

Федеральное государственное автономное образовательное учреждение  
высшего образования

**«Национальный исследовательский университет  
«Высшая школа экономики»**

Факультет компьютерных наук  
ООП «Прикладная математика и информатика»  
**Отчёт о прохождении практики**

**Студент:** Кобелев Максим Олегович

**Группа:** 165

**Место прохождения  
учебной практики:** НИУ ВШЭ

Руководитель:

доцент: Факультет компьютерных наук / Департамент больших данных и информационного поиска, к.ф.-м.н.  
Конушин Антон Сергеевич

Москва, 2018

# ОГЛАВЛЕНИЕ

---

<b>АННОТАЦИЯ .....</b>	<b>3</b>
<b>1. СОВМЕЩЕНИЕ КАНАЛОВ ИЗОБРАЖЕНИЯ .....</b>	<b>4</b>
1.1    ПОСТАНОВКА ЗАДАЧИ.....	4
1.1    БАЗОВОЕ СМЕЩЕНИЕ .....	4
1.2    Совмещение с помощью пирамиды.....	5
<b>2 КОНТЕКСТНО-ЗАВИСИМОЕ МАСШТАБИРОВАНИЕ ИЗОБРАЖЕНИЙ .....</b>	<b>6</b>
2.1    ПОСТАНОВКА ЗАДАЧИ.....	6
2.2    ИДЕЯ АЛГОРИТМА .....	6
2.3    РАБОТА С МАСКОЙ ИЗОБРАЖЕНИЯ .....	7
2.4    Сжатие и растяжение изображения .....	7
<b>3 ЗАКЛЮЧЕНИЕ .....</b>	<b>8</b>
<b>4 СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ .....</b>	<b>9</b>

## АННОТАЦИЯ

---

Цель данной практики: получение базовых знаний в областях компьютерного зрения, а также знакомство с методами машинного обучения, применимых в этих областях. Прохождение практики подразумевает собой освоение 5 видеолекций курса «Введение в компьютерное зрение и глубинное обучение», состоящего из следующих тем:

- История и предмет компьютерного зрения
- Основы обработки изображений
- Особые точки и сопоставление изображений
- Введение в машинное обучение
- Классификация изображений

В качестве закрепления знаний, полученных после просмотра видеолекций и ознакомления с дополнительными материалами по каждой из тем, было предложено выполнить 3 задания.

- Реконструкция изображений Прокудина-Горского
- Изменение размеров изображения с сохранением пропорций контекста
- Распознавание дорожных знаков с помощью алгоритма классификации SVM

Каждое задание, необходимо было сдать в тестирующую систему, оценивающую его правильность в соответствии с критериями, указанными в задании.

# 1. СОВМЕЩЕНИЕ КАНАЛОВ ИЗОБРАЖЕНИЯ

## 1.1 ПОСТАНОВКА ЗАДАЧИ

Первым цветным фотографом России является Михаил Сергеевич Прокудин-Горский, сделавший единственный цветной портрет Льва Толстого. Каждый его снимок представляет из себя три изображения в градациях серого, соответствующие синему, зеленому и красному цветовым каналам. Сейчас коллекция его снимков находится в американской библиотеке конгресса, скан-копии фотопластинок доступны в интернете. Предлагается реализовать алгоритм, который будет совмещать изображения, полученные с фотопластинок Прокудина-Горского.



## 1.1 БАЗОВОЕ СМЕЩЕНИЕ

Наиболее примитивная реализация алгоритма совмещения красного, зеленого и синего цветовых каналов используется в алгоритме базового смещения. Идея алгоритма следующая – входное изображение, состоящее из трёх фотопластинок в градациях серого, разделяется на три разных изображения – красный, зеленый и синий слои. Для работы с более информативными частями изображения, исключив различные засветы и дефекты фотоплёнки – каждое из полученных изображений-слоёв обрезается по контуру на заранее заданную величину, зависящую от размеров всего изображения. Затем необходимо найти сдвиги красного и синего слоёв относительно зелёного, по которым наложение слоёв на выбранной метрике (среднеквадратичное отклонение или нормализованная кросс-корреляция) будет показывать наиболее оптимальное значение. Поиск сдвигов осуществляется простейшим перебором среди всевозможных пар сдвигов каждого слоя в окне от -15 до +15 пикселов, с вычислением значения метрики на каждом шагу и выявлением оптимальной конфигурации. Получив искомые значения сдвигов для красного и синего каналов относительно зеленого, несложно посчитать координаты соответствующих точек красного и синего слоёв, для заданных координат точки зелёного.

В этой части задания и был осуществлён выбор метрики использующейся в процессе выполнения оставшегося задания, а именно: при увеличении величины окна для поиска сдвигов нормализованная кросс-корреляция не давала существенного прироста визуальной оценки правильности выполнения алгоритма, в отличии от обычного среднеквадратического отклонения.



Рисунок 1: кросс-корреляция.



Рисунок 2: среднеквадратическое отклонение.

## 1.2 Совмещение с помощью пирамиды

Очевидно, что при совмещении больших изображений, размер которых исчисляется в тысячах пикселей, использование базового подхода может оказаться неэффективным и время обработки таких изображений станет очень долгим, потому что число пикселей в окне увеличивается пропорционально размерам нового изображения. Совмещение «с помощью пирамиды» - метод, который решает проблему долгой работы перебора с помощью алгоритма, работающего по схеме «разделяй и властвуй». Его реализация подразумевает собой написание рекурсивной функции, работающей следующим образом: для выхода из рекурсии - маленькие изображения с высотой не более 300 пикселей предлагается обрабатывать при помощи базового алгоритма нахождения сдвигов в окне размерами от -15 до +15 пикселей. В свою очередь, углубление в рекурсию происходит путём уменьшения размеров большого изображения в два раза, и вызова рекурсивной функции уже для уменьшенной копии. На пути наверх предлагаются увеличивать сдвиги  $dx$  и  $dy$ , найденные на очередной глубине вызова рекурсивной функции и производить поиск уже новых сдвигов в окне размерами  $[2dx - 1, 2dx + 1] \times [2dy - 1, 2dy + 1]$ .

Эмпирическим путём было установлено, что вышеописанные значения: пороговый размер 300 пикселей и уточнение сдвига на  $\pm 1$  пиксель на выходе из рекурсии, дают достойные результаты при хорошем приросте в скорости работы алгоритма.



Рисунок 3: 4k-изображение, алгоритм "с помощью пирамиды".

## 2 КОНТЕКСТНО-ЗАВИСИМОЕ МАСШТАБИРОВАНИЕ ИЗОБРАЖЕНИЙ

### 2.1 ПОСТАНОВКА ЗАДАЧИ

При стандартном подходе изображение равномерно деформируется по всей длине при изменении размера (объекты на изображении уменьшаются вместе со всем изображением). Это очень эффективно в плане потребления ресурсов (такое сжатие происходит довольно быстро), однако совсем не удовлетворяет требованиям, т.к. при его использовании получаются сильно искаженные объекты. Предлагаемый алгоритм учитывает важность объектов на изображении, и деформация происходит так, что объекты сохраняют свои первоначальные размеры.



Предлагается изучить алгоритм контекстно-зависимого масштабирования изображений (seam-carving) и реализовать функцию изменяющую размер изображения на 1 пиксель в длину или в высоту. Наиболее неочевидное преимущество данного алгоритма заключается в том, что при помощи изменения маски изображения возможно выделять области, подлежащие к удалению, или наоборот, защищать их от него. Известно, что такие популярные программы как Adobe Photoshop (начиная с версии CS4), и GIMP, используют open-source реализацию этого алгоритма от Mitsubishi Research Electric Laboratories (MERL) в качестве одного из методов для сжатия/растяжения изображения. Существуют также более сложные версии алгоритма, приспособленные для масштабирования видеозаписей.

### 2.2 ИДЕЯ АЛГОРИТМА

Идея алгоритма заключается в удалении/добавлении специальных кривых – швов, исчезновение или появление которых будет наименее заметно визуально, и для метрики. Шов – связанный кривая, соединяющая верхний и нижний (правый и левый) края изображения. В качестве метрики, для определения наиболее незаметного шва, среди всех ищем тот, который проходит через меньшее количество перепадов яркости. На каждом шаге алгоритма необходимо находить и удалять такой шов, имеющий минимальную энергию, где энергия точки – это модуль градиента в этом пикселе, а энергия шва – соответственной сумма энергий его точек. Заметим также, что благодаря тому, что в реальности изображения имеют сложную структуру, чем фон, швы, проходящие через объекты, будут иметь более высокую энергию, за счёт больших перепадов яркости и их количества. Значит такие сложные швы удаляться не будут, поэтому масштабирование начинает называться контекстно- зависимым.

Как было сказано выше, энергия пикселя – модуль градиента яркости в этой точке. Яркость изображения – компонента Y в цветовой модели YUV. Конверсия из RGB осуществлялась по следующей формуле:

$$Y = 0.299 * R + 0.587 * G + 0.114 * B$$

Далее необходимо найти градиент изображения. Для этого необходимо найти частные производные по каждому из направлений. Например, для того чтобы найти производную по направлению  $X$ , необходимо для каждого пикселя из правого соседа вычесть левый (нужно быть аккуратными на границах изображения). В общем же случае будем придерживаться таких формул:

$$I'_{Ox}(x, y) = Y(x + 1, y) - Y(x - 1, y)$$

$$I'_{Oy}(x, y) = Y(x, y + 1) - Y(x, y - 1)$$

Далее в каждой точке находится норма градиента и заполняется матрица энергии:

$$E_{x,y} = \sqrt{I'_{Ox}_{x,y}^2 + I'_{Oy}_{x,y}^2}$$

Для нахождения шва с наименьшей энергией, создается матрица энергии (таких же размеров, что и само изображение) и начинает динамически заполняться сверху-вниз (слева-направо) согласно правилу. Первая строка/столбец остаются нетронутыми. Затем очередной пиксель смотрит на 3 предыдущих, заполненных, связанных с ним стороной или углами пикселей, выбирает значение наименьшего из них и суммирует со своим, перезаписывая. После динамического пересчёта матрицы, необходимо найти наименьшее значение энергии в последней строке/столбце и обратно-восстановить его. Найденный шов и будет швом с наименьшей энергией.

### 2.3 РАБОТА С МАСКОЙ ИЗОБРАЖЕНИЯ

Если с помощью маски выделить какой-нибудь объект, то его можно или удалить из изображения, или наоборот защитить от этого. Таким образом изменяя параметры входной маски мы можем влиять на выбор швов нашим алгоритмом. Внутри это реализовано достаточно примитивно – матрица маски искусственно увеличивает/уменьшает значения в соответствующих пикселях матрицы энергии до пороговых (верхняя/нижняя граница) следующим образом:

$$E[x][y] \pm= mask[x][y] * image_{height} * image_{width} * 256$$

### 2.4 СЖАТИЕ И РАСТЯЖЕНИЕ ИЗОБРАЖЕНИЯ

После того как шов с минимальной энергией найден, для сжатия или растяжения изображения осталось выполнить лишь пару шагов. Подробнее рассмотрим каждую из подзадач:

#### 1. Сжатие.

Чтобы произвести сжатие изображения, осталось сделать наиболее очевидные вещи. Для сжатия на 1 пиксель, достаточно просто удалить найденный шов с минимальной энергией. На следующих итерациях для продолжения сжатия изображения, искомые швы будут пересчитываться, так как минимальные были удалены.

#### 2. Растяжение.

Для растяжения изображения необходимо справа от найденного минимального шва вставить новый, пиксели которого будут являться усреднением (в модели RGB) минимального и следующего за ним справа швов. Важно не забыть изменить значения маски добавив в неё минимальный шов, иначе при растяжении минимальный шов будет постоянно одним и тем же, а вставленные на каждой итерации соседние швы будут отличаться не более чем в 2 раза по интенсивности цвета (в модели RGB) между собой в пересчёте с минимальным швом. Данный приём поможет избежать появления размытых областей.

Ниже представлены примеры для каждой операции – сжатия и растяжения в 3 видах:

- Исходное изображение
- Применено контекстно-зависимое масштабирование
- Применено обычное масштабирование



Рисунок 4: Демонстрация сжатия

Рисунок 5: Демонстрация растяжения

### 3 ЗАКЛЮЧЕНИЕ

В результате прохождения практики были приобретены базовые знания из области компьютерного зрения, а также методами машинного обучения, которые были примененные в процессе работы. Были реализованы алгоритмы, использованные для реконструкции изображений Прокудина-Горского, и контекстно-зависимого масштабирования изображений.

Оба алгоритма показали отличный результат во время тестирования, однако если скорость алгоритма с совмещением каналов на больших изображениях была довольно высокой, то для крупного масштабирования больших изображений требовалось гораздо больше времени. Это одновременно и естественно при текущей реализации, и заставляет пересмотреть некоторые из используемых подходов в пользу более быстрых решений, что в свою очередь оставляет почву для дальнейших размышлений и улучшений уже реализованных алгоритмов.

## 4 СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ

---

- [1] Library of Congress. “Чудеса фотографии: восстановление фотографического наследия С.М.Прокудина-Горского” [<http://www.loc.gov/exhibits/empire/empire-ru.html>]
- [2] Navnet Dalal, Bill Triggs (2005) “Histograms of Oriented Gradients for Human Detection”. В: CVPR., pp.886-893
- [3] Satya Mallick. *Histogram of Oriented Gradients*. [<https://www.learnopencv.com/histogram-of-oriented-gradients>]