

Совмещение каналов изображения

Константин Кожемяков, Влад Шахуро, Александр Сергеев



Обзор задания

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, сканкопии фотопластинок доступны в интернете. В данном задании мы предлагаем вам создать программу, которая будет совмещать изображения, полученные с фотопластинок Прокудина-Горского.



Описание задания

1. Базовое совмещение

Реализация базовой части программы реализуется в несколько этапов:

1. Загрузка изображения и разделение изображения на три канала. Достаточно разделить изображение на три равные части по высоте.
2. Удаление рамок пленки. Каждый из каналов изображения нужно обрезать на 5% с каждой стороны.
3. Поиск наилучшего сдвига для совмещения каналов. Для того, чтобы совместить два изображения, будем сдвигать одно изображение относительно другого по горизонтали и по вертикали в некоторых пределах, например, от -15 до 15 пикселей. Далее, для перекрывающихся областей изображений посчитаем метрику. Оптимальным будет тот сдвиг, при котором метрика принимает наибольшее или наименьшее значение (в зависимости от метрики). Предлагается реализовать две метрики и выбрать ту, которая позволяет получить более качественный результат при совмещении:

(а) Среднеквадратичное отклонение для изображений I_1 и I_2 :

$$MSE(I_1, I_2) = \frac{1}{width \cdot height} \sum_{x,y} (I_1(x, y) - I_2(x, y))^2,$$

где $width, height$ — ширина и высота изображений соответственно. Для нахождения оптимального сдвига нужно взять минимум по всем сдвигам.

(b) Нормализованная кросс-корреляция для изображений I_1 и I_2 :

$$I_1 \star I_2 = \frac{\sum_{x,y} I_1(x,y) I_2(x,y)}{\sqrt{\sum_{x,y} I_1^2(x,y) \cdot \sum_{x,y} I_2^2(x,y)}}.$$

Для нахождения оптимального сдвига нужно взять максимум по всем сдвигам.

Совмещение больших изображений при базовом подходе будет проходить очень медленно. Для ускорения совмещения можно использовать пирамиду изображений.

В пирамиде изображений исходное изображение последовательно уменьшается в 2 раза до некоторого размера (например, чтобы обе стороны были не больше 500 пикселей в длину). Поиск оптимального сдвига начинается с самого маленького изображения, а затем на пути к исходному изображению уточняется на уменьшенных копиях изображения. Таким образом, оригинальное изображение совмещается не в диапазоне $-15 \dots 15$ пикселей, а в меньшем, уточненном с помощью уменьшенных копий изображения.

2. Совмещение с помощью преобразования Фурье

Когда ищутся достаточно большие сдвиги (например, при сшивании изображений или поиске части изображения на целом), пирамидальный подход может работать не очень хорошо: для больших сдвигов нужно сильнее уменьшать исходные изображения, что приводит к потере деталей. В этом случае хорошо работает подход, основанный на преобразовании Фурье.

Рассмотрим нормализованную кросс-корреляцию для двух изображений I_1 и I_2 . Для нахождения оптимального сдвига нужно взять максимум по всем сдвигам:

$$(u^*, v^*) = \operatorname{argmax}_{u,v} \frac{\sum_{x,y} I_1(x,y) I_2(x+u, y+v)}{\sqrt{\sum_{x,y} I_1^2(x,y) \cdot \sum_{x,y} I_2^2(x,y)}}.$$

Для сдвига изображения будем использовать циклический сдвиг. Выражение в знаменателе становится константой, на значение argmax не влияет:

$$(u^*, v^*) = \operatorname{argmax}_{u,v} \sum_{x,y} I_1(x,y) I_2(x+u, y+v).$$

Заметим, что справа находится операция корреляции. Преобразование Фурье позволяет заменить дорогую операцию корреляции на более дешевую операцию произведения матриц. Обозначим

$$C(u, v) = \sum_{x,y} I_1(x,y) I_2(x+u, y+v) = (I_1 \star I_2)(u, v).$$

Матрица C содержит значения кросс-корреляции двух изображений для всех возможных сдвигов. Координаты максимального значения матрицы C и есть искомый сдвиг. Вычислить эффективным образом матрицу C можно с помощью преобразования Фурье:

$$\mathcal{F}\{C(u, v)\} = \mathcal{F}\{(I_1 \star I_2)(u, v)\} = \left(\mathcal{F}\{I_1\} \cdot \overline{\mathcal{F}\{I_2\}} \right)(u, v),$$

$$C(u, v) = \mathcal{F}^{-1} \left(\mathcal{F}\{I_1\} \cdot \overline{\mathcal{F}\{I_2\}} \right)(u, v).$$

Здесь \mathcal{F} и \mathcal{F}^{-1} — прямое и обратное преобразование Фурье, $\overline{\mathcal{F}\{\cdot\}}$ — комплексное сопряжение. Таким образом, для нахождения оптимального сдвига необходимо вычислить

$$\operatorname{argmax} \mathcal{F}^{-1} \left(\mathcal{F}\{I_1\} \cdot \overline{\mathcal{F}\{I_2\}} \right).$$

Интерфейс программы, данные и скрипт для тестирования

Необходимо реализовать функцию `align`, принимающую на вход изображение, полученное сканированием фотопластинки, и возвращающую совмещенное изображение. Для полного решения задания необходимо реализовать *либо* базовое и пирамидальное совмещение, *либо* совмещение с помощью преобразования Фурье.

Данные для тестирования — 10 картинок в двух разрешениях для тестирования обычной реализации и реализации с пирамидой соответственно. На каждой картинке размечены три точки, по одной на каждом канале. Функция совмещения `align` должна по точке `(g_row, g_col)` зеленого канала определить координаты соответствующих ей точек синего и красного каналов: `(b_row, b_col)`, `(r_row, r_col)`. Для возвращенных функцией координат точек и координат разметки вычисляется метрика l_1 , которая затем сравнивается с порогом. Если метрика не превосходит порог, то изображение считается качественно совмещенным. Для маленьких изображений порог равен 5, для больших — 10. Скрипт для тестирования также проверяет, что возвращаемое функцией `align` изображение достаточно большого размера. Оценка за задание определяется как количество качественно совмещенных картинок, деленное на 2, максимум 10 баллов.

Полезные ресурсы

[Выставка](#) о Прокудине-Горском на сайте библиотеки конгресса.