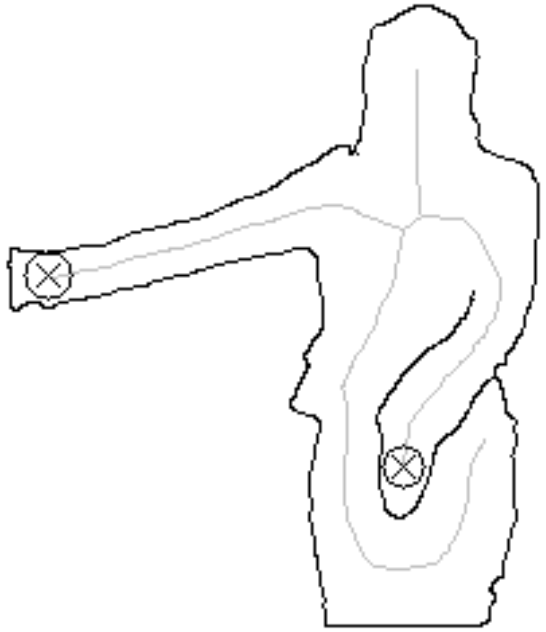
3. Построение медиального представления



4. Генерация признаков

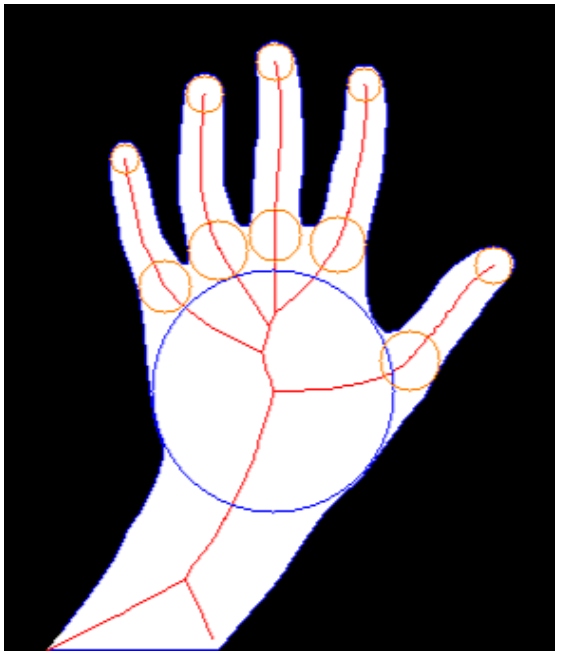
- Признаки = координаты «ключевых» точек объекта
- Ключевые точки: кончики пальцев, руки



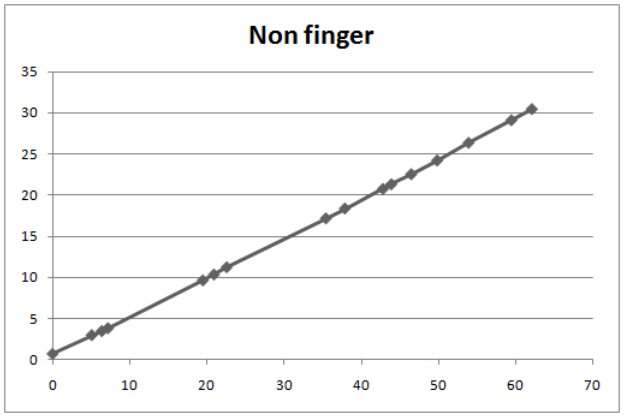


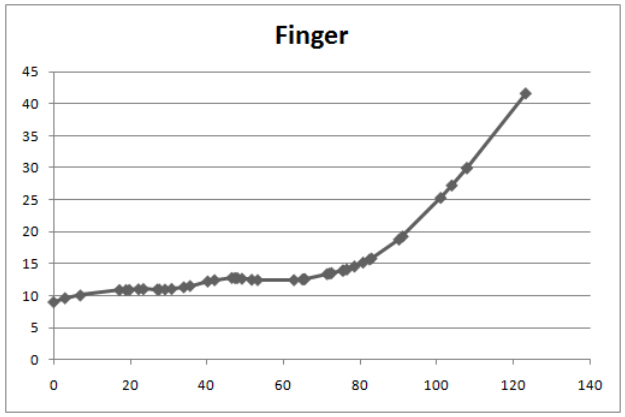
4. Генерация признаков

Ключевые точки среди *терминальных вершин скелета*



Идея классификации терминальных вершин – использовать радиальную функцию вдоль ветви





4. Генерация признаков

Для обнаружения ключевых точек:

рассматриваем ветви соединяющие вершины степени 1 и 3:

AD, BC, FC, ED.

Классифицируем каждую ветвь на два класса:

Класс 1 = есть ключевая точка

Класс 0 = нет ключевой точки

4. Генерация признаков

Признаки для классификации ветви скелета:

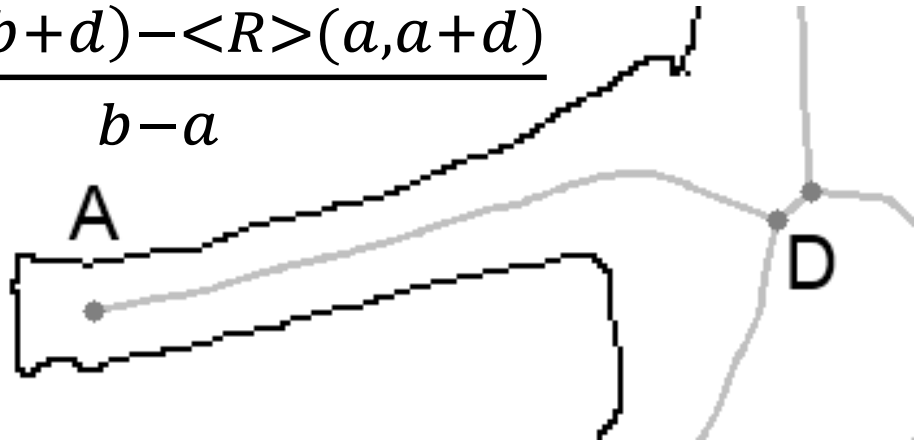
1) Значение радиальной функции R(x)
2) Среднее значение радиальной функции

$$< R > (a, b) = \int_a^b R(x)dx$$

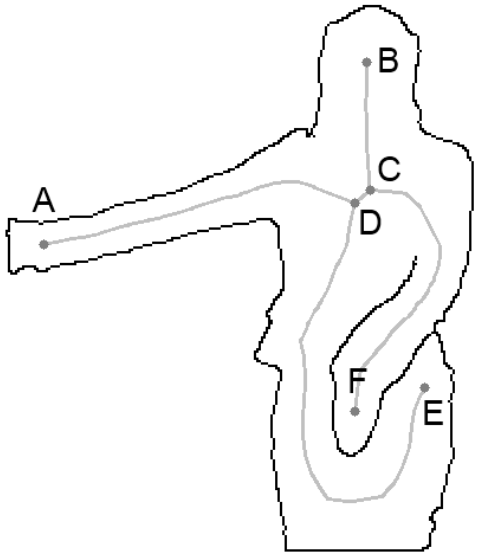
3) Скорость роста радиальной функции

$$Gr(a, b, d) = \frac{<R>(b,b+d) - <R>(a,a+d)}{b-a}$$

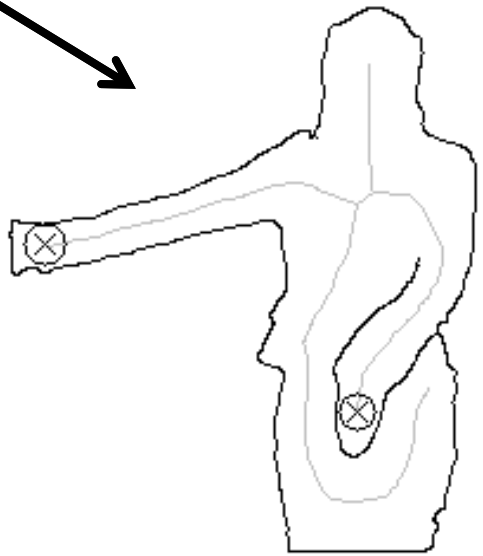
4) Длина ветви L



4. Генерация признаков



Обнаружение ключевых точек

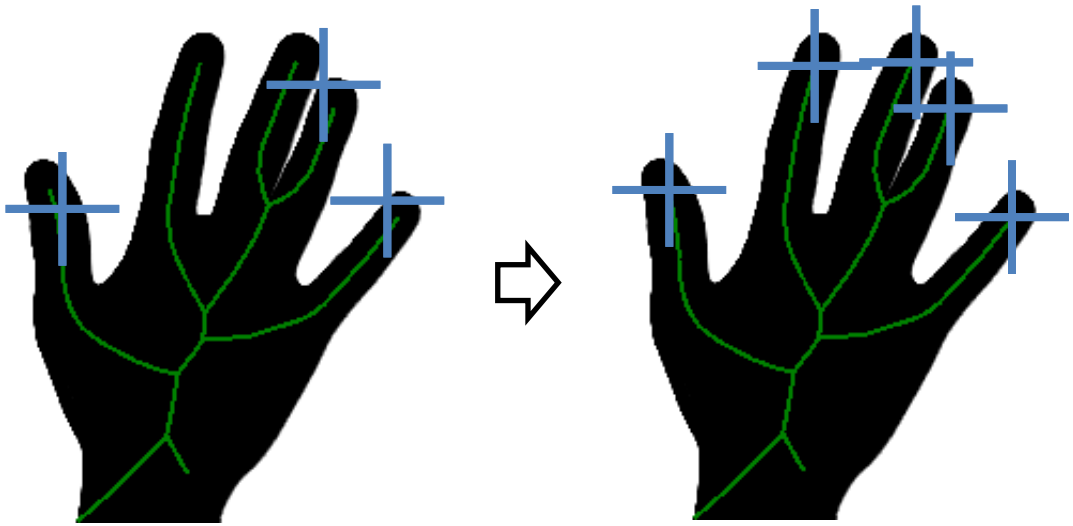


$$\begin{pmatrix} x_1 \\ y_1 \\ z_1 \\ x_2 \\ y_2 \\ z_2 \end{pmatrix}$$

Вектор признаков

Размерность признакового описания отличается от кадра к кадру!

5. Межкадровая фильтрация

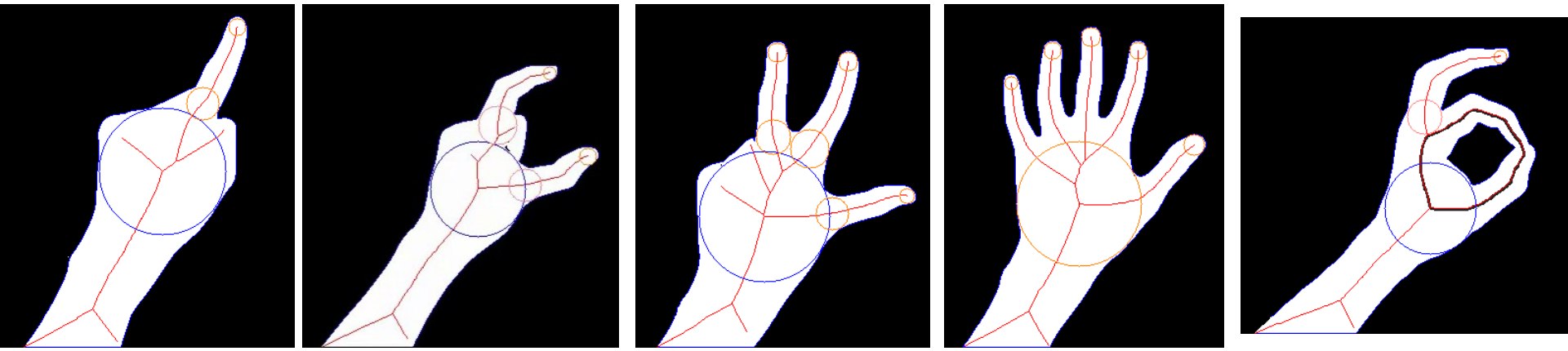


1. Прослеживание траекторий ключевых точек
2. Заполение пропусков в траекториях
3. Фильтрации координат в траекориях

6. Распознавание жестов

- Статические жесты или простые дин. жесты:
  - Набор эвристических правил
- Сложные динамические жесты:
  - Сравнение с образцом на основе метрики

Распознавание жестов на основе правил

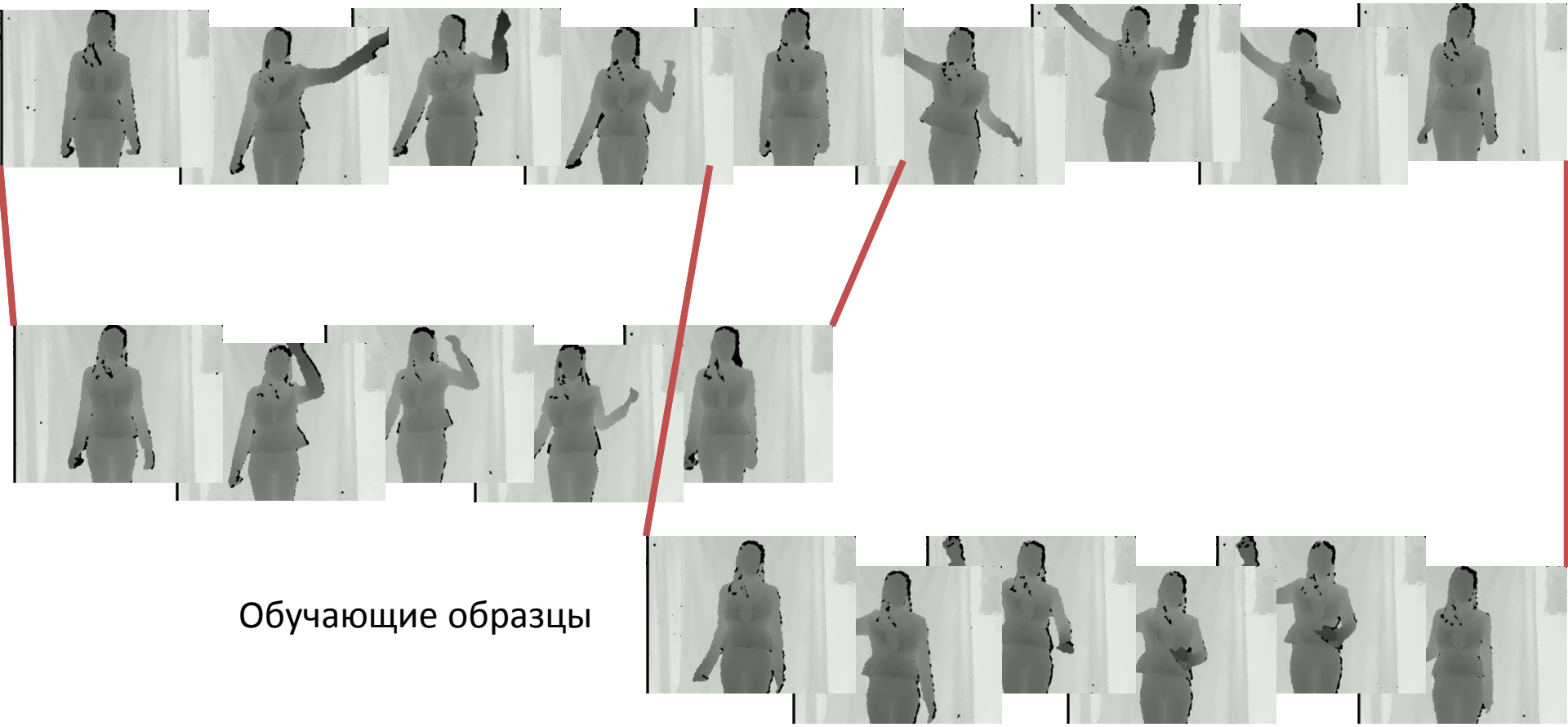


Практическая задача – управление мышью и объектами на экране компьютера с помощью рук.

Жесты различаются количеством видимых пальцев.
Координаты пальцев – координаты курсора.

Метрическое распознавание жестов

Контрольное видео с несколькими жестами



Обучающие образцы

Метрическое распознавание жестов

Мера сходства непрерывных кривых **F** и **F’**:

$$similarity_{cont}(\mathbf{F}(\bullet), \mathbf{F}'(\bullet)) = \min_{\substack{w(t) - \text{монотонная} \\ w(0)=0,\ w(l)=l'}} \int_0^l \|\mathbf{F}(t) - \mathbf{F}'(w(t))\| dt$$

Для жестов G и G’ с дискретными траекториями ($\mathbf{F}_1$, …, $\mathbf{F}_{|G|}$) и ($\mathbf{F'}_1$, …, $\mathbf{F'}_{|G'|}$) мера сходства:

$$similarity(G, G') = \frac{1}{m} \min_{m, u(\bullet), v(\bullet)} \sum_{k=1}^{m} d(\mathbf{F}_{u(k)}, \mathbf{F}'_{v(k)})$$

при u(1) = v(1) = 1, u(m) = |G|, v(m) = G’,
u(k) <= u(k+1) <= u(k+1),
v(k) <= v(k+1) <= v(k) + 1
u(k) < u(k+1) или v(k) < v(k+1)

Метрическое распознавание жестов

G_1, … G_N – множество эталонных жестов,
составляющих обучающую совокупность.
V – видео для распознавания.
Распознавание производиться методом ближайшего соседа:

$$\hat{i} = \underset{i}{argmin}\ similarity(V, G_i)$$

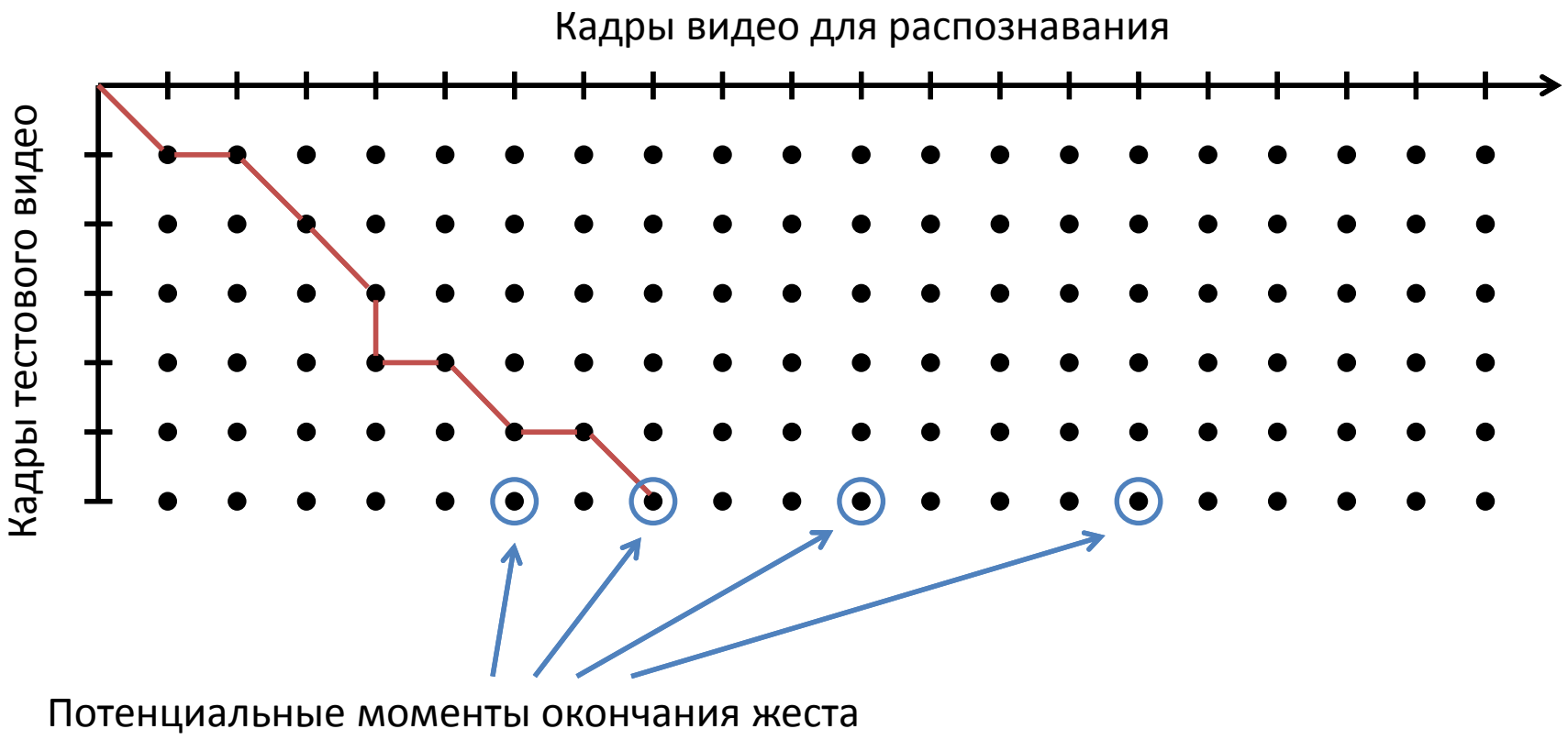
Одновременное распознавание жестов и
определение момента окончания:

$$\hat{i} = \underset{i}{argmin} \underset{j \in endings(G_i)}{min} similarity(V_j, G_i)$$

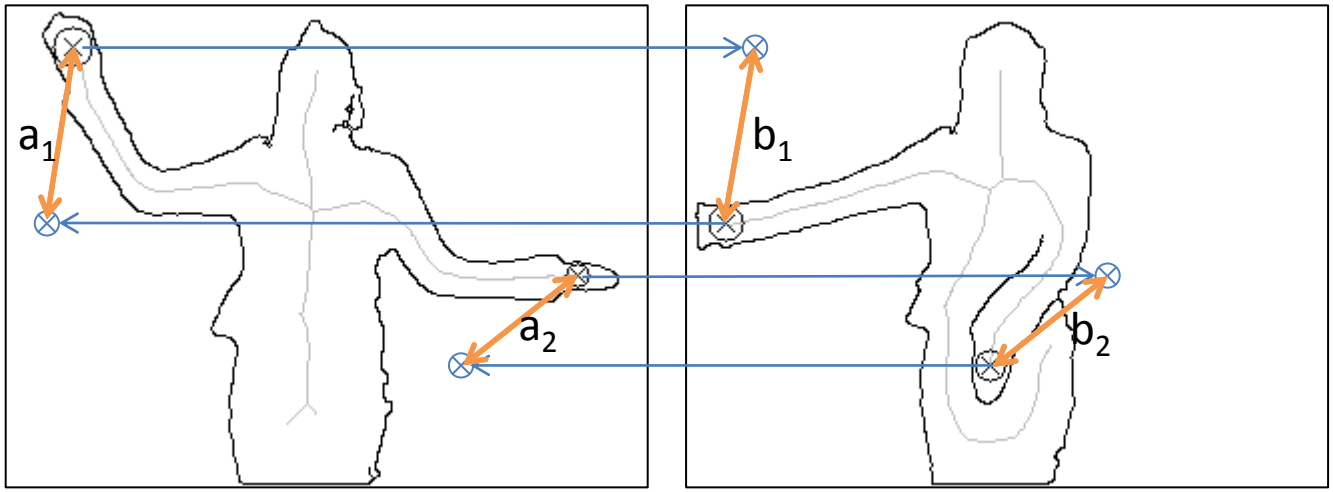
где V_j – первые j кадров видео, endings(G_i) – потенциальные моменты окончания эталонного жеста i в видео.

Метрическое распознавание жестов

Одновременная сегментация и распознавание с помощью дин. программирования



Мера сходства кадров



n – число ключевых точек на первом кадре
m – число ключевых точек на втором кадре

$$dist = \sum_{i=1}^{n} a_i + \sum_{i=1}^{m} b_i + C(n-m)$$

Результаты экспериментов

База жестов ChaLearn Gesture Challenge:

- База разбита на независимые пакеты
- Каждый пакет содержит 10 эталонных жестов и порядка 35-40 контрольных видео.
- Одно контрольное видео содержит от 1 до 5 жестов, сегментация контрольных видео на жесты неизвестна

Критерий качества (≈ доля ошибок классификации):

$$Q = \frac{\sum_{i=1}^{N} Levenshtein(c_i, t_i)}{\sum_{i=1}^{N} |t_i|}$$

где Levenstein(a, b) – расстояние Левенштейна между последовательностями a и b, c_i – результат классификации, t_i – истинные метки классов для видео i

Результаты экспериментов

Эксперименты на базе ChaLearn Gesture Challenge

Пакет	Качество классификации Q	Доля корректно сегментир. видео	Качество сегментации Q_S
devel01	0,067	89% (33 из 37)	0,96
devel02	0,23	83% (30 из 36)	0,93
devel04	0,23	65% (24 из 37)	0,84
devel07	0,15	92% (35 из 38)	0,97
devel01,02,04,07	0,17	82% (122 из 148)	0,92
valid01-20	0,44	-	-

Пакеты devel01,02,04,07 содержали жесты совершаемые за счет перемещения ладоней

Пакеты valid01-20 использовались для ранжирования участников соревнования, у лидеров качество классификации на этих пакетах было порядка 0,15 – 0,23