

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Математико-механический факультет

Кафедра системного программирования

Калитеевский Василий Николаевич

ИСПОЛЬЗОВАНИЕ СКРЫТЫХ МАРКОВСКИХ МОДЕЛЕЙ ДЛЯ РАСПОЗНАВАНИЯ

ЖЕСТОВ ГОЛОВЫ

Курсовая работа

студента 371 группы

Научный руководитель:

профессор Граничин Олег Николаевич

Санкт-Петербург

2014

ОГЛАВЛЕНИЕ

Введение.....	3
1. Подготовка признаков.....	5
2. Скрытые марковские модели	8
2.1 Обучение СММ.....	9
2.2 Этап распознавания.....	10
3. Реализация.....	12
4. Заключение.....	13
5. Список используемой литературы.....	14

Введение

Задача распознавания жестов хорошо рассматривается на примере создания интерфейса для взаимодействия с компьютером.

Общепринятые способы взаимодействия с компьютером почти не изменились со времён появления первых персональных компьютеров. Клавиатура и мышь, сразу став основными устройствами ввода, практически не получили развития. Управление голосом широкого применения не нашло. Большой шаг вперед сделали производители смартфонов, внедрив touch-интерфейс, оказавшийся естественным и интуитивно понятным, однако, для персональных компьютеров он не подходит, поэтому в настоящее время компьютерная индустрия проявляет большой интерес к интерфейсам, основанным на машинном зрении. Действительно, такие интерфейсы содержат значительный потенциал, для реализации и внедрения которого должны быть выполнены следующие требования:

- **Безошибочность.** Не должно возникать неоднозначностей. Нужно использовать только те жесты, которым можно гарантировать точное распознавание;
- **Естественность жестов.** Все жесты должны быть естественны и интуитивно понятны;
- **Стандартизация.** Должна быть единая служба, встроенная в операционную систему, к которой могли бы обращаться другие приложения.

Интерфейс взаимодействия, подчиненный этим правилам может применяться также для смартфонов или, полюбившихся пользователю, планшетов и может иметь различные области применения. С помощью такого интерфейса можно управлять браузером или же беспилотным летательным аппаратом смотря в камеру телефона и подавая команды.

В рамках проекта разработки интерфейса взаимодействия с персональным компьютером при помощи веб-камеры в качестве расширения для браузера Internet Explorer была поставлена задача о составлении карты движений и последующем их распознавании. Реализованный алгоритм, схожий с EM-методом не давал хорошего процента детектирования, поэтому требовалось найти и применить другой алгоритм распознавания.

1. Постановка задачи

Была определена следующая цель: провести обзор существующих решений и взяв за основу подходящий алгоритм добиться хорошего процента детектирования, согласно карте жестов.

Карта жестов:

- Кивок
- Наклон головы влево + возврат
- Наклон головы вправо + возврат
- Поворот головы влево + возврат
- Поворот головы вправо + возврат
- Приближение
- Отдаление

Необходимо создать программу, способную:

- Принять видео-поток с веб-камеры
- Произвести необходимую обработку кадров
- Выделить признаки с помощью библиотеки intel pcsdk
- На основе выделенных признаков детектировать команды с помощью скрытых моделей Маркова

2. Подготовка признаков

Движения головы продолжаются определенный интервал времени и являются непрерывным процессом, однако мы будем рассматривать дискретную модель и представлять жесты в виде последовательности определенных положений головы в пространстве, задаваемых признаками.

Основные изменения в положении головы во время жестов это приближение(удаление), наклон и поворот , эти соображения задают минимальное количество признаков, которое нужно использовать. Чтобы отслеживать изменение расстояния до лица нужно задать метрику расстояния на лице, отслеживать угол для наклона и высчитывать изменение горизонтальной координаты как характеристику поворота.

С помощью библиотеки Intel Perceptual Computing SDK будем получать координаты положения частей лица, среди которых нос, глаза и рот.

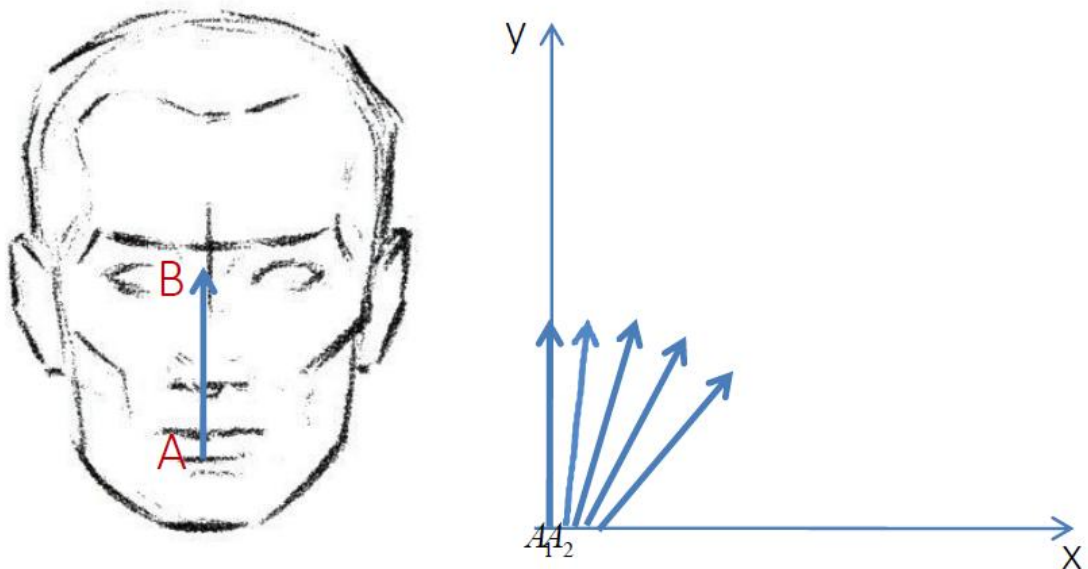
Пример извлеченных опорных точек лица с помощью Intel pcsdk:



Итак, процесс распознавания жестов начинается с обработки каждого конкретного кадра:

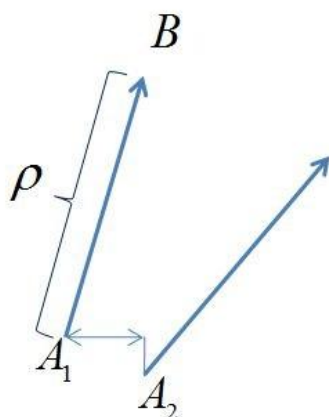
- Детектирование лица
- Определение координат опорных точек лица
- Извлечение признаков

Проекция наклона головы на оси двухмерных координат:



Введем следующие обозначения: точка А - координата центра рта, точка В - координата середины отрезка, соединяющего глаза, тогда угол α есть отклонение АВ по вертикали, ρ фиксирует изменение расстояния между точками А и В, а x горизонтальное изменение положения точки А. Таким образом, параметры $\langle \alpha, \rho, x \rangle$ и возьмем в качестве базовой тройки признаков.

Выделенные признаки на лице:

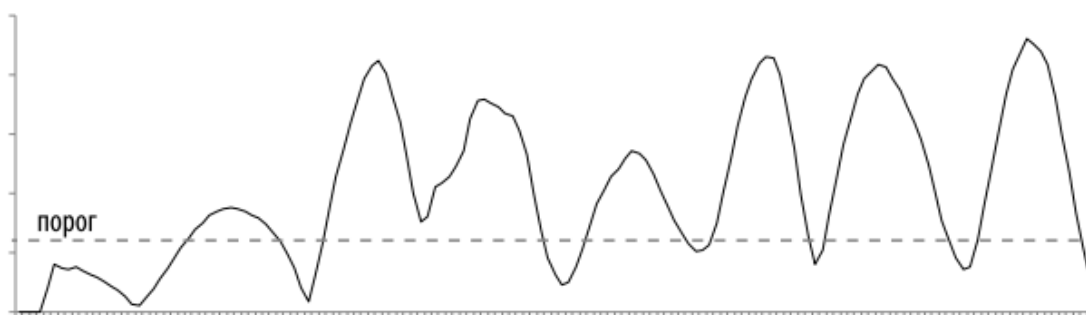


Обозначим координаты точек: $A_1(x_1, y_1)$, $A_2(x_2, y_2)$, $B(x_3, y_3)$. Тогда признаки $\langle \alpha, \rho, x \rangle$ можно вычислить по следующим формулам:

$$\rho = \sqrt{(x_3 - x_1)^2 + (y_3 - y_1)^2}$$

$$\alpha = \text{sign}(y_3 - y_1) \arccos\left(\frac{x_3 - x_1}{\rho}\right), \quad x = x_2 - x_1$$

Для детектирования начала жеста высчитывается суммарное изменение каждого из трех признаков за последние десять кадров. Хранить признаки было решено в относительных, а не абсолютных величинах. В случае преувеличения одного из признаков порогового значения в 8% начинается покадровая запись жеста и продолжается до возвращения параметров в исходное положение. Затем берется среднее изменение каждого из параметров на участках жеста, равных количеству состояний скрытой марковской модели и в качестве извлеченных признаков подаются на вход алгоритму обучения.



2. Скрытые марковские модели

Теоретически любая последовательность может быть сгенерирована некоторым параметризованным случайным процессом. Моделирование такого процесса может быть организовано с помощью скрытой Марковской модели, параметры которой могут быть настроены на основе обучающих последовательностей [1].

Скрытые марковские модели описывают стохастические процессы, состоящие из множества состояний, каждое из которых связано с другим стохастическим процессом [2]. Формально могут быть описаны следующим образом [3]:

1. N - общее количество состояний в модели. Переход в любое выбранное состояние возможен из любого состояния всей системы (в том числе и само в себя). Алфавит ненаблюдаемой последовательности мы обозначим как $\Omega = \{\omega_1, \dots, \omega_N\}$.

2. M - количество возможных символов в наблюдаемой последовательности, размер алфавита наблюдаемой последовательности. Алфавит наблюдаемой последовательности мы обозначим как $O = \{o_1, \dots, o_M\}$.

3. Вероятностное распределение смены состояний $A = \{\alpha_{ij}\}$, где α_{ij} представляет вероятность смены состояния из ω_i во время t в состояние ω_j во время $t+1$.

$$\alpha_{ij} = P(\omega_j(t+1) | \omega_i(t)), \quad 1 \leq i, j \leq N.$$

4. Вероятностное распределение выбора части распознаваемого жеста $B = \{\beta_{il}\}$, где β_{il} представляет вероятность выбора части жеста o_l во время t в состоянии ω_i .

$$\beta_{il} = P(o_l(t) | \omega_i(t)), \quad \sum_{l=1}^M \beta_{il} = 1 \quad \text{для } \forall i.$$

5. Первоначальное распределение состояний $\pi = \{\pi_i\}$. $\pi_i = P(\omega_i(1))$, $1 \leq i \leq N$.

Полный набор параметров СММ будем обозначать как $\lambda = \{A, B, \pi\}$.

2.1 Обучение СММ

Движения головы продолжаются определенный интервал времени и являются непрерывным процессом, однако мы будем рассматривать дискретную модель и представлять жесты в виде последовательности определенных положений головы в пространстве.

Обучение скрытой марковской модели является одной из основных задач СММ и формально формулируется следующим образом:

Дано: наблюдаемая последовательность $O = \{o_1, \dots, o_M\}$. Подобрать параметры модели $\lambda = \{A, B, \pi\}$ таким образом, чтобы максимизировать вероятность $P(o | \lambda)$.

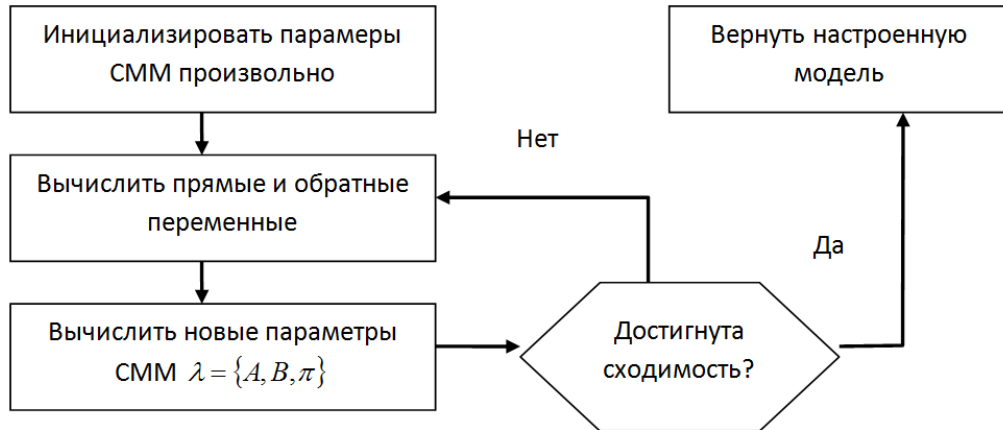
Решается данная задача с помощью алгоритма Баума-Уелша [4]. Этот метод рассматривает значение вероятности перехода из состояния ω_i в ω_j как порядок между ожидаемым количеством переходов из состояния ω_i в ω_j и ожидаемым общим количеством переходов из ω_i . Таким же способом оценка вероятности выбора части жеста o_l в состоянии ω_i получается за счет вычисления порядка между ожидаемым количеством раз, когда будет выбрана часть жеста o_l в состоянии ω_i и ожидаемым общим количеством раз в состоянии ω_i . Оценка начальной вероятности в состоянии ω_i есть ожидаемая частота в ω_i во время $t = 1$. Эти частоты могут быть посчитаны через вероятности перехода из $\omega_i(t)$ в $\omega_i(t+1)$ как

$$v_{ij}(t) = P(\omega_i(t), \omega_j(t+1) | o, \lambda) = \frac{\delta_i(t) \alpha_{ij} \beta_{j o(t+1)} \eta_j(t+1)}{P(o | \lambda)}.$$

Используя $v_{ij}(t)$, становится возможным определить значения параметров СММ:

$$\hat{\pi}_i = \sum_{j=1}^N v_{ij}(1), \quad \hat{\alpha}_{ij} = \frac{\sum_{t=1}^{T-1} v_{ij}(t)}{\sum_{t=1}^{T-1} \sum_{j=1}^N v_{ij}(t)}, \quad \hat{\beta}_{il} = \frac{\sum_{t=1, o(t)=o_l}^T \sum_{j=1}^N v_{ij}(t)}{\sum_{t=1}^T \sum_{j=1}^N v_{ij}(t)}.$$

Схема работы алгоритма Баума-Уеллиа:



2.2 Этап распознавания

Задача распознавания скрытой марковской моделью имеет следующую формальную постановку:

Дано: наблюдаемая последовательность $O = \{o_1, \dots, o_M\}$ и модель $\lambda = \{A, B, \pi\}$. Необходимо подобрать последовательность состояний системы $\Omega = \{\omega_1, \dots, \omega_N\}$, которая лучше всего соответствует наблюдаемой последовательности.

Решается путем применения алгоритма Витерби [5], суть которого заключается в том, чтобы сохранять путь до текущего наилучшего состояния ω_j во время $t-1$, который имеет наибольшую вероятность получения наблюдаемой последовательности $\{o(1), o(2), \dots, o(t)\}$.

Эта вероятность может быть вычислена рекурсивно $\zeta_i(t) = \left(\max_j \zeta_j(t-1) \alpha_{ji} \right) \beta_{io}(t)$. Соответствующее наилучшее состояние сохраняется в переменной $\zeta_i(t)$.

Алгоритм Витерби:

- Инициализация. $t=1$, $\zeta_i(1) = \pi_i \beta_{io(1)}$, $1 \leq i \leq N$, $\zeta_i(1) = 0$, $1 \leq i \leq N$.
- Рекурсия. ($t=2, \dots, T$). $\zeta_i(t) = \left(\max_j \zeta_j(t-1) \alpha_{ji} \right) \beta_{io(t)}$, $1 \leq i \leq N$,

$$\zeta_i(t) = \arg \max_j \zeta_j(t-1) \alpha_{ji}, \quad 1 \leq i \leq N.$$

- Остановка. $\zeta^* = \max_j \zeta_j(T)$, $\omega^*(T) = \arg \max_j \zeta_j(T)$.
- Обратная связь. ($t=T-1, T-2, \dots, 1$). $\omega^*(t) = \zeta_{\omega^*(t+1)}(t+1)$.

3. Реализация

Реализация данного инструмента осуществлялась с помощью языка программирования С# и библиотеки Intel Perceptual SDK (для извлечения опорных точек лица).

В начале работы программы создается база обучающих последовательностей, после чего пользователю предлагается либо создать свою, либо использовать существующую.

После чего начинается распознавание новых жестов по следующему алгоритму:

- Создается скрытая марковская модель;
- Создается классификатор скрытой марковской модели;
- Классификатор обучается на основе базы жестов по алгоритму Баума-Уелша, используя нормальное распределение;
- Полученный жест распознается по алгоритму Витерби.

4. Заключение

До перехода к скрытым марковским моделям был реализован алгоритм, основанный на идее EM-метода, измеряющий отклонение наблюдаемого движения от некоторой эталонной траектории для каждого жеста. В таблице ниже приводится сравнение вероятностей правильного детектирования жестов двумя алгоритмами.

	Алгоритм, основанный на идее EM-метода	СММ
Кивок	60%	80%
Наклон влево	70%	90%
Наклон вправо	70%	90%
Поворот влево	70%	90%
Поворот вправо	70%	90%
Приближение	95%	95%
Удаление	95%	95%

При использовании скрытых марковских моделей наблюдается существенный подъем вероятности правильного детектирования движений.

Цель следующих исследований - сравнить описанные в курсовой работе алгоритмы с одной из вариаций метода рандомизированной стохастической аппроксимации.

5. Список используемой литературы

1. M. Jones and P. Viola, "Fast multi-view face detection," Mitsubishi Electric Research Laboratories, Tech. Rep. 096, 2003
2. Yang J., Xu Y. Hidden Markov Model for Gesture Recognition: Technical Report CMU.
3. Местецкий Л. М., «Математические методы распознавания образов», МГУ, ВМиК, Москва, 2002–2004.
4. Rabiner L. A tutorial on hidden Markov models and selected applications in speech recognition . Proceedings of the IEEE , 77 : 257 – 286, 1989
5. Viterbi A. Error bounds for convolutional codes and an asymptotically optimal decoding algorithm . IEEE Transactions on Information Theory, IT - 13 : 260 – 269, 1967.