

ДИНАМИЧЕСКОЕ РАСПОЗНАВАНИЕ ЖЕСТОВ РУКИ НА ОСНОВЕ СКРЫТЫХ МОДЕЛЕЙ МАРКОВА

Бауэр А.В.

Научный руководитель Болотова Ю.А. к.т.н., доцент
Томский политехнический университет, 634050, Россия, г. Томск, пр. Ленина 30
E-mail: anton.bauer@inbox.ru

Разрабатываемый метод позволяет в реальном времени захватывать и интерпретировать жесты руки. В таблице 1 приведена последовательность обработки жеста с описанием входных данных для каждого этапа.

Таблица 1. Описание метода.

| Этап | Входные данные |
|-----------------------------|--|
| Захват и отслеживание жеста | На вход алгоритма подается видеопоток, на каждом кадре которого отыскивается изображение руки. Для каждого изображения рассчитывается его дескриптор |
| Распознавание жеста | На вход классификатора поступает последовательность дескрипторов изображений. Классификатор возвращает идентификатор команды. |
| Исполнение команды | На вход фабрики команд поступает идентификатор команды. Команда выбирается из конечного множества поддерживаемых команд. |

В рассматриваемом методе для захвата жеста используется алгоритм Виолы-Джонса. Затем каждый жест представляется как последовательность HOG-дескрипторов.

Для распознавания жеста используются скрытые модели Маркова, которые являются эффективным методом представления пространственно-временной информации.

Алгоритм Виолы-Джонса для захвата жеста был выбран по следующим причинам:

- Этот алгоритм является одним из лучших по соотношению эффективность распознавания / скорость работы.
 - Метод Виолы-Джонса обладает очень низкой вероятностью ложного срабатывания.
 - Алгоритм в состоянии распознавать искомый объект, даже если на изображении он повернут на небольшой угол (до 30 градусов).
- Принципы работы алгоритма Виолы-Джонса:
- Интегральное представление изображения – используется для ускорения расчетов, связанных с яркостью пикселей изображения.
 - Использование признаков Хаара, с помощью которых происходит поиск объекта.
 - Обучение основывается на алгоритме AdaBoost для выбора наиболее подходящих признаков Хаара искомого объекта.

- Все признаки поступают на вход классификатора, который возвращает один из двух вариантов – «истина» для искомого объекта и «ложь» в противном случае.

- Используются каскады признаков для быстрого отбрасывания участков изображения, на которых не найден объект.

Задача интерпретации жестов решается с помощью скрытых моделей Маркова (Hidden Markov models, HMM).

Скрытые Марковские модели – это двойной стохастический процесс, управляемый двумя вероятностями:

1. Вероятность перехода, которая определяет, с какой вероятностью произойдет переход в каждое состояние.

2. Вероятность выхода определяет, какой из символов конечного алфавита будет на выходе сети в текущем состоянии.

Скрытые Марковские сети являются эффективной моделью представления пространственно-временной информации естественным образом. Модель называется «скрытой», так как все, что видно стороннему наблюдателю – лишь последовательность выходных символов, в то время как состояние, в котором находится модель, наблюдателю неизвестно. Таким образом, распознавание жеста основано на предположении о жесте и нет 100% уверенности в правильности предположения.

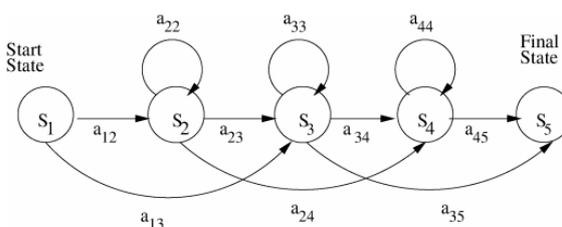


Рис. 1. Скрытая сеть Маркова.

Скрытая Марковская модель описывается следующим образом:

- Набор входных значений $O = \{O_1, \dots, O_T\}$ где $t = 1, \dots, T$.
- Множество состояний $\{s_1, \dots, s_N\}$.
- Дискретный набор наблюдаемых символов $\{v_1, \dots, v_k\}$.
- Матрица состояний-переходов $A = \{a_{ij}\}$, где a_{ij} - вероятность перехода из состояния s_i в момент времени t в состояние s_j в момент времени $t+1$.
- Вероятность наблюдаемого символа

$V = \{b_{jk}\}$, где b_{jk} - вероятность генерации v_k из состояния s_j

• Начальные вероятности $\Pi = \{\pi_j\}, j = 1, 2, \dots, N$ где π_j - вероятность нахождения сети в состоянии s_j в начальный момент времени.

Для решения задачи интерпретации жестов, для каждого жеста определяется наиболее коррелирующая с ним последовательность наблюдений, получаемых от скрытой модели Маркова. На каждой из последовательностей обучается одна Марковская модель, после чего этот набор моделей становится способен распознавать заданное множество жестов. Задача обучения скрытой марковской модели решается с помощью алгоритма Баума-Велша.

Гистограмма направленных градиентов (Histogram of Oriented Gradients, HOG) – дескрипторы особых точек изображения.

Основной идеей использования HOG-дескрипторов является допущение, что внешний вид и форма объекта на участке изображения могут быть описаны распределением градиентов интенсивности или направлением краев. Реализация этих дескрипторов может быть произведена путем деления изображения на маленькие связные области, именуемые ячейками, и расчетом для каждой ячейки гистограммы направлений градиентов или направлений краев для пикселей, находящихся внутри ячейки.

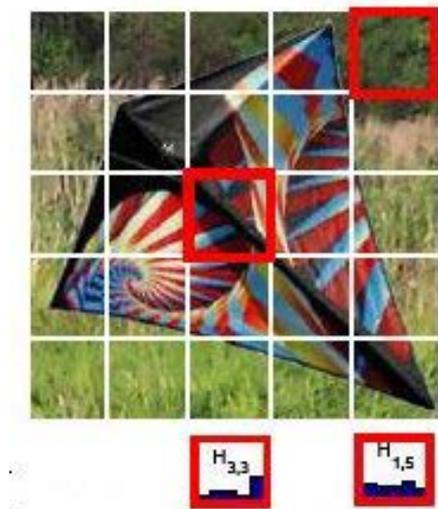


Рис. 2. HOG-дескрипторы.

Комбинация этих гистограмм и является дескриптором. Для увеличения точности локальные гистограммы подвергаются нормализации по контрасту. С этой целью вычисляется мера интенсивности на большем фрагменте изображения, который называется блоком, и полученное значение используется для нормализации. Нормализованные дескрипторы обладают лучшей инвариантностью по отношению к освещению.

Для тестирования системы используется Кембриджская база данных жестов.



Рис. 3. Пример кадра из базы жестов.

Подборка представляет собой 900 последовательностей изображений в 9 классах жестов. Класс определяется одним из 3-х примитивных движений и 3 формами ладони. Каждый класс содержит 100 последовательностей изображений – 5 различных типов освещения, 10 произвольных движений, 2 субъекта. Каждая последовательность была записана перед фиксированной камерой и изображает строго изолированный жест в пространстве и времени.

Литература

1. Vladimir I. Pavlovic, Rajeev Sharma and Thomas S. Huang, “Visual Interpretation of Hand Gestures for Human-Computer interaction.” in *Pattern Analysis and Machine Intelligence, IEEE Transaction, Vol.19 no. 7, pp 677-695, Jul. 1997*
2. Sushmita Mitra and Tinku Acharya, “Gesture recognition : A Survey”, *IEEE Transaction on systems, man, and cybernetics—part c: applications and reviews, vol. 37, no. 3, may 2007*