

Отчет по второму практическому заданию по курсу "Обработка и распознавание изображений": генерация признаков формы объектов на изображении

Выполнил студент 317 группы, Демьянов Иван

8 мая 2022 г.

Постановка задания

Данное задание было направлено на ознакомление с методами генерации признаков формы объектов на изображении. В нем было необходимо разработать и реализовать программу для классификации изображений ладоней. Программа должна обеспечить ввод и отображение на экране исходных изображений ладоней в формате TIF, их сегментацию на основе точечных и пространственных преобразований, а также определение «позы» ладоней и определение «линии пальцев» ладоней в качестве генерации признаков формы объектов на изображении.

Поза ладони определяется по расположению сомкнутых пальцев. Пальцы нумеруются от 1 до 5, начиная с большого пальца против часовой стрелки. Поза ладони описывается кодом $1*2*3*4*5$, где значок «*» обозначает «-», если пальцы разомкнуты, или «+», если они прижаты друг к другу.

Линия пальцев – ломаная линия, соединяющая точки на кончиках пальцев (tips) с точками в основаниях пальцев (valleys).

Описание данных

Для отладки и обучения алгоритма к заданию прилагался набор из 67 цветных изображений ладоней разных людей, полученных с помощью сканера, в формате 489×684 с разрешением 72 dpi. На изображениях представлены сканы левых рук. Ниже приведен пример входных данных для алгоритма.



Рис. 1: Изображения из обучающего набора данных

Как видно, на изображениях различаются не только «позы» рук, но и положение: оно может быть горизонтальным, вертикальным, наклонным.

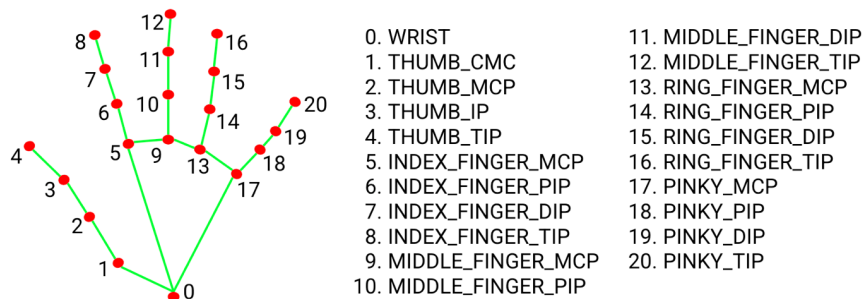
Описание метода решения

Поставленную задачу легче разбить на базовую - нахождение координат пикселей кончиков пальцев (5 пар координат) и других 4 пар координат, образующих вместе линию пальцев, а затем решать задачу определения «позы» ладоней.

Определение точек, образующих линию пальцев

К счастью, базовая задача уже решалась компанией Google и присутствует программная реализация данного решения (подробнее об этом в пункте "Описание программной реализации"). В частности [здесь](#) решалась похожая задача.

Understanding the hand landmark model



Для решения базовой задачи нам остается только взять интересующие нас точки. В качестве оснований ладони берется среднее между двумя ближайшими точками.

Преимущество такого подхода состоит в отсутствии необходимости сегментации изображения с последующей бинаризацией.

Ниже приведена визуализация результата работы алгоритма на нескольких объектах обучающей выборки.



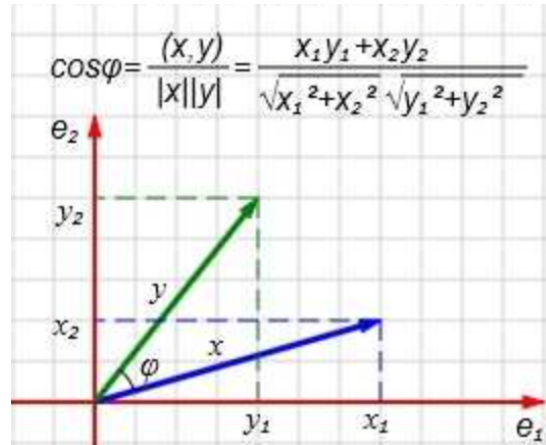
Рис. 2: Изображения из обучающего набора данных с размеченными линиями ладони

Как видно по результату, алгоритм хорошо справляется с поставленной задачей. Визуально качество на порядок лучше, чем при изначальных попытках поиска этих точек по контуру ладони.

Определение позы ладони

После решения базовой задачи мы имеем координаты точек, образующих линию пальцев. По этим координатам можно найти величины углов, которые образуются между кончиками двух соседних пальцев и основанием ладони, лежащем между ними.

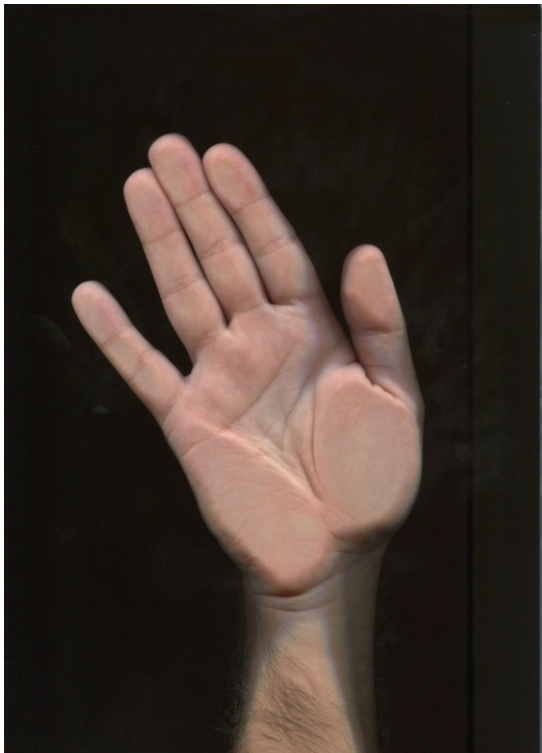
Для начала соответствующие 3 точки преобразуются в 2 вектора, далее находится угол между этими векторами при помощи скалярного произведения:



Затем производится подбор порогового значения, которое определяет сомкнуты ли пальцы.

Стоит отметить, что для улучшения качества прогноза пороговое значение для определения сомкнутости большого и указательного пальцев следует сделать чуть больше, это может быть объяснено особенностью строения ладони.

Ниже приведены результаты работы алгоритма на нескольких объектах, на которых производилась настройка порогового значения угла.



1+2+3+4+5



1-2-3+4+5



1-2-3-4-5



1-2-3+4-5

Рис. 3: Изображения из обучающего набора данных с размеченным кодом позы ладони

По итогу можно сказать, что данная стратегия имеет место быть, так как алгоритм хорошо справился с задачей на обучающих данных.

Описание программной реализации

Для реализации метода решения, описанного выше, был выбран язык Python с использованием библиотек `os`, `numpy`, `cv2`, `mediapipe`. В обработке и распознавании изображений значительный вклад внесли две последние библиотеки. Первая служит для получения доступа к данным, вторая предназначена для векторных вычислений.

Библиотека `mediapipe` внесла большой вклад в решение базовой задачи этого задания. Документацию библиотеки можно посмотреть [здесь](#).

Программа разбита на функции, которые решают определенную подзадачу, краткое описание их работы приведено ниже.

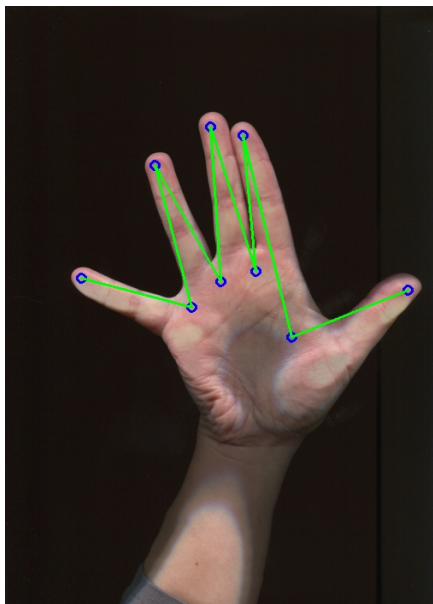
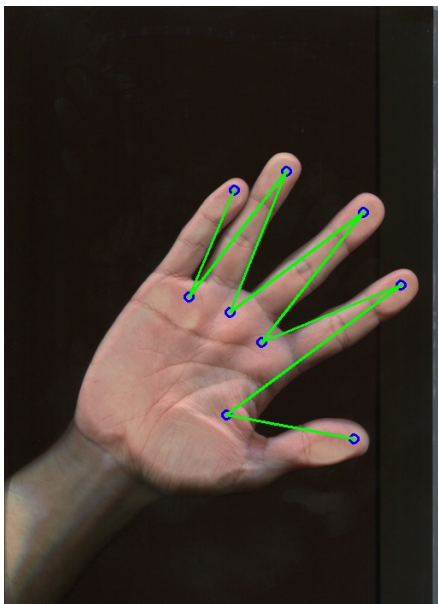
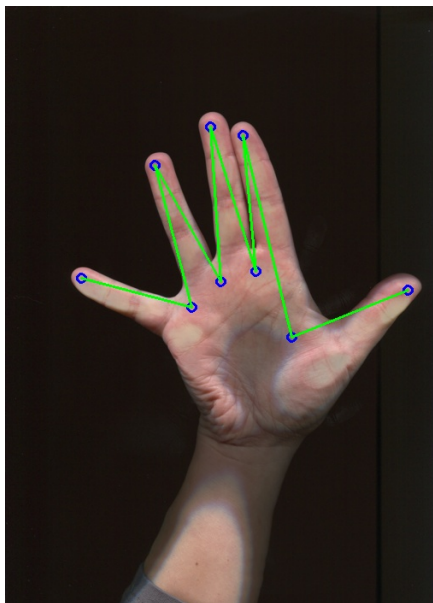
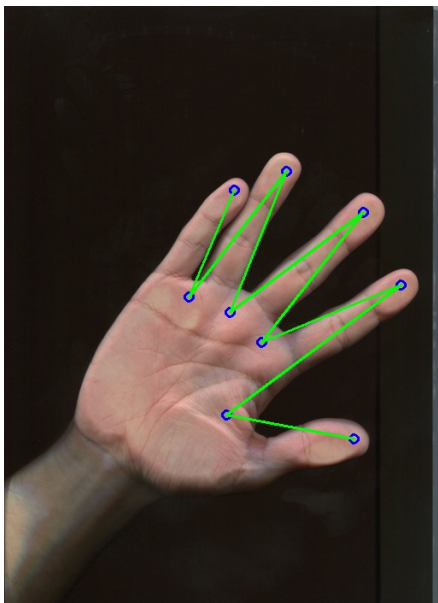
- Функция `get_handList_and_image(image_path)` принимает на вход путь до изображения и возвращает список координат особых точек, указанных выше, а также само считанное изображение в виде трехмерного тензора.
- Функция `make_actual_dots_list(dots_list)` отбирает из всех особых точек ладони нужные для решения поставленной задачи, а также делает преобразование, описанное в методе решения, для получения точек основания ладони.
- Функции `draw_circles(image, hand_lst)`, `draw_lines(image, hand_lst)` отвечают за визуализацию решения базовой задачи.
- Функция `expert_image_processing(image_path)` возвращает обработанное бинарное изображение, которое изначально имело пестрый фон.
- Функция `pose_definition(act_hand_List)` выдает прогноз кода позы ладони по списку расположения особых точек.
- Функции `make_vectors(a, b, c)`, `find_angle(a, b)` производят вспомогательные вычисления для предыдущей функции.
- Функции `big_finger_pose_predict(a, b)`, `pose_predict(a, b)` возвращают прогноз положения пальцев.

Демонстрация работы функций показана в приложенном `ipynb`-файле.

Эксперименты

В качестве экспериментов исследуем работу алгоритма, описанного выше на валидационных данных. Далее приведены ответы на изображениях, на которых алгоритм не обучался (не настраивались параметры).

Визуализация работы алгоритма на валидации при решении базовой задачи



Модель безошибочно работает на валидационных данных, так как часть решающая эту задачу была реализована с готовой библиотекой, которая не обучалась на наших данных. Это дает возможность качественного прогноза для задачи определения позы ладони.

Визуализация работы алгоритма на валидации при определении позы ладони



1-2-3-4+5



1-2+3-4-5



1-2-3+4-5



1-2-3-4-5

Как видно, модель делает вполне логичные предсказания, однако сложности возникают при близком, но не сомкнутом расположении соседних пальцев. В данном случае было бы хорошим тоном узнать у заказчика цену каждой ошибочной классификации позы.

Также возникают проблемы с классификацией положения большого и указательного пальцев. Это связано с особенностью строения ладони.

Для более точной классификации было бы неплохо сделать разметку данных и обучить классификатор классическими методами машинного обучения. Однако у этого подхода есть один большой минус - заметно большее время формирования прогноза, что может быть решающим фактором в пользу исходной модели при определении положения ладони в динамике.

В файле `Results.txt` находятся прогнозы модели на всех выданных данных. Стоит уточнить, что координаты основания выводились начиная с большого пальца.

Выводы

В данном отчете были представлены методы генерации признаков формы объектов на изображении для решения задачи распознавания линии ладони с последующей классификацией ее позы. Данные методы показали внушительное качество, однако из минусов такого подхода следует отметить ручной и поэтому порой весьма долгий подбор параметров.

Также в отчете присутствует описание программной реализации, в котором отмечена библиотека `cv2` как очень полезная в обработке изображений, благодаря ее большому функционалу. Немалый вклад в решение внесла и библиотека `mediapipe`, которая качественно решает базовую задачу.

И наконец, в отчете была проверена работа алгоритма на валидационных данных.