Санкт-Петербургский Государственный Университет

Математико-механический факультет

Кафедра системного программирования

Распознавание языка жестов на видео потоке

Курсовая работа студента 445 группы

Землянской Светланы Андреевны

Научный руководитель: ..... д-р физ.-мат. наук О. Н. Граничин

Санкт-Петербург

2012

Содержание

Глава 1

Введение

1.1 Актуальность

В последнее время все больше внимания уделяется автоматическому распознаванию жестов с помощью визуальных систем. Такой интерес вызван природным характером и удобством использования интерфейса на основе жестов, а также возможностью его применения в большинстве областей человеческой деятельности. Это может быть как способ ввода информации в компьютер, в том случае, если другие способы неудобны или просто недоступны. Простым примером является ввод данных в мобильное устройство: ввод текста сильно замедляется из-за небольшого размера клавиш, но при введении системы распознавания жестов становится возможным ввод данных с помощью виртуальной клавиатуры. Другой важной областью использования систем распознавания жестов является разработка искусственного интеллекта в робототехнике. Умение робота распознавать не только голосовые команды, но и отслеживать жесты человека, существенно упрощает его обучение и, как следствие, ускоряет его адаптацию.

Отдельной задачей в распознавании жестов стоит задача распознавания жестов языка глухонемых. По статистике, нарушениями слуха страдает каждый девятый человек, в России их количество составляет примерно 12 млн. И приложение, способное распознавать язык жестов, сильно упростило бы коммуникации для глухонемых людей.

1.2 Цель работы

В качестве распознаваемого набора жестов было решено использовать не весь набор жестов из языка глухонемых, а только лишь латинский алфавит. Обусловлено это несколь-

кими причинами:

- Выбор алфавита, как результирующего набора жестов, позволяет проводить распознавания с использованием только одной камеры.
- Появляется возможность проверить состоятельность данного способа классификации жестов.

В связи с этим, конечной целью данной работы является написание приложения по распознаванию латинского алфавита. Данное приложение на вход должно получать видео поток, снятый с одной вебкамеры, и на выход выдавать полученный текст.

Требования на конечное приложение:

- В качестве инструмента предполагается использовать только одну камеру
- Работа приложение должна в режиме реального времени
- Не должна требоваться отдельная "настройка"приложения под человека
- Не должна делаться ставка на цвет кожи

1.3 Метод исследования

В исследовании предполагается задействовать следующие направления:

- *Компьютерное зрение*

  Использование алгоритмов компьютерного зрения для распознавания образов на изображении.

- *Машинное обучение*

  Необходимо разбиение изображения на кластеры для отслеживания каждой ладони в отдельности. И построения классификатора для определения является ли данный кластер ладонью или нет.

- *Стохастическая оптимизация*

  Предполагается, что алгоритмы будут работать недостаточно быстро для обработки изображения в режиме реального времени. Планируется использовать методы стохастической оптимизации для их ускорения.

1.4 Существующие исследования

Существует достаточно много разработок, связанных с распознаванием жестов и управлением компьютером с их помощью. Приведу наиболее успешные из них.

1 **MIT**

Одной из первых крупных разработок по этой тематике является разработка в Масачусецком Институте. Для построения модели руки на основании изображения было предложено использовать маркеры - разноцветные перчатки. Такой подход во многом упростил задачу и сделал возможным её решение практически в режиме реального времени, всего лишь с небольшой задержкой. Однако, использование дополнительных атрибутов для распознавания, является принципиальным недостатком такого метода, т. к. усложняет его применение.

2 **Microsoft**

Другой способ решения был предложен компанией Microsoft, где делалась ставка на использование не только двумерной картинки, но и показателя глубины изображения. Для её измерения был сконструирован аппарат Kinect, излучающий инфрокрасное излучение. С его использованием снижаются требования к освещенности помещения, но появляются требования к влажности и температуре в комнате. И, конечно, важным его недостатком является стоимость.

3 **Терком**

Из отечественных исследований, стоит отметить исследование компании Терком. В этом решении настраиваются не одна, а две камеры и по ним строится трехмерное изображение. Плюсом такого подхода, несомненно, является отказ от использования дополнительных атрибутов, что принципиально снижает его стоимость.

В случае с целенаправленным распознаванием жестов языка глухонемых, дела обстоят несколько хуже.

1 **Университеты Осаки и Шиншу**

Японские исследователи разработали систему, способную преобразовывать жесты языка глухонемых в символы. Несмотря на хорошие результаты по распознаванию, главным недостатком данной системы является использование специальной перчатки с постоянными магнитами, закреплёнными на кончике каждого пальца.

2 Чешско-русская разработка

Несколько другие результаты дала совместная работа Университета Западной Богемии и русской огранизации "Академия фантазии". В данной разработке, при распознавании жестов, учитывается не только движение рук, но и изменение мимики человека. Для достижения такого результата вводятся дополнительные сенсоры, существенно усложняющие его применение на практике.

Глава 2

Теоретическое обоснование

2.1 Классификатор

2.1.1 Постановка задачи

Для классификации объекта выделяется набор признаков, который считается достаточным для идентификации класса объекта, набор признаков представляется в виде n-мерного вектора. Классификатор может давать положительный, в случае принадлежности объекта к главному множеству, или отрицательный результат, в противном случае. Задача классификации является задачей с учителем, поэтому на вход даётся тренировочное множество из m векторов: $x^{(1)} = [x_1^{(1)}, x_2^{(1)}, ..., x_n^{(1)}], x^{(2)}, ..., x^{(m)}$. Каждому вектору ставится в соответствие ожидаемый результат классификации: $y^{(1)}, y^{(2)}, ..., y^{(m)}$.

2.1.2 Логистическая регрессия

Метод логистической регрессии основан на представлении классификатора как параметрически заданной функции:

$h_\Theta(x) = \frac{1}{1+e^{-\Theta^T x}}$, где $\Theta = [\Theta_1, \Theta_2, ..., \Theta_n]$

С помощью тренировочного множества подбирается набор параметров наиболее оптимальный для данной классификации. Определяется функция ошибки и с помощью метода градиентного спуска находится её минимум:

$J(\Theta) = \frac{1}{2m} \sum_{i=1}^{n} (h_\Theta(x^{(i)}) - y^{(i)})^2 \to min$, где $y^i \in \{0, 1\}$

2.1.3 Support Vector Machines

Классификация в данном методе происходит с помощью разделения точек различных классов гиперплоскостью. Таких гиперплоскостей может быть много, поэтому в качестве меры качества выбранной является зазор между классами. Если существует гиперплоскость, разделяющая классы с максимальным зазором, то она называется оптимальной разделяющей гиперплоскостью, а соответствующий ей линейный классификатор называется оптимально разделяющим классификатором.

Строим разделяющую гиперплоскость, которая имеет вид (где w - перпендикуляр к разделяющей гиперплоскости):

$$w * x - b = 0$$

Отдельно введём понятие ошибка классификации для данного элемента - ξ_i. Результат классификации, в таком случае, может принимать два значения: $y_i \in \{-1, 1\}$. Согласно данной классификации каждый вектор x должен удовлетворять такому условию:

$$y_i * (w * x_i - b) \geq 1 - \xi_i$$

Для нахождения максимально возможного зазора между классами при таком линейном разделении необходимо минимизировать следующую функцию:

$$\frac{1}{2} * ||w||^2 + C * \sum_{i=1}^{n} \xi_i \to min$$

2.1.4 Нелинейная SVM

Идея классификации с помощью нелинейной SVM практически полностью повторяет идеи линейной, с той лишь разницей, что каждое скалярное произведение заменяется нелинейной функцией ядра.

В данной работе в использовалось ядро Гаусса:

$k(x_1, x_2) = exp(-\frac{||x_1 - x_2||^2}{2*\sigma^2})$

2.2 Мультиклассификатор

Основой мультиклассификации был выбран алгоритм OneVsAll.

1 Необходимо произвести классификацию с m различными исходами.

2 Для каждого $i \in \{1 : m\}$ происходит классификация, где положительным исходом считается принадлежность данного объекта к i-ому классу и отрицательной в противном случае.

3 В результате получается m-мерный вектор вероятностей, принадлежности к каждому классу, на основе которого и происходит классификация.

Глава 3

Реализация

3.1 Алгоритм

Входная информация - видео поток с одной вебкамеры. Изображения получаются по одной в секунду, данная скорость была выбрана по причине того, что она позволяет существенно увеличить время на обработку изображения без потери важной информации.

1 *Предварительная обработка изображения*

На данном этапе изображение уменьшается и с помощью алгоритма Канни, параметры для которого были выбраны эмпирическим путём, на нём выделяются границы.

2 *Построение предположений*

К бинарному изображению применяется метод Хаффа, нахождения дуг окружностей, исходя из результатов которого строится ряд предположений. Данное место является наиболее узким в разработке, т.к. требует полного обхода изображения. На выход получается список из областей, где согласно предположению, может находиться палец.

3 *Классификация пальцев*

Было сделано предположение, что в большинстве жестах, пальцы, либо находятся в непосредственной близости от каркаса ладони, т.е. многоугольника, описанного вокруг неё, либо близко друг от друга. На основе этого было выделено два признака классификации:

- расстояние до ближайшего потенциального пальца
- расстояние до каркаса ладони

В качестве метода классификации, для нахождения наилучшего результата, была испробована логистическая регрессия и svm с линейным и нелинейным ядром.

4 *Классификация жестов*

На вход поступает список пальцев, уже прошедших первый этап классификации. Нормированные координаты данных пальцев и являются вектором признаков, по которому происходит классификация.

В качестве классификаторов были испробованы три, описанных выше способа.

Глава 4

Результат работы

Для оценки классификации использовались стандартные метрики:

$precission = \frac{tp}{tp+fp}$

$recall = \frac{tp}{tp+fn}$

tp(true positive) - количество объектов на которые был дан положительный ответ в том случае, когда он и требовался; fp(false positive) - положительный ответ, когда был верен отрицательный; fn(false negative) - отрицательный ответ, когда требовался положительный.

Finger Gestes	Logistic regression	Linear SVM	Gaussian SVM
Logistic regression	0.35714	0.26667	0.47619
Linear SVM	-	-	-
Gaussian SVM	0.43478	0.071429	0.30435

Таблица 4.1: Precission классификатора

Finger Gestes	Logistic regression	Linear SVM	Gaussian SVM
Logistic regression	0.27778	0.22222	0.47619
Linear SVM	-	-	-
Gaussian SVM	0.47619	0.076923	0.46667

Таблица 4.2: Recall классификатора

Глава 5

Заключение

Было написано приложение по распознаванию латинской азбуки жестов. На вход приложение получает видеопоток с одной вебкабмеры, на выход текст.

5.1 Требования к приложению

Требования	Исполнение
Одна веб-камера	Приложение на вход получает данные с одной камеры (+)
Работа в режиме реального времени	Устроить работу в режиме реально времени не получилось, требуются доработки (-)
"Настройка"под человека	Настройка под человека не проводилась (+)
Зависимость от цвета кожи	В алгоритме работы приложения цвет кожи не учитывался (+)

Таблица 5.1: Выполнение требований к задаче

5.2 Дальнейшее развитие

Задачи, по продолжению разработок в данной области, можно условно разделить на несколько областей:

1 *Увеличение скорости работы приложения*

Требование относительно работы приложения в режиме реального времени выполнить не удалось, распознавание происходит с задержкой и, как следствие, возможно

только на записи видео. Данная проблема может быть решена аппаратным образом. Наиболее затратным является этап, на котором строятся предположения относительно расположения пальцев, т.к. для этого необходимо обойти каждый пиксель изображения. Распараллеливание данного процесса должно привести к существенному выигрышу в скорости.

2 *Расширение числа распознаваемых жестов*

На данном этапе приложение проводит распознавание алфавита жестов. Данный выбор был обусловлен тем, что для распознавания алфавита достаточно изображения одной руки. Полноценный язык жестов более сложен и в нём задействованы обе руки человека, голова, туловище, также значение сказанному может придавать мимика. Распознавание данного комплекса сигналов является сложной, но интересной и важной задачей, и в перспективе хотелось бы коснуться этого вопроса.

Литература

[1] А. Л. Воскресенский, С. Н. Ильин, M. Zelezny "О распознавании жестов языка глухих"

[2] В.Г. Абакумов, Е.Ю. Ломакина "Автоматическое распознавание жестов в интеллектуальных системах"/ статья

[3] "Язык жестов":- URL: http://jestov.net/

[4] Л. Шапиро, Дж. С. Стокман "Компьютерное зрение"

[5] Девид Форсайт, Жан Понс "Компьютерное зрение"

[6] Tom M. Mitchell "Machine Learning"

[7] A. Barr, Edvard A. Feigenbaum "The Handbook of Artifical Intelligence"

[8] Gary Bradski, Adrian Kaehler "Learning OpenCV"