

Информационная технология моделирования жестовой речи

Суслов Александр Юрьевич

Научный руководитель:

к.т.н., доц. Филиппович Юрий Николаевич

МГТУ им. Н.Э. Баумана
Кафедра ИУБ

8 июня 2010 г.

В ходе исследования использовались материалы курсового проекта
«Лингвистический процессор (видео) автоматизированного перевода жестового языка»
Авторы: Мелик-Сафарянц Д. Ю., Михайлова С.А., Шумилова М.С.,
а также материалы сайта <http://surdoserver.ru/>
Разработчики: коллектив Лаборатории №17 ИПУ РАН (<http://surdo.asmon.ru/>)



Актуальность

Актуальность работы обусловлена:

- ▶ Необходимостью создания новых интерфейсов взаимодействия человек-компьютер;
- ▶ Необходимостью создания банка жестов языка глухонемых;
- ▶ Исследованиями языка жестов и телодвижений.

Возможные области применения технологии моделирования жестов

Задача распознавания:

- ▶ Автоматический сурдоперевод

Задача управления:

- ▶ Замена традиционным устройствам дистанционного управления
- ▶ Компьютерные игры
- ▶ Среда виртуальной реальности

Цель и задачи исследования

Цель работы:

Разработка методов моделирования и распознавания динамических жестов.

Задачи:

- ▶ Анализ современных подходов к моделированию жестов;
- ▶ Разработка модели формального описания динамических жестов;
- ▶ Разработка методов и алгоритмов распознавания динамических жестов;
- ▶ Проектирование реализующего разработанные алгоритмы программного комплекса.

Обзор аналогов и прототипов

Существующие аналоги и прототипы:

- ▶ Мультимедийный толковый словарь русского жестового языка

Воскресенский А.Л., Хахалин Г.К.

АНО «Колледж управления, права и информационных технологий МЭСИ»

- ▶ Сурдосервер (<http://surdoserver.ru/>)

Лаборатория №17 ИПУ РАН

- ▶ Технология CamSpace

- ▶ Технологии Softkinetic's iisu, Microsoft Natal, Trendy Vision Play

Формализация поставленной задачи

Обозначим:

- $Gestures = \{gest\}$;
- $Videos(gest) = \{video\}$,
 $gest \in Gestures$;
- $Videos = \bigcup_{i=0}^n Videos(gest_i)$,
 $gest_i \in Gestures$;
- $Z(video)$, $video \in Videos$;
- $z \in Z(video)$,
 $video \in Video(gest)$,
 $gest \in Gestures$.

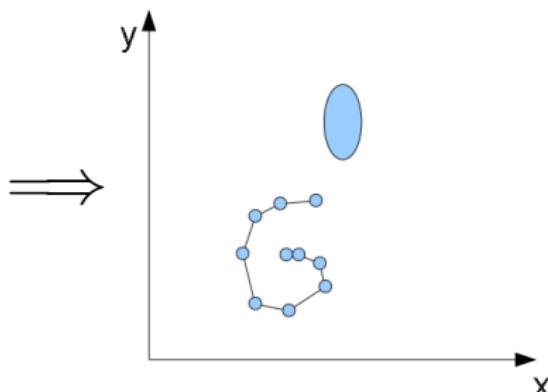
Дано:

- $Gestures_0 \subset Gestures$, $Gestures_0 = \{g_i\}$,
 $i = 0, 1, \dots, n$;
- $Videos_0(gest_i) \subset Videos$,
 $Videos_0(gest_i) = \{video_j\}$,
 $j = 0, 1, \dots, m$;
- $Videos_0 = \bigcup_{i=0}^n Videos_0(gest_i)$;
- $V(W, H) \in Videos$,
 $V(W, H) = (l_0, l_1, \dots, l_K)$;
- $V(W, H) \notin Videos_0$.

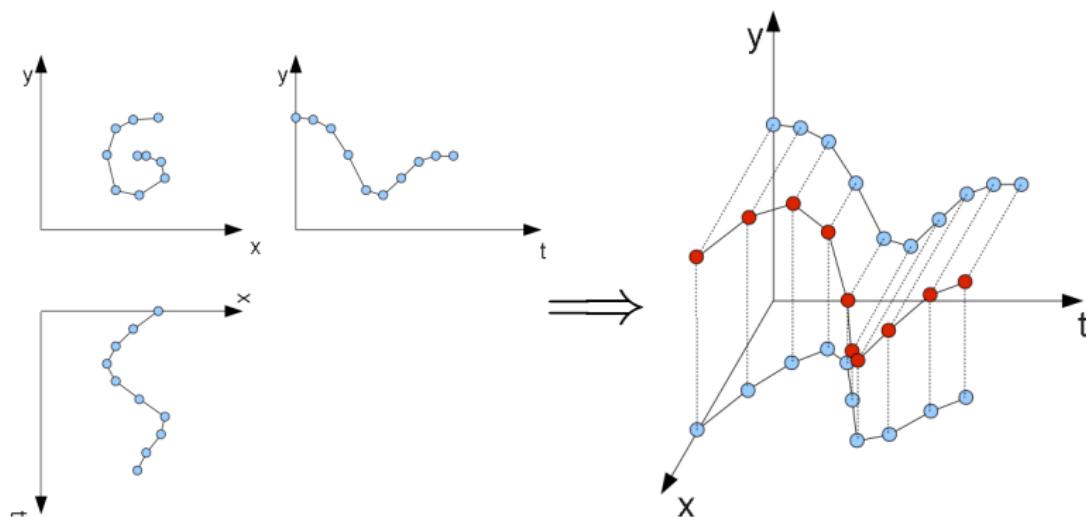
Найти: $Z_0 : Z_0(V) \subseteq Z(V)$, $Z_0(V) = \{z_0\}$.

$$gest_i : \arg \max_{video_i} (Z_0 \cap Z(video_i)) = video \in Videos_0(gest_i).$$

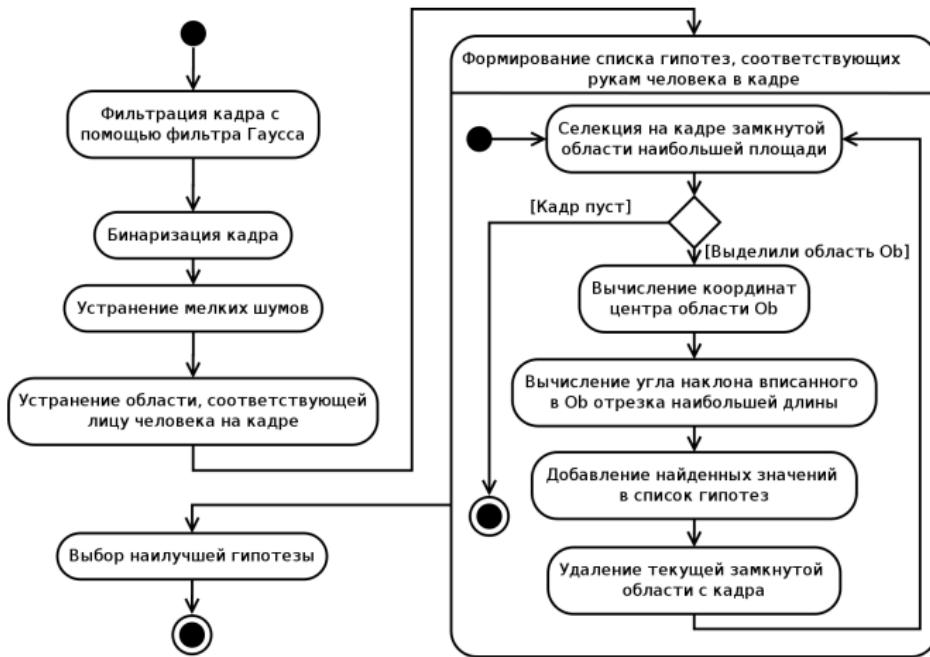
Концепт динамического жеста



Концепт динамического жеста



Алгоритм захвата рук на кадре



Бинаризация кадра

$$R' = \begin{cases} R(C_{x,y}) & , \text{ если } ParR_1 < R(C_{x,y}) < ParR_2, \\ 0 & , \text{ иначе; } \end{cases}$$

$$G' = \begin{cases} G(C_{x,y}) & , \text{ если } ParG_1 < G(C_{x,y}) < ParG_2, \\ 0 & , \text{ иначе; } \end{cases}$$

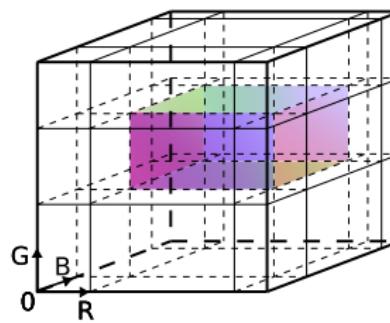
$$B' = \begin{cases} B(C_{x,y}) & , \text{ если } ParB_1 < B(C_{x,y}) < ParB_2, \\ 0 & , \text{ иначе; } \end{cases}$$

Бинаризация кадра

$$R' = \begin{cases} R(C_{x,y}) & , \text{ если } ParR_1 < R(C_{x,y}) < ParR_2, \\ 0 & , \text{ иначе;} \end{cases}$$

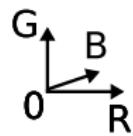
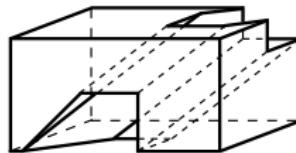
$$G' = \begin{cases} G(C_{x,y}) & , \text{ если } ParG_1 < G(C_{x,y}) < ParG_2, \\ 0 & , \text{ иначе;} \end{cases}$$

$$B' = \begin{cases} B(C_{x,y}) & , \text{ если } ParB_1 < B(C_{x,y}) < ParB_2, \\ 0 & , \text{ иначе;} \end{cases}$$



Бинаризация кадра

$$C'_{x,y} = \begin{cases} 1 & , \text{ если } \max\{R', G', B'\} - \min\{R', G', B'\} > ParBal \\ & \text{и } R(C_{x,y}) > G(C_{x,y}) \text{ и } R(C_{x,y}) > B(C_{x,y}), \\ 0 & , \text{ иначе.} \end{cases}$$

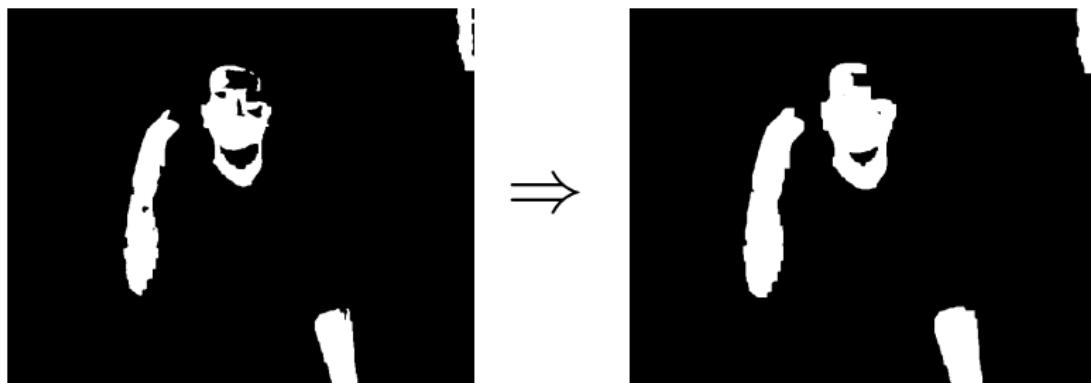


Бинаризация кадра

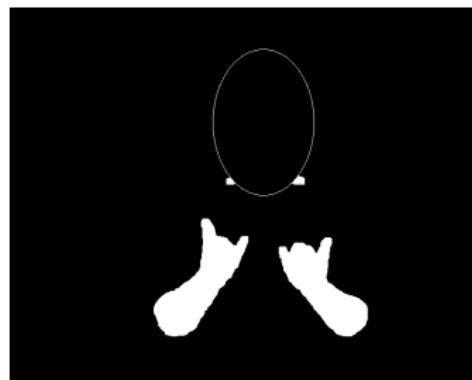


Устранение мелких шумов

$$M_k = \text{dilate}(\text{erode}(M_k, ParE), ParD).$$



Удаление лица с кадра

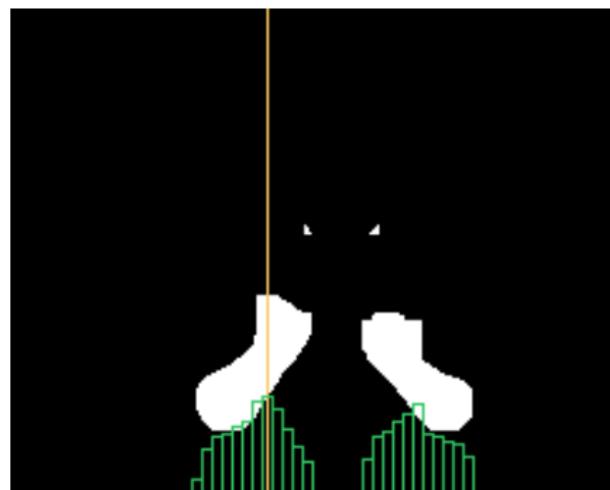


Селекция замкнутой области наибольшей площади

$$h_x(a) = \sum_{i=a \cdot S_{pix}}^{(a+1) \cdot S_{pix}-1} \sum_{j=0}^{H-1} M_k(i, j),$$

$$a \in \left[0, \frac{W}{S_{pix}} - 1\right],$$

$$x_c = \left(\arg \max_a \{h_x(a)\} + \frac{1}{2}\right) \cdot S_{pix},$$

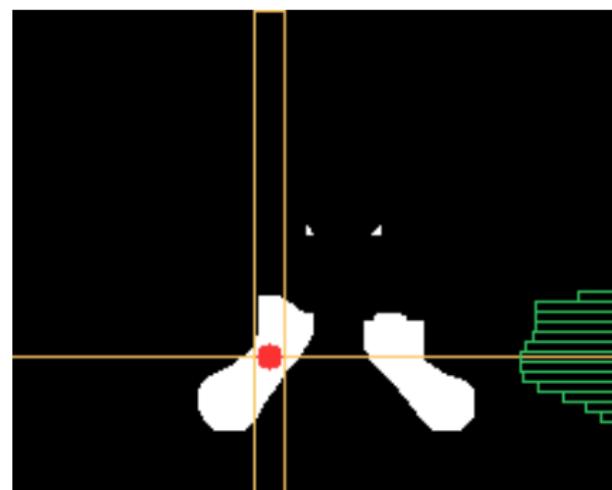


Вычисление координат центра области

$$h_y(b) = \sum_{i=x_c - \frac{S_{pix}}{2}}^{x_c + \frac{S_{pix}}{2} - 1} \sum_{j=b \cdot S_{pix}}^{(b+1) \cdot S_{pix} - 1} M_k(i, j),$$

$$b \in \left[0, \frac{H}{S_{pix}} - 1\right],$$

$$y_c = \left(\arg \max_b \{h_y(b)\} + \frac{1}{2} \right) \cdot S_{pix}.$$



Вычисление угла наклона

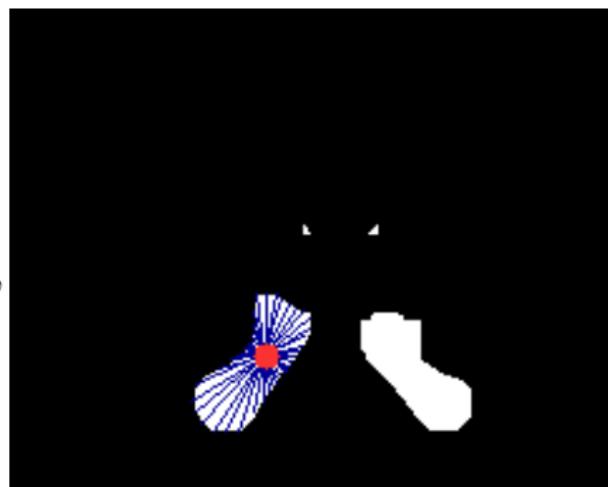
$$Ob \cap l_i = \{(x_{1i}, y_{1i}), (x_{2i}, y_{2i})\},$$

$$\angle(l_i, OX) \in [0, 180^\circ),$$

$$length_i = \sqrt{(x_{1i} - x_{2i})^2 + (y_{1i} - y_{2i})^2},$$

$$angle_i = \angle(l_i, OX).$$

$$n = \arg \max_i (length_i),$$



Удаление текущей замкнутой области с кадра

$$h_y(b) = \sum_{i=x_c - \frac{S_{pix}}{2}}^{x_c + \frac{S_{pix}}{2} - 1} \sum_{j=b \cdot S_{pix}}^{(b+1) \cdot S_{pix} - 1} M_k(i, j),$$

$$b \in \left[0, \frac{H}{S_{pix}} - 1\right],$$

$$y_c = \left(\arg \max_b \{h_y(b)\} + \frac{1}{2} \right) \cdot S_{pix}.$$



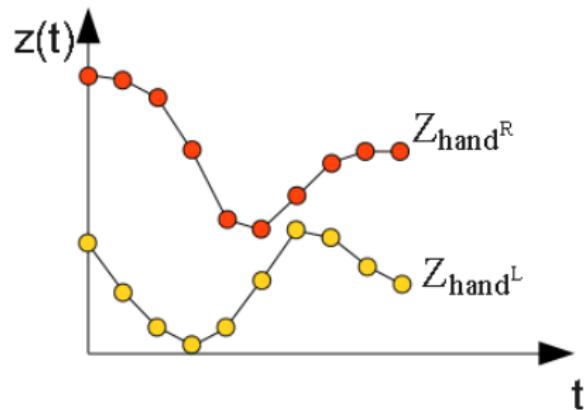
Концепт динамического жеста

$$Z = (z_0, z_1, \dots, z_K),$$

$$z_i = (hand_i^L, hand_i^R),$$

$$i = 0, 1, \dots, K,$$

$$hand = (x, y, length, angle).$$

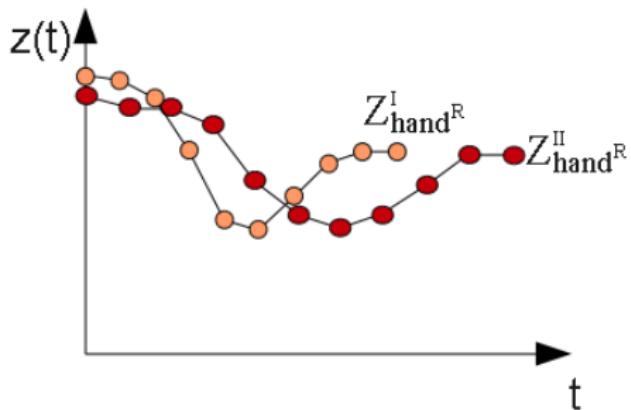


Сравниваемые концепты жестов

$$Z^I = (z_0^I, z_1^I, \dots, z_{K_1}^I),$$

$$Z^{II} = (z_0^{II}, z_1^{II}, \dots, z_{K_2}^{II}),$$

$$K_1 \neq K_2.$$

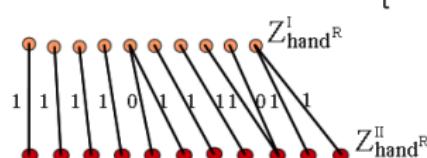
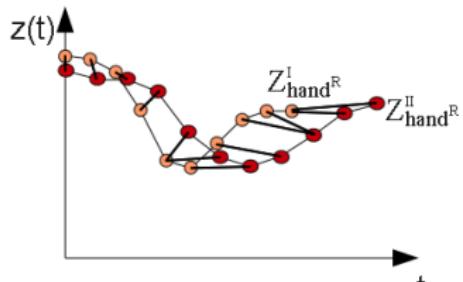


Мера близости двух концептов

$$\rho(hand_i^I, hand_j^{II}) = \begin{cases} 1, & \begin{aligned} |x_i^I - x_j^{II}| &\leq \varepsilon_x \\ \text{и } |y_i^I - y_j^{II}| &\leq \varepsilon_y \\ \text{и } (|\gamma_i^I - \gamma_j^{II}| &\leq \varepsilon_\gamma \\ \text{или } |\gamma_i^I - \gamma_j^{II}| &\geq 360^\circ - \varepsilon_\gamma) \\ \text{и } |len_i^I - len_j^{II}| &\leq \varepsilon_{len} \\ \text{и } |i \cdot FPS^I - j \cdot FPS^{II}| &\leq \varepsilon_t \end{aligned} \\ 0, & \text{иначе.} \end{cases}$$

$$v(Z', Z'') = \max_S \mu(S)$$

$$\begin{aligned} \mu(S) = & \sum_{(i,j) \in S} \rho(\textit{hand}_i^{IR}, \textit{hand}_j^{IR}) + \\ & + \sum_{(i,j) \in S} \rho(\textit{hand}_i^{IL}, \textit{hand}_j^{IL}) \end{aligned}$$



Программная реализация

Особенности реализации

Язык реализации:

C++

Подсистема графического интерфейса:

библиотека Qt

Система управления базой данных:

SQLite

Подсистема обработки видео:

библиотека OpenCV

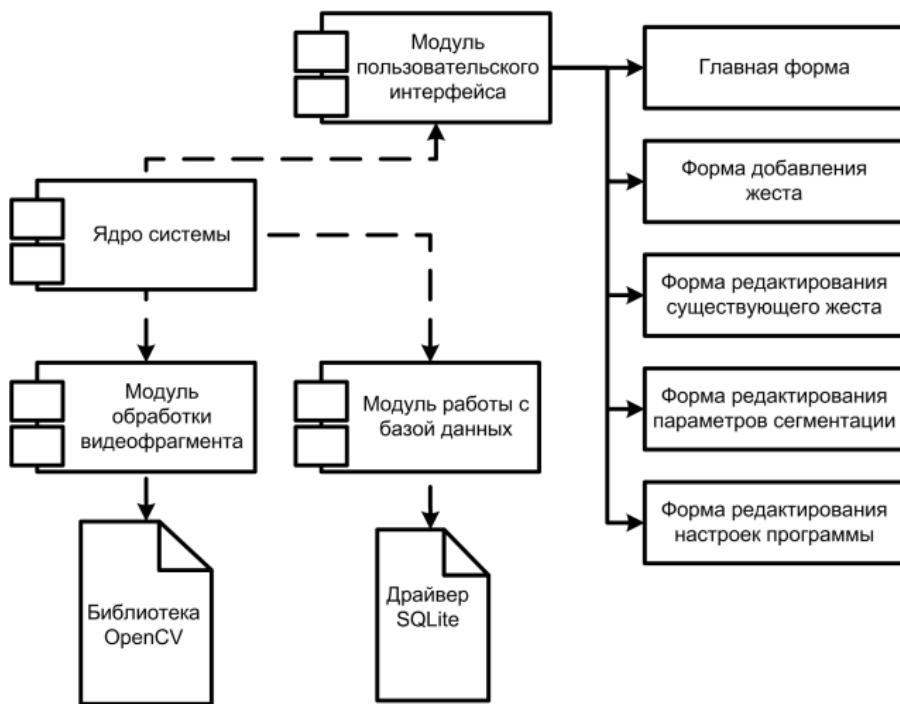
Поддерживаемые ОС:

Windows, Linux,
Mac OS X

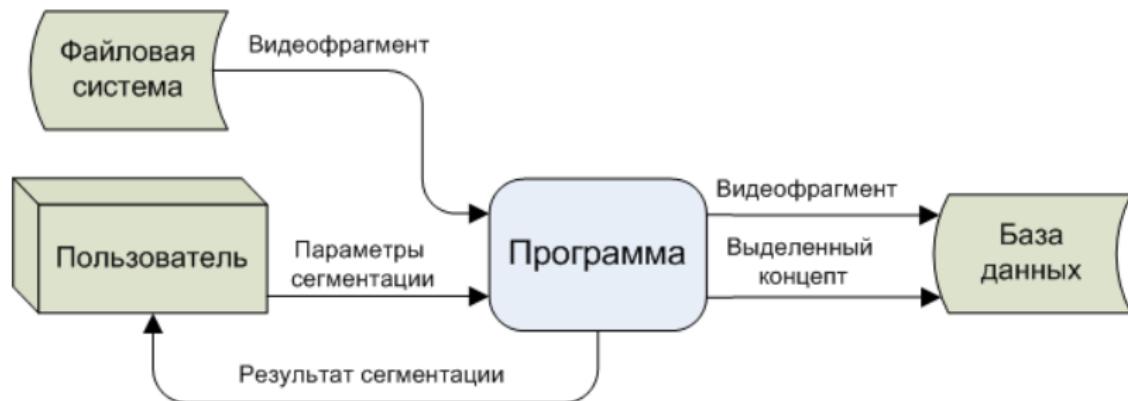
Средства разработки

Интегрированная среда разработки: Qt Creator

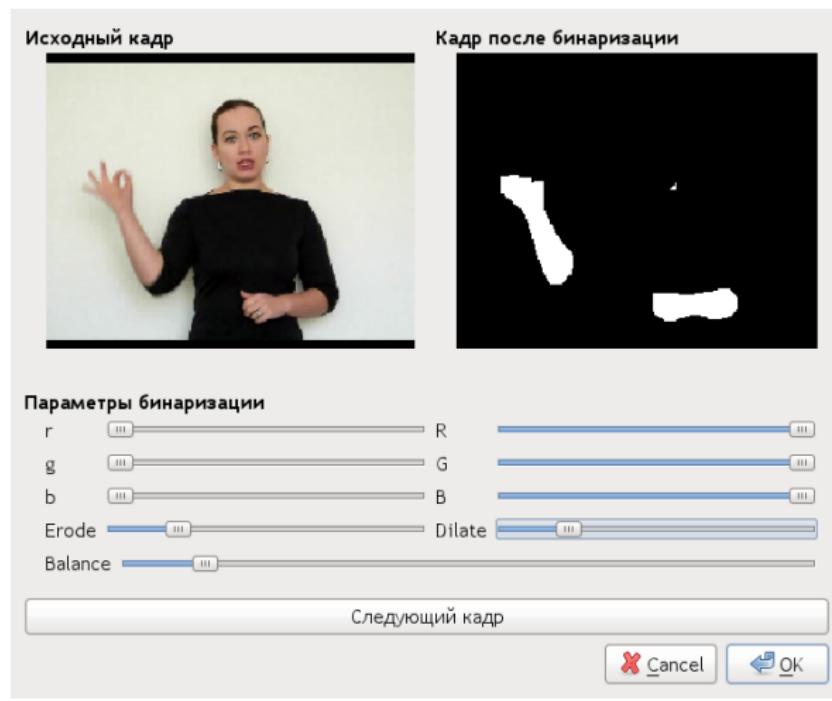
Основные компоненты системы



Функциональная схема



Окно настройки параметров бинаризации



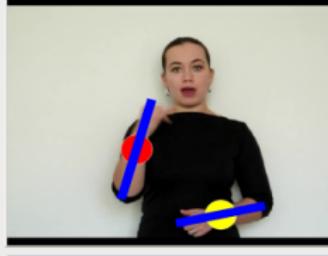
Главное окно программы

Файл

Жесты

- ▷ DNS
- ▷ Схема
- ▷ Явление
- ▷ Системный администратор
- ▷ FTP
- ▷ IP-адрес
- ▷ Байт
- Video_1**
- ▷ Бит
- ▷ Бесплатное ПО
- ▷ Архив
- ▷ Алгоритм сжатия без потерь
- ▷ Алгоритм маршрутизации
- ▷ Deface
- ▷ Коллизия
- ▷ Маршрутизатор
- ▷ Настройка
- ▷ Оперативная память
- ▷ Отладчик
- ▷ Перезагрузка

Видеофрагмент



Добавить...
Редактировать...
Распознать...
Выход

Описание

Байт (англ. byte) – единица хранения и обработки цифровой информации. Чаще всего байт считается равным восьми битам, в этом случае он может принимать одно из 256 (2^8) различных значений. Для того, чтобы подчеркнуть, что имеется в виду восьмивитный байт, в описании сетевых протоколов используется термин «октет» (лат. octet).

Резюме

Основные выводы и результаты работы:

1. Проведён анализ основных методов и систем моделирования и распознавания жестов.
2. Разработана модель формального описания динамических жестов.
3. Разработан метод сегментации видеоизображения.
4. Разработан алгоритм выделения концепта из видеофрагмента.
5. Разработан алгоритм распознавания динамических жестов.
6. Создан программный комплекс, реализующий разработанные методы и алгоритмы.