

1 Глава 2. Введение в предмет Компьютерное зрение

1.1 Предмет классической обработки изображений

Предмет компьютерное зрение включает в себя изучение и применение фото изображений и их видео последовательностей для анализа и понимания реальных объектов или сцен, а также вопросы, относящиеся к реализации предложенных алгоритмов и их методологические проблемы [1].

Цифровое изображение можно описать как оцифрованное (дискретизация и квантование) представление непрерывных аналоговых данных в пространственной области. Цифровое изображение состоит из прямоугольного массива пикселей (x, y, u) . Каждый пиксель является комбинацией местоположения $(x, y) \in \mathcal{Z}^2$ и значения интенсивности u . Здесь \mathcal{Z} - множество всех целых чисел. Точки (x, y) образуют регулярную сетку. Отметим, что формально изображение \mathcal{I} определено на прямоугольном множестве Ω , которое называется носителем (carrier) \mathcal{I} и содержит узлы сетки:

$$\Omega = (x, y) : 1 \leq x \leq W_{\text{кол}} \wedge 1 \leq y \leq H_{\text{стр}} \subset \mathcal{Z}^2, \quad (1)$$

где $W_{\text{кол}}, H_{\text{стр}}$ узлы сетки по горизонтали и вертикали с равномерным шагом; $|\Omega| = W_{\text{кол}} \cdot H_{\text{стр}}$ - мощность носителя [1].

Если изображение цветное, то изображение может представлять собой набор нескольких значений u , соответствующих нескольким цветовым каналам. Для любого оцифрованного изображения можно ввести его размер $W \times H \times C$, где W, H - это ширина и высота изображения, C - число каналов.

Отметим, что как правило значения x, y, u - дискретные. Причем часто u выбираются в диапазоне $[0, 255]$ - то есть в 8-ми битном диапазоне, а шаг изменений пространственных координат по x и y равномерный. Таким образом можно сказать, что оцифрованное изображение, как правило, имеет 8-битное квантование и дискретизацию с равномерным шагом $\Delta x = \Delta y = x_2 - x_1$. Если число каналов $C = 1$ - то говорят об **одно-тоновых (монохромных) изображениях**, часто принято визуализировать такие изображения в оттенках серого, однако на самом деле оттенки будут отражать значения яркости u безотносительно какого бы ни было цвета. Также в ряде случаев мы будем иметь дело с бинарными изображениями. **Бинарные изображения** - это изображения для которых $C = 1$, $u = \{1, 0\}$, такие изображения принято визуализировать с использованием только двух цветов: черного и белого [1].

Таким образом, каждое оцифрованное изображение может быть представлено в виде матрицы или нескольких матриц, объединенных в так называемый тензор (в упрощенном смысле этого слова). Отметим, что если изображение описывается одной матрицей, то это не цвета, а светимость (физическая интенсивность) или яркость (воспринимаемая интенсивность). Для описания цветных изображений основными способами описания являются формат RGB (red, green, blue) и HSI (hue, saturation, lightness (intensity) - тон, насыщенность, светлота), а также некоторые другие. Иллюстрация представления цвета q в форматах RGB и HSI приведена на рисунке 1[1].

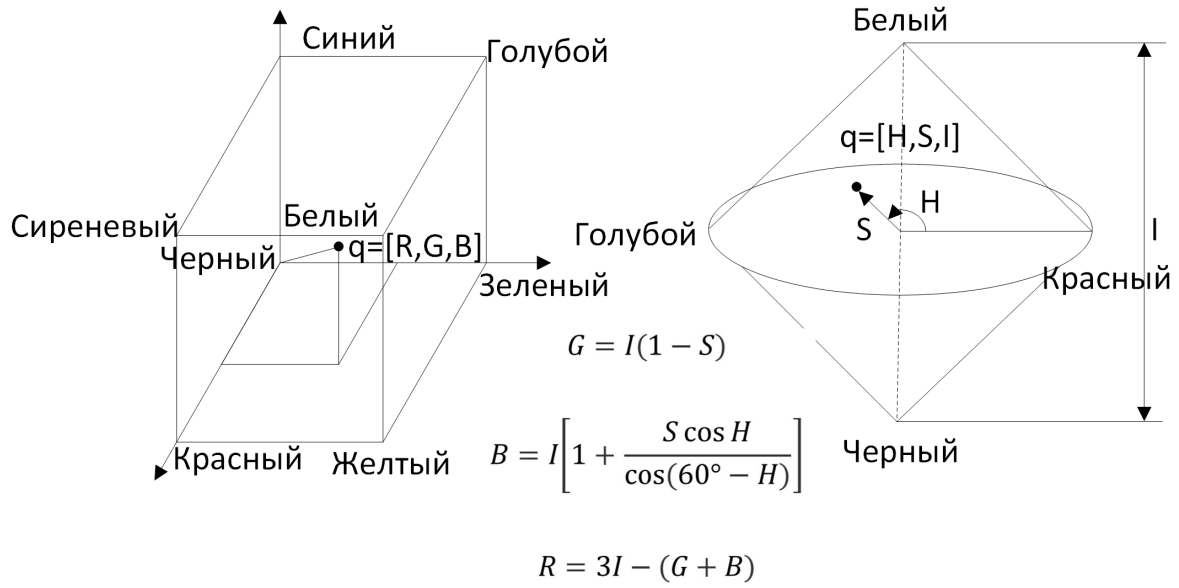


Рисунок 1 – Иллюстрация представления цвета q в форматах RGB и HSI

Для изображения можно ввести такие характеристики, как яркость, контрастность, гистограмма яркости и другие [1].

- **Средняя яркость изображения** может быть вычислена как

$$\mu_I = \frac{1}{|\Omega|} \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} I(x, y), \quad (2)$$

где μ_I - средняя яркость; $|\Omega| = W \cdot H$.

- **Контрастность** - средняя разность значений яркости каждого пикселя и соседних к нему

$$\nu = \frac{1}{|\Omega|} \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} |I(x, y) - \mu_w(x, y)|, \quad (3)$$

где ν - контрастность; $\mu_w(x, y)$ - среднее по окну пикселей вокруг пикселя с координатами.

- **Гистограмма яркости** может быть выражена как подсчет числа раз, когда встречается то или иное значение яркости канала u :

$$H(u) = |\{(x, y) \in \Omega : I(x, y) = u\}|, \quad h(u) = \frac{H(u)}{|\Omega|}, \quad CDF(u) = \sum_{v=0}^{u-1} h(v) \quad (4)$$

где $H(u)$ - абсолютная гистограмма; $h(u)$ - относительная гистограмма; $CDF(u)$ - кумулятивная функция гистограммы.

Гистограмма определяет распределение яркости по каждому оттенку цветового канала (число раз сколько встречается тот или иной оттенок u). Кумулятивная функция гистограммы показывает сколько раз встречаются все значения яркости вплоть до выбранного. Иллюстрации гистограммы одного и того же изображения, имеющего различную яркость приведены на рисунке [?].

- **Частотное представление изображения** (спектральное представление). Частотным является представление изображения в виде его разложение по двух-мерному ортогональному базису, например Фурье базису. То есть, в наиболее простом случае, частотное представление это двух-мерное преобразование Фурье от изображения. Результатом такого преобразования в общем случае является набор двух-мерных комплексных величин $\mathbf{I}(k, j)$, где k, j - это т.н. пространственные частоты. **Двухмерное прямое дискретное преобразование Фурье** может быть записано как

$$\mathbf{I}(k, j) = \frac{1}{|\Omega|} \sum_{x=0}^{W-1} \sum_{y=0}^{H-1} I(x, y) \exp \left(-i2\pi \left(\frac{xk}{W} + \frac{yj}{H} \right) \right), \quad (5)$$

где $i = \sqrt{-1}$.

Преобразование Фурье показывает, что изображение может быть разложено на набор гармонических базисных функций вида $\exp(-i2\pi(\frac{xk}{W} + \frac{yj}{H}))$. Как правило рассматривается только амплитудная часть спектра $|\mathbf{I}(k, j)|$. Для амплитудного спектра можно сделать следующие выводы. Наличие интенсивности изображения для функций с наиболее низкой (около-нулевой) частотой соответствует однородным частям массива. Наличие интенсивности для низких частот будет соответствовать медленно меняющимся частям изображения (например, размытие). Наличие интенсивности на высоких частотах, напротив, будет соответствовать резко-изменяющимся частям массива, например, контрастным границам. Примеры иллюстраций амплитудных спектров для некоторых изображений приведены на рисунке 3. Следует отметить, что на практике преобразование Фурье как правило реализуется при помощи т.н. алгоритма **Быстрого преобразования Фурье (БПФ, Fast Fourier Trans-**

form, FFT).

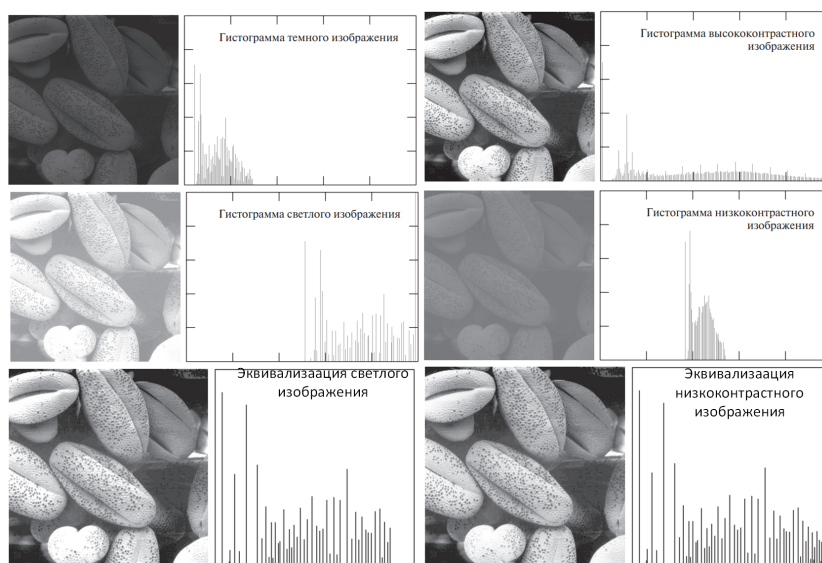


Рисунок 2 – Иллюстрации гистограммы яркости изображений

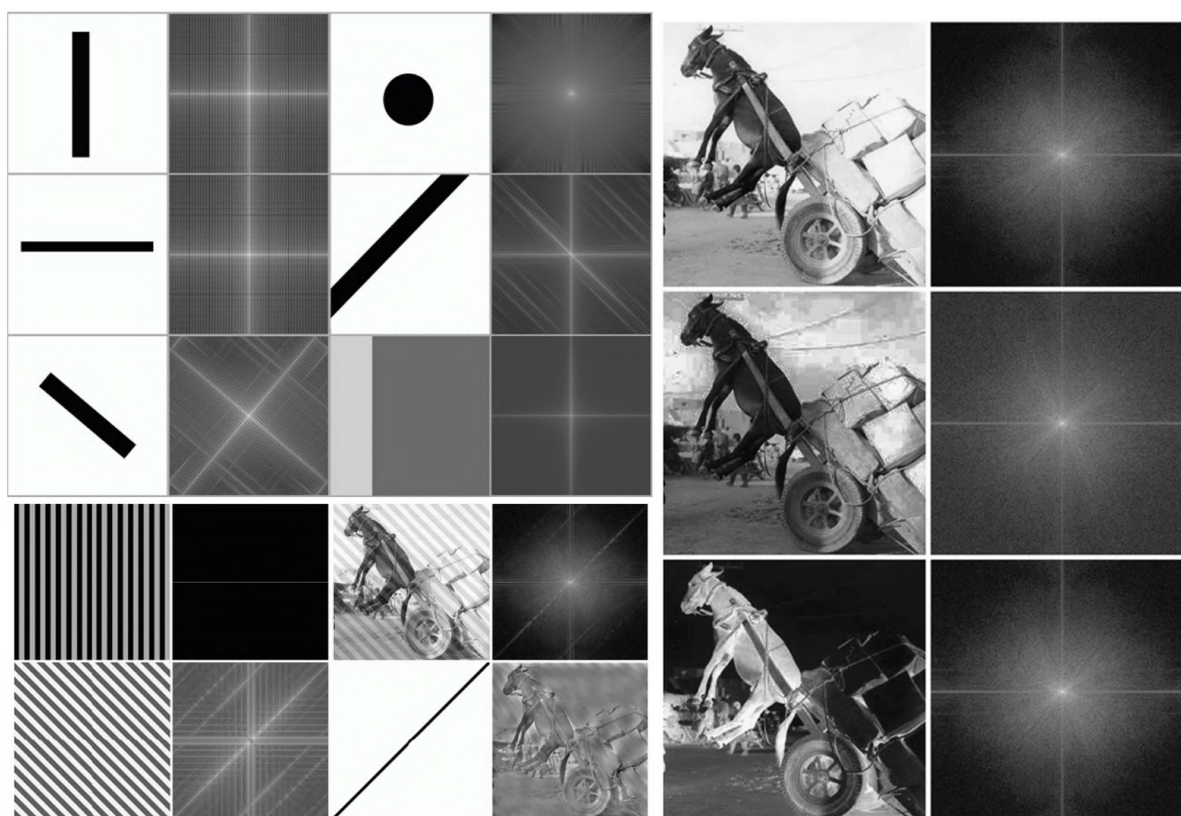


Рисунок 3 – Иллюстрации амплитудных спектров некоторых изображений

Отметим, что дискретное представление изображений приводит к ряду ограничений. Наиболее существенное из которых - частотное ограничение. Данное ограничение связано с условиями Теоремы Котельникова.

Теорема Котельникова (теорема Котельникова-Шенона-Найквиста) для цифровых изображений теорема сводится к тому, что максимальная пространственная частота в изображении не может быть выше чем $1/2\Delta x$. Таким образом спектр цифрового изображения ограничен. Это можно представить как ограничение резкости оцифрованного изображения. Так, для изображений низкого разрешения нельзя получить высокое качество выделения границ, особенно для небольших объектов. Если в спектре изображения должны появиться высокие частоты, то это приводит к т.н. артефактам наложения - то есть эффектам точечного искажения. Такой эффект называется **эффект алиасинга**.

1.2 Операции классической обработки изображений

Предмет классической обработки изображений - это достаточно большой курс и широко развитая область знаний. В данной книге эти вопросы не будут рассмотрены. Однако, интересующийся читатель может найти информацию, например, в таких книгах, как [2, 3, 4] и многих других.

Операции классической обработки изображений включают следующие типы [2].

- **Градационные преобразования** (попиксельные преобразования яркости) например,: усиление контраста, инверсия (получение негатива), степенная, кусочно-линейная, логорифмическая или другая нелинейная коррекция яркости.
- **Гистограммные преобразования**, например,: локальное или глобальное выравнивание гистограммы (эквивализация), приведение гистограммы к заданному виду или ее масштабирование.
- **Пространственная фильтрация изображений**. Данный тип операций основан на т.н. операции пространственной свертки, речь о которой пойдет ниже.
- **Нелинейная фильтрация изображений**. Данный тип операций основан, например, на использовании порядковых статистик (медианный фильтр), операциях расчета градиента от изображения или на использовании нечетких множеств.
- **Передискретизация изображений**. Данный тип операций прореживания и/или интерполяции значений пикселей с целью изменения общего размера изображения.

Операция пространственной свертки (двухмерная свертка, **convolution**) является широким понятием, которое возникает в различных приложениях статистической обработки данных. По существу, операция сводится к оценки "схожести" двух функций, скажем, $I(x, y)$ и $G(-x, -y)$, одна из которых смещается относительно другой как по оси x , так и по оси y . В данном определении операция свертки соответствует взаимокорреляционной функции между $I(x, y)$ и $G(x, y)$. Свертка может быть использована, например,

- для поиска в изображении $I(x, y)$ некоторых шаблонов (паттернов, образов, patterns) похожих на шаблоны, заданные в $G(x, y)$;
- для проведение пространственной фильтрации, например, высокочастотная фильтрация оставит только наиболее резкие участки изображения, такие, как границы объектов.

Математически операция свертки может быть описана как

$$r(x, y) = (G * I)(x, y) = \sum_{h=0}^{H_G-1} \sum_{w=0}^{W_G-1} G(h, w)I(x + h, y + w), \quad (6)$$

где:

- $G(h, w)$ - значение ядра свертки с для пикселя с координатами (h, w) (G имеет размерность $H_G \times W_G$);
- $I(x, y)$ - значение входного двухмерного массива с координатами (x, y) (I имеет размерность $H_I \times W_I$);
- $r(x, y)$ - значение результата свертки с координатами (x, y) (r имеет размерность $H_I - H_G + 1 \times W_I - W_H + 1$);
- $*$ - операция свертки.

На практике операция пространственной свертки как правило проводится между изображением и т.н. окном (window), которое также может быть названо ядро свертки (kernel). Тогда операция свертки может быть представлена так, как это показано на рисунках 4 и 5 в двух разных формах. Отметим, что здесь и далее предполагается, что направление движения окна фильтра слева на права и сверху вниз. Также отметим, что операция корреляции соответствует операции фильтрации с окном перевернутым на 180 градусов.

Как можно заметить из выражения (7) в общем случае операция свертки может привести к сокращению размеров изображения. Для того, чтобы этого не было перед сверткой изображение может быть дополнено границей из нулей с шириной с каждого блока по горизонтали $H_G/2-1$; по вертикали $W_G/2-1$. Операция дополнения нулями называется **padding** или zero padding если быть более конкретным.

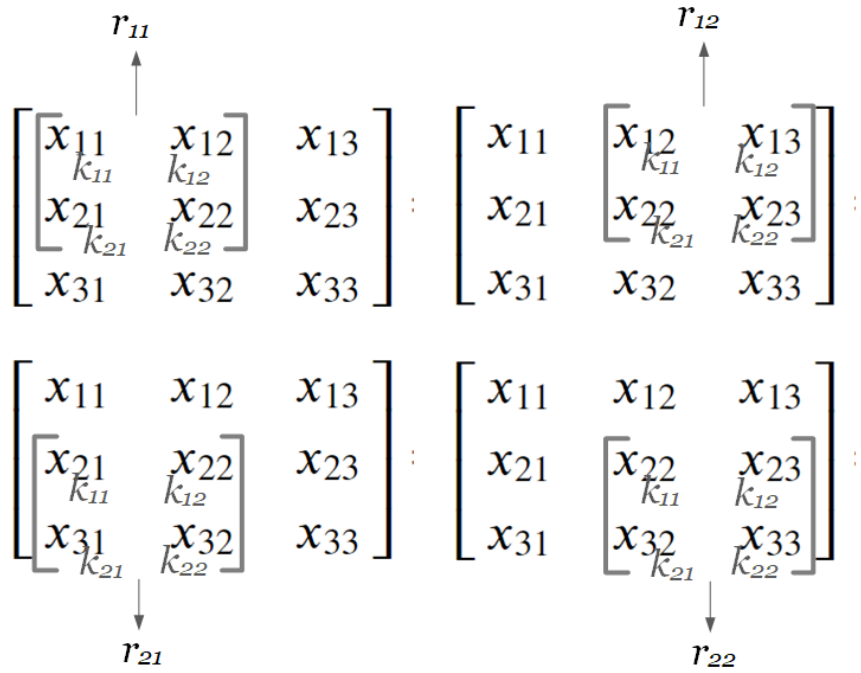


Рисунок 4 – Иллюстрация двухмерной свертки 2×2

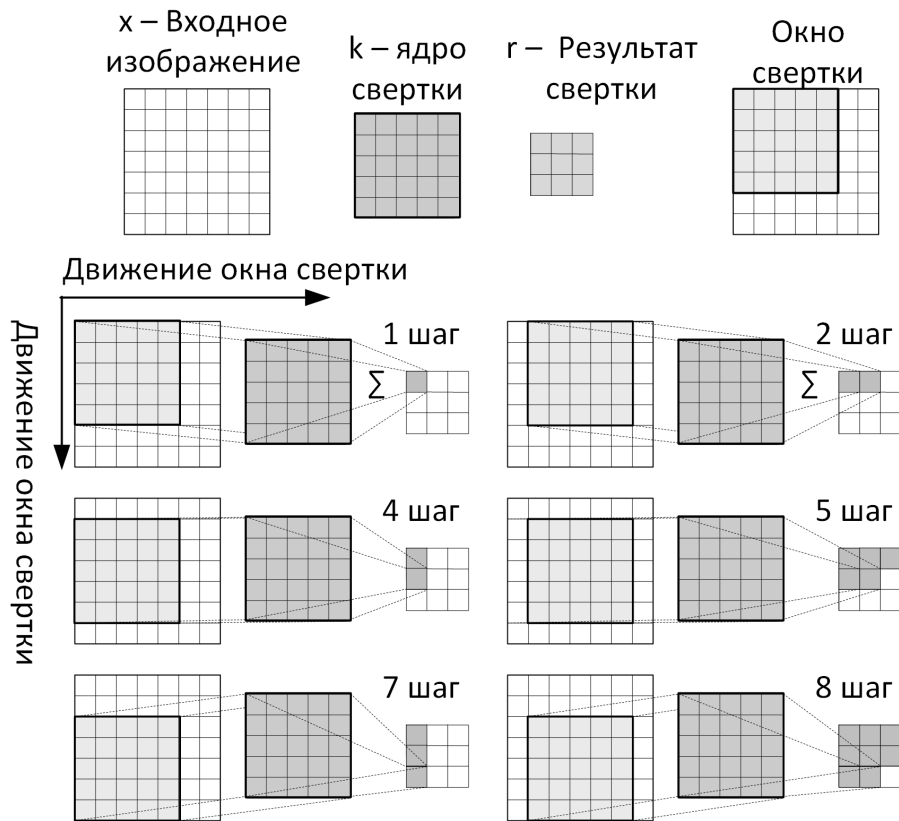


Рисунок 5 – Иллюстрация двухмерной свертки 5×5

Часто, особенно в классической обработке изображений, для вычисления свертки вместо формулы (7) используют важное свойство свертки, которое может быть записано как

$$r = (G * I) = \text{FFT}^{-1}(\mathbf{I} \cdot \mathbf{G}), \quad (7)$$

где \mathbf{I} и \mathbf{G} - это Фурье-спектры изображения и ядра свертки, имеющие равный размер; FFT^{-1} - операция обратного преобразования Фурье. Если операция выполняется между функциями разной размерности, то изображение меньшей размерности должно быть дополнено нулями с краев до равных значений со вторым изображением. Также операция преобразования Фурье выполняется для изображений четной размерности.

<https://arxiv.org/ftp/arxiv/papers/1910/1910.13796.pdf> Deep Learning vs. Traditional Computer Vision

СПИСОК ЛИТЕРАТУРЫ

- [1] Клетте Рейнхард. Компьютерное зрение. Теория и алгоритмы. Litres, 2019.
- [2] Гонсалес Рафаэл, Вудс Ричард. Цифровая обработка изображений. Litres, 2019.
- [3] Яне Бернд. Цифровая обработка изображений. 2007.
- [4] Solomon Chris, Breckon Toby. Fundamentals of Digital Image Processing: A practical approach with examples in Matlab. John Wiley & Sons, 2011.