

Рейнхард Клетте

Компьютерное зрение

Теория и алгоритмы

 Springer


Издательство

Рейнхард Клетте

Компьютерное зрение

Теория и алгоритмы

Reinhard Klette

Concise Computer Vision

An Introduction into Theory and Algorithms

Рейнхард Клетте

Компьютерное зрение

Теория и алгоритмы



Москва, 2019

УДК 004.8
ББК 32.81
К48

К48 Рейнхард Клетте

Компьютерное зрение. Теория и алгоритмы / пер. с англ. А. А. Слинкин. – М.: ДМК Пресс, 2019. – 506 с.: ил.

ISBN 978-5-97060-702-2

В этой книге рассмотрены основные аспекты компьютерного зрения: обработка и анализ изображений, анализ плотного движения, сегментация изображений, работа с камерами, трехмерная реконструкция, сопоставление стереоизображений, обнаружение объектов и др. Материал дополняется историческими справками, рекомендациями по дальнейшему чтению и сведениями о рассматриваемых математических понятиях. В конце каждой главы имеются проверенные на практике упражнения и вопросы на понимание материала.

Издание предназначено широкому кругу специалистов по анализу данных и изображений, а также может использоваться в качестве учебника для студентов старших курсов и для самообразования.

УДК 004.8
ББК 32.81

First published in English under the title Concise Computer Vision; Copyright © Springer-Verlag London, 2014. This edition has been translated and published under licence from Springer-Verlag London Ltd., part of Springer Nature. Springer-Verlag London Ltd., part of Springer Nature takes no responsibility and shall not be made liable for the accuracy of the translation. © 2019 by DMK Press. All rights reserved.

Все права защищены. Любая часть этой книги не может быть воспроизведена в какой бы то ни было форме и какими бы то ни было средствами без письменного разрешения владельцев авторских прав.

Материал, изложенный в данной книге, многократно проверен. Но, поскольку вероятность технических ошибок все равно существует, издательство не может гарантировать абсолютную точность и правильность приводимых сведений. В связи с этим издательство не несет ответственности за возможные ошибки, связанные с использованием книги.

ISBN 978-1-4471-6319-0 (англ.)
ISBN 978-5-97060-702-2 (рус.)

© Springer-Verlag London, 2014
© Оформление, перевод на русский язык, издание,
ДМК Пресс, 2019

Всем, кто любит мечтать



С помощью компьютерного зрения можно посчитать деревья, оценить расстояние до островов, но нельзя понять, что грезится людям, заглянувшим в эту бухточку.

Оглавление

Предисловие	11
Предмет книги.....	11
Характер изложения.....	11
Целевая аудитория.....	12
Использование материала на практике.....	12
Замечания для преподавателя и рекомендуемый порядок использования книги.....	13
Дополнительные ресурсы.....	13
Благодарности (в алфавитном порядке по фамилии).....	13
Обозначения	16
Глава 1. Данные изображения	18
1.1. Изображения в пространственной области.....	18
1.1.1. Пиксели и окна.....	19
1.1.2. Значения и основные статистики изображения.....	21
1.1.3. Пространственные и временные меры данных.....	26
1.1.4. Ступенчато-граничная модель.....	29
1.2. Изображения в частотной области.....	34
1.2.1. Дискретное преобразование Фурье.....	34
1.2.2. Обратное дискретное преобразование Фурье.....	35
1.2.3. Комплексная плоскость.....	37
1.2.4. Данные изображения в частотной области.....	39
1.2.5. Фазово-конгруэнтная модель признаков изображения.....	44
1.3. Цвет и цветные изображения.....	47
1.3.1. Определения цвета.....	49
1.3.2. Цветовое восприятие, дефекты зрения и уровни серого.....	51
1.3.3. Представления цвета.....	56
1.4. Упражнения.....	61
1.4.1. Упражнения по программированию.....	61
1.4.2. Упражнения, не требующие программирования.....	64
Глава 2. Обработка изображений	65
2.1. Точечные, локальные и глобальные операторы.....	65
2.1.1. Градационные функции.....	65
2.1.2. Локальные операторы.....	68
2.1.3. Фильтрация Фурье.....	71
2.2. Три процедурных компонента.....	75
2.2.1. Интегральные изображения.....	75
2.2.2. Регулярные пирамиды изображений.....	76
2.2.3. Порядок обхода.....	78

2.3. Классы локальных операторов	80
2.3.1. Сглаживание	80
2.3.2. Повышение резкости	85
2.3.3. Простые детекторы границ	86
2.3.4. Простые детекторы углов	91
2.3.5. Удаление артефактов освещения	95
2.4. Более сложные детекторы границ	98
2.4.1. LoG и DoG и их пространства масштабов	98
2.4.2. Встроенная уверенность	103
2.4.3. Алгоритм Ковеси	107
2.5. Упражнения	113
2.5.1. Упражнения по программированию	113
2.5.2. Упражнения, не требующие программирования	115
Глава 3. Анализ изображений	117
3.1. Основы топологии изображений	117
3.1.1. 4- и 8-смежность в бинарных изображениях	118
3.1.2. Топологически непротиворечивая смежность пикселей	123
3.1.3. Трассировка краев	126
3.2. Анализ двумерных геометрических фигур	130
3.2.1. Площадь	130
3.2.2. Длина	133
3.2.3. Кривизна	136
3.2.4. Дистанционное преобразование	140
3.3. Анализ значений изображения	147
3.3.1. Матрицы совместной встречаемости и метрики	148
3.3.2. Анализ участков с привлечением моментов	151
3.4. Поиск прямых и окружностей	153
3.4.1. Прямые	153
3.4.2. Окружности	160
3.5. Упражнения	162
3.5.1. Упражнения по программированию	162
3.5.2. Упражнения, не требующие программирования	167
Глава 4. Анализ плотного движения	169
4.1. 3D-движение и двумерный оптический поток	169
4.1.1. Локальное смещение и оптический поток	169
4.1.2. Проблема апертуры и градиентный поток	173
4.2. Алгоритм Хорна–Шанка	175
4.2.1. Подготовительная часть	175
4.2.2. Алгоритм	183
4.3. Алгоритм Лукаса–Канаде	187
4.3.1. Линейное решение методом наименьших квадратов	188
4.3.2. Оригинальный алгоритм и алгоритм с весами	191

4.4. Алгоритм ВВРВ	193
4.4.1. Исходные предположения и функция энергии.....	193
4.4.2. Краткое описание алгоритма	195
4.5. Оценка качества алгоритмов вычисления оптического потока	197
4.5.1. Стратегии тестирования	197
4.5.2. Меры ошибки для сравнения с контрольными данными.....	200
4.6. Упражнения	202
4.6.1. Упражнения по программированию.....	202
4.6.2. Упражнения, не требующие программирования	204
Глава 5. Сегментация изображений	206
5.1. Простые примеры сегментации изображений	207
5.1.1. Бинаризация изображения.....	209
5.1.2. Сегментация путем выращивания семян.....	212
5.2. Сегментация методом сдвига среднего.....	218
5.2.1. Примеры и подготовка	218
5.2.2. Модель сдвига среднего	221
5.2.3. Алгоритмы и оптимизация по времени.....	224
5.3. Сегментация изображений как задача оптимизации	230
5.3.1. Метки, пометка и минимизация энергии.....	231
5.3.2. Примеры членов данных и гладкости.....	234
5.3.3. Передача сообщений.....	237
5.3.4. Алгоритм распространения доверия.....	239
5.3.5. Распространение доверия в задаче о сегментации изображений.....	245
5.4. Сегментация видео и прослеживание сегментов.....	247
5.4.1. Использование согласованности признаков изображений.....	248
5.4.2. Использование временной согласованности	249
5.5. Упражнения	254
5.5.1. Упражнения по программированию.....	254
5.5.2. Упражнения, не требующие программирования	257
Глава 6. Камеры, координаты и калибровка	262
6.1. Камеры	263
6.1.1. Свойства цифровой камеры	263
6.1.2. Центральная проекция.....	268
6.1.3. Система с двумя камерами.....	271
6.1.4. Системы панорамных камер.....	273
6.2. Координаты	276
6.2.1. Мировые координаты.....	276
6.2.2. Однородные координаты.....	279
6.3. Калибровка камеры.....	281
6.3.1. Калибровка камеры с точки зрения пользователя.....	282
6.3.2. Ректификация пар стереоизображений.....	286
6.4. Упражнения	292

6.4.1. Упражнения по программированию.....	292
6.4.2. Упражнения, не требующие программирования.....	294
Глава 7. Трехмерная реконструкция	296
7.1. Поверхности	296
7.1.1. Топология поверхности	297
7.1.2. Локальные параметризации поверхности	301
7.1.3. Кривизна поверхности	304
7.2. Структурная подсветка.....	308
7.2.1. Проекция световой плоскости.....	308
7.2.2. Анализ световой плоскости	311
7.3. Стереоскопическое зрение.....	314
7.3.1. Эпиполярная геометрия.....	314
7.3.2. Бинокулярное зрение в канонической геометрии стереоскопической системы.....	316
7.3.3. Бинокулярное зрение в конвергентной системе.....	319
7.4. Фотометрический метод анализа стереоизображений.....	324
7.4.1. Ламбертовская отражательная способность.....	325
7.4.2. Восстановление градиентов поверхности.....	327
7.4.3. Интегрирование градиентных полей.....	331
7.5. Упражнения.....	340
7.5.1. Упражнения по программированию.....	340
7.5.2. Упражнения, не требующие программирования	343
Глава 8. Сопоставление стереоизображений	345
8.1. Сопоставление, стоимость данных и уверенность.....	346
8.1.1. Общая модель сопоставления.....	347
8.1.2. Функции стоимости данных.....	351
8.1.3. От глобального сопоставления к локальному.....	354
8.1.4. Тестирование функций стоимости данных.....	358
8.1.5. Меры уверенности.....	360
8.2. Сопоставление методом динамического программирования	362
8.2.1. Динамическое программирование	363
8.2.2. Ограничение порядка.....	364
8.2.3. Сопоставление методом динамического программирования с использованием ограничения порядка	367
8.2.4. Алгоритм DPM с ограничением гладкости	373
8.3. Сопоставление методом распространения доверия.....	379
8.4. Техника третьего глаза.....	382
8.4.1. Построение виртуального изображения, снятого третьей камерой.....	383
8.4.2. Сходство между виртуальным и третьим изображениями	388
8.5. Упражнения	391
8.5.1. Упражнения по программированию.....	391
8.5.2. Упражнения, не требующие программирования	393

Глава 9. Обнаружение и прослеживание признаков	396
9.1. Инвариантность, признаки и множества признаков	396
9.1.1. Инвариантность.....	396
9.1.2. Особые точки и векторы трехмерного потока	398
9.1.3. Множества особых точек в соседних кадрах.....	402
9.2. Примеры признаков	407
9.2.1. SIFT-признаки	407
9.2.2. SURF-признаки	410
9.2.3. ORB-признаки	411
9.2.4. Оценка признаков.....	414
9.3. Прослеживание и обновление признаков	417
9.3.1. Прослеживание – задача разреженного соответствия.....	419
9.3.2. Прослеживатель Лукаса–Канаде.....	420
9.3.3. Фильтр частиц.....	426
9.3.4. Фильтр Калмана.....	435
9.4. Упражнения	442
9.4.1. Упражнения по программированию.....	442
9.4.2. Упражнения, не требующие программирования	447
Глава 10. Обнаружение объектов	449
10.1. Локализация, классификация и оценка	449
10.1.1. Дескрипторы, классификаторы и обучение	450
10.1.2. Качество детекторов объектов.....	456
10.1.3. Гистограмма ориентированных градиентов	458
10.1.4. Вейвлеты и признаки Хаара.....	460
10.1.5. Метод Виолы–Джонса.....	464
10.2. AdaBoost.....	467
10.2.1. Алгоритм	468
10.2.2. Параметры	470
10.2.3. Почему именно такие параметры?.....	474
10.3. Случайные решающие леса	476
10.3.1. Энтропия и информационный выигрыш	476
10.3.2. Применение леса	479
10.3.3. Обучение леса.....	481
10.3.4. Леса Хафа.....	486
10.4. Обнаружение пешеходов.....	488
10.5. Упражнения.....	491
10.5.1. Упражнения по программированию	491
10.5.2. Упражнения, не требующие программирования	493
Предметный указатель	495

Предисловие

Это учебник для студентов третьего и четвертого года обучения, посещающих курс компьютерного зрения – одной из научно-технических дисциплин.

Предмет книги

Задача компьютерного зрения – применение фото- и видеокамер для анализа или понимания реальных сцен. Эта дисциплина изучает методологические и алгоритмические проблемы, а также вопросы, относящиеся к реализации предложенных решений.

В компьютерном зрении нас может интересовать, на каком расстоянии от камеры находится здание, движется ли автомобиль посередине своей полосы, сколько человек присутствует на изображении и даже присутствует ли на нем конкретный человек. На все эти вопросы нужно дать ответ, имея лишь сохраненные фотографии или видео. Благодаря недавним достижениям область применимости компьютерного зрения значительно расширилась. Прогресс затронул как технологии изготовления камер и методы вычислений, так и теоретические основания компьютерного зрения.

За последние годы компьютерное зрение стало ключевой технологией во многих областях. Если речь идет о потребительской электронике, то достаточно упомянуть мобильные телефоны, помощь водителям или взаимодействие пользователя с компьютерными играми. В автоматизации промышленности компьютерное зрение повсеместно применяется для контроля качества и технологических процессов. Существенный вклад компьютерное зрение внесло в киноиндустрию (вспомним, например, аватары или создание виртуальных миров на основе отснятых изображений, дополнение исторических видеоданных или высококачественные презентации фильмов). И это лишь малая часть областей применения, связанных с обработкой и анализом фотографий и видео.

Характер изложения

В книге дается общее введение в основы компьютерного зрения, которые находят применения в самых разных прикладных областях. Математика играет важную роль в изложении, обсуждаются также алгоритмы. Конкретные приложения в книге не рассматриваются.

На врезках представлены историческая информация, ссылки на источники излагаемого материала и сведения о математических объектах в том месте, где они впервые встречаются. Врезки следует рассматривать как *дополнения* к основному тексту.

Книга *не* является справочником по текущим исследованиям в области компьютерного зрения, и объем библиографических ссылок *очень ограничен*; читатель легко найдет интересующий его материал в сети. Компьютерное зрение – активно развивающаяся дисциплина, количество ссылок огромно, поэтому любая попытка втиснуть полную библиографию в ограниченное пространство книги обречена на провал. Но мы, по крайней мере, дадим совет: на странице homepages.inf.ed.ac.uk/rbf/CVonline/ вы найдете онлайн-овое введение в различные проблемы компьютерного зрения.

Целевая аудитория

Книга задумывалась как вводный курс для студентов третьего-четвертого курсов, специализирующихся в научных или технических дисциплинах. Предварительные знания в области обработки и анализа изображений или компьютерной графики были бы полезны, но первые две главы содержат начальные сведения о компьютерной обработке изображений.

Использование материала на практике

Материалы, представленные в книге, частично использовались в моих лекциях по мехатронике и информатике в Оклендском университете (Новая Зеландия), в Центре исследований и математики (CIMAT) в Гуанахуато (Мексика), в университете Фрайбурга и Гёттингена (Германия), в Кордобском техническом университете (Аргентина), в Тайваньском национальном нормальном университете (Тайвань) и в Уханьском университете (Китай).

В книге использовались также материалы из предыдущих публикаций автора: [R. Klette and P. Zamperoni. Handbook of Image Processing Operators. Wiley, Chichester, 1996], [R. Klette, K. Schlüns and A. Koschan. Computer Vision. Springer, Singapore, 1998], [R. Klette and A. Rosenfeld. Digital Geometry. Morgan Kaufmann, San Francisco, 2004] и [F. Huang, R. Klette and K. Scheibe. Panoramic Imaging. Wiley, West Sussex, 2008].

Первые две из перечисленных выше книг использовались как дополнение к лекциям автора по компьютерному зрению в Германии и Новой Зеландии в 1990-х и в начале 2000-х г., а третья – в лекциях, прочитанных в более поздний период.

Замечания для преподавателя и рекомендуемый порядок использования книги

Книга содержит больше материала, чем можно охватить в односеместровом курсе. Преподавателю следует отбирать материал, сообразуясь с контекстом: предварительными знаниями студентов и направленностью последующих курсов.

Каждая глава заканчивается упражнениями, в том числе предполагающими программирование. В книге не отдается предпочтение какой-либо конкретной среде. Но обычно применение функций, имеющихся в таких библиотеках, как OpenCV, упрощает решение. Упражнения по программированию намеренно сформулированы так, чтобы у студентов было несколько вариантов ответа. Например, в упражнении 2.5 к главе 2 можно визуализировать результаты с помощью Java-апплетов (хотя в тексте этого не требуется), использовать изображения большого или малого размера (в тексте размер не указан) и ограничить движение курсора центральной частью входного изображения, например квадратом 11×11 с центром в точке p , который всегда целиком содержится в изображении (или рассмотреть также особые случаи, когда курсор может приближаться к границе изображения). Поэтому каждый студент может дать свое решение упражнений, и творческий подход следует всячески приветствовать.

Дополнительные ресурсы

На сопроводительном сайте книги имеются дополнительные материалы (данные, исходный код, примеры, презентации). См. www.cs.auckland.ac.nz/~rklette/Books/K2014/.

Благодарности (в алфавитном порядке по фамилии)

Я благодарен перечисленным ниже коллегам, бывшим и нынешним студентам и друзьям (если упоминается только рисунок, то я выражаю признательность за совместную работу или контакты, относящиеся к теме рисунка).

- **A-Kn**

Ali Al-Sarraf (рис. 2.32), Hernan Badino (рис. 9.25), Anko Börner (различные комментарии по черновым вариантам книги и вклад в

раздел 5.4.2), Hugo Carlos (помощь во время работы над книгой в CIMAT), Diego Caudillo (рис. 1.9, 5.28, 5.29), Gilberto Chávez (рис. 3.39 и 5.36, верхний ряд), Chia-Yen Chen (рис. 6.21 и 7.25), Kaihua Chen (рис. 3.33), Ting-Yen Chen (рис. 5.35, вклад в раздел 2.4, в главу 5, предоставление исходного кода), Eduardo Destefanis (вклад в пример 9.1 и рис. 9.5), Uwe Franke (рис. 3.36, 6.3 и 9.23, правый нижний), Stefan Gehrig (замечания по вопросу анализа стереоизображений и рис. 9.25), Roberto Guzmán (рис. 5.36, нижний ряд), Wang Han (привлечение студентов к проверке черновика книги), Ralf Haeusler (вклад в раздел 8.1.5), Gabriel Hartmann (рис. 9.24), Simon Hermann (вклад в разделы 5.4.2 и 8.1.2, рис. 4.16 и 7.5), Václav Hlaváč (предложения по улучшению состава глав 1 и 2), Heiko Hirschmüller (рис. 7.1), Wolfgang Huber (рис. 4.12, правый нижний), Fay Huang (вклад в главу 6, особенно в раздел 6.1.4), Ruyi Jiang (вклад в раздел 9.3.3), Waqar Khan (рис. 7.17), Ron Kimmel (предложения о способе изложения локальных операторов и оптического потока – они вынужденно запланированы в основном для следующего издания книги), Karsten Knoerppel (вклад в раздел 9.3.4).

- **Ko-Sc**

Andreas Koschan (замечания к различным частям книги и рис. 7.18, справа), Vladimir Kovalevsky (рис. 2.15), Peter Kovesi (вклад в главы 1 и 2, касающийся фазовой конгруэнтности, включая разрешение на публикацию рисунков), Walter Kropatsch (предложения по главам 2 и 3), Richard Lewis-Shell (рис. 4.12, левый нижний), Fajie Li (упражнение 5.9), Juan Lin (вклад в раздел 10.3), Yizhe Lin (рис. 6.19), Dongwei Liu (рис. 2.16), Yan Liu (разрешение на публикацию рис. 1.6), Rocío Lizárraga (разрешение на публикацию рис. 5.2, нижний ряд), Peter Meer (замечания к разделу 2.4.2), James Milburn (замечания к разделу Sect. 4.4), Pedro Real (замечания на темы геометрии и топологии), Mahdi Rezaei (вклад в материал о распознавании лиц в главе 10, включая текст, рисунки и упражнение 10.2), Bodo Rosenhahn (рис. 7.9 справа), John Rugis (определение кривизны подобия и упражнения 7.2 и 7.6), James Russell (вклад в раздел 5.1.1), Jorge Sanchez (вклад в пример 9.1, рис. 9.1 справа и 9.5), Konstantin Schauwecker (замечания о детекторах признаков и обнаружении плоскостей методом RANSAC, рис. 6.10 справа, 7.19, 9.9 и 2.23), Karsten Scheibe (вклад в главу 6, особенно в раздел 6.1.4, и рис. 7.1), Karsten Schlüns (вклад в раздел 7.4).

- **Sh-Z**

Vok-Suk Shin (предложения по редактированию в Latex, замечания к различным частям книги, вклад в разделы 3.4.1 и 5.1.1 и рис. 9.23 с относящимися к нему комментариями), Eric Song (рис. 5.6 слева), Zijiang Song (вклад в главу 9, особенно в раздел 9.2.4), Kathrin Spiller (вклад в изложение трехмерного случая в разделе 7.2.2), Junli Tao (вклад в материал об обнаружении пешеходов в главе 10, включая текст, рисунки, упражнение 10.1 и замечания о структуре этой главы), Akihiko Torii (вклад в раздел 6.1.4), Johan Van Horebeek (замечания к главе 10), Tobi Vaudrey (вклад в раздел 2.3.2 и рис. 4.18, вклад в раздел 9.3.4, упражнение 9.6), Mou Wei (замечания к главе 4), Shou-Kang Wei (совместная работа над вопросами, относящимися к разделу 6.1.4), Tiangong Wei (вклад в раздел 7.4.3), Jürgen Wiest (рис. 9.1 слева), Yihui Zheng (вклад в раздел 5.1.1), Zezhong Xu (вклад в раздел 3.4.1 и рис. 3.40), Shenghai Yuan (замечания к разделам 3.3.1 и 3.3.2), Qi Zang (упражнение 5.5, рис. 2.21, 5.37 и 10.1), Yi Zeng (рис. 9.15), Joviša Žunić (вклад в раздел 3.3.2).

Автор особенно признателен *Сандино Моралесу* (Sandino Morales) (федеральный округ Мехико) за реализацию и тестирование алгоритмов, предоставление многих рисунков, вклад в главы 4 и 8 и многочисленные замечания, касающиеся различных частей книги, *Владиславу Скарбеку* (Władysław Skarbek) (Варшава, Польша) за разнообразные предложения по улучшению содержания книги и за предложенные упражнения 1.9, 2.10, 2.11, 3.12, 4.11, 5.7, 5.8, 6.10 и *Гарри Ту* (Garry Tee) (Окленд, Новая Зеландия) за тщательную редактуру, замечания, части врезки 5.9, сноску на стр. 480 и еще много ценных советов.

Я благодарю свою жену, *Гизелу Клетте* (Gisela Klette), за написание раздела 3.2.4 о дистанционном евклидовом преобразовании и за критические замечания о структуре и деталях книги во время работы над ней в Центре исследований и математики (СИМАТ) в Гуанахуато в период между серединой июля и началом ноября 2013 г. в период творческого отпуска, предоставленного Оклендским университетом (Новая Зеландия).

Рейнхард Клетте

Гуанахуато, Мексика
3 ноября 2013 г.

Обозначения

$ S $	мощность множества S
$\ \mathbf{a}\ _1$	норма L_1
$\ \mathbf{a}\ _2$	норма L_2
\wedge	логическое «И»
\vee	логическое «ИЛИ»
\cap	пересечение множеств
\cup	объединение множеств
\square	конец доказательства
a, b, c	вещественные числа
A	множество смежности
$\mathcal{A}(\cdot)$	площадь измеримого множества (как функция)
$\mathbf{a}, \mathbf{b}, \mathbf{c}$	векторы
$\mathbf{A}, \mathbf{B}, \mathbf{C}$	матрицы
α, β, γ	углы
b	длина базы в системе стереокамер
\mathbb{C}	множество комплексных чисел вида $a + i \cdot b$, где $i = \sqrt{-1}$, $a, b \in \mathbb{R}$
d	диспаратность
d_1	метрика L_1
d_2	метрика L_2 , известная также как евклидова
e	постоянная $e = \exp(1) \approx 2,7182818284$
ε	вещественное число, большее 0
f	фокусное расстояние
f, g, h	функции
G_{\max}	максимальная яркость в полутоновом изображении
γ	кривая в евклидовом пространстве (например, прямая линия, ломаная или гладкая кривая)
\mathbf{H}	матрица Гессе (гессиан)
i, j, k, l, m, n	натуральные числа; координаты пикселей (i, j) в окне
$I, I(\cdot, \cdot, t)$	изображение, один из последовательных кадров, кадр в момент времени t
L	длина (вещественное число)

$\mathcal{L}(\cdot)$	длина спрямляемой кривой (функция)
λ	вещественное число; значение по умолчанию от 0 до 1
n	натуральное число
N	окрестность (в сетке, наложенной на изображение)
N_{cols}, N_{rows}	количество столбцов, количество строк
\mathbb{N}	множество натуральных чисел $\{0, 1, 2, \dots\}$
$\mathcal{O}(\cdot)$	асимптотическая верхняя граница
Ω	носитель изображения (image carrier), множество всех позиций пикселей $N_{cols} \times N_{rows}$
p, q	точки в \mathbb{R}^2 с координатами x и y
P, Q, R	точки в \mathbb{R}^3 с координатами X, Y, Z
π	вещественная постоянная $\pi = 4 \times \arctan(1) \approx 3,14159265358979$
Π	многогранник
r	радиус окружности или сферы; точка в \mathbb{R}^2 или \mathbb{R}^3
\mathbb{R}	множество вещественных чисел
\mathbf{R}	матрица поворота
ρ	путь с конечным числом вершин
s	точка в \mathbb{R}^2 или \mathbb{R}^3
S	множество
t	момент времени; точка в \mathbb{R}^2 или \mathbb{R}^3
\mathbf{t}	вектор параллельного переноса
T, τ	пороговое значение (вещественное число)
u, v	компоненты оптического потока; вершины или узлы; точки в \mathbb{R}^2 или \mathbb{R}^3
\mathbf{u}	вектор оптического потока $\mathbf{u} = (u, v)$
W, W_p	окно внутри изображения, окно с начальным пикселем p
x, y	вещественные переменные; координаты пикселя изображения (x, y)
X, Y, Z	координаты в \mathbb{R}^3
\mathbb{Z}	множество целых чисел

Глава 1

Данные изображения

В этой главе мы познакомимся с базовыми обозначениями и математическими понятиями для описания изображения на регулярной сетке в пространственной или частотной области. Здесь же мы опишем способы задания цвета и познакомимся с цветными изображениями.

1.1. Изображения в пространственной области

Цифровое *изображение* определяется посредством *дискретизации* непрерывных аналоговых данных в пространственной области. Оно состоит из прямоугольного массива *пикселей* (x, y, u) , каждый из которых является комбинацией местоположения $(x, y) \in \mathbb{Z}^2$ и значения u , представляющего *отсчет* в точке (x, y) . Здесь \mathbb{Z} – множество всех целых чисел. Точки $(x, y) \in \mathbb{Z}^2$ образуют *регулярную сетку*. Формально изображение I определено на прямоугольном множестве Ω , которое называется *носителем* (carrier) I :

$$\Omega = \{(x, y) : 1 \leq x \leq N_{cols} \wedge 1 \leq y \leq N_{rows}\} \subset \mathbb{Z}^2 \quad (1.1)$$

и содержит узлы сетки, или местоположения пикселей для $N_{cols} \geq 1$ и $N_{rows} \geq 1$.

Мы предполагаем левостороннюю систему координат, показанную на рис. 1.1. *Строка* y содержит узлы сетки $\{(1, y), (2, y), \dots, (N_{cols}, y)\}$ для $1 \leq y \leq N_{rows}$, а *столбец* x – узлы сетки $\{(x, 1), (x, 2), \dots, (x, N_{rows})\}$ для $1 \leq x \leq N_{cols}$.

В этом разделе мы начнем знакомство с предметом обработки цифровых изображений с обсуждения способов представления и описания данных изображения в пространственной области, определяемой носителем Ω .

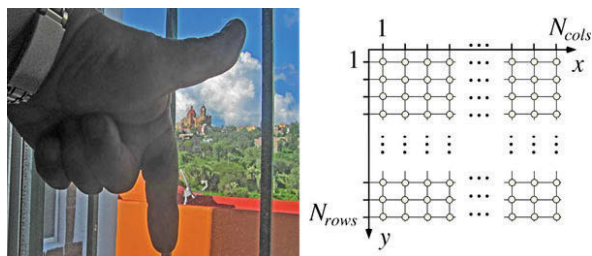


Рис. 1.1. Левосторонняя система координат. Большой палец указывает положительное направление оси x , указательный палец – положительное направление оси y , если смотреть с внутренней стороны ладони. (На левом рисунке видна также барочная церковь в городке Валенсиана, которая всегда была перед глазами автора, когда он писал эту книгу в Центре исследований и математики в Гуанахуато)

1.1.1. Пиксели и окна

На рис. 1.2 показаны два способа геометрической интерпретации пикселей, т. е. отборов изображения на регулярной сетке.

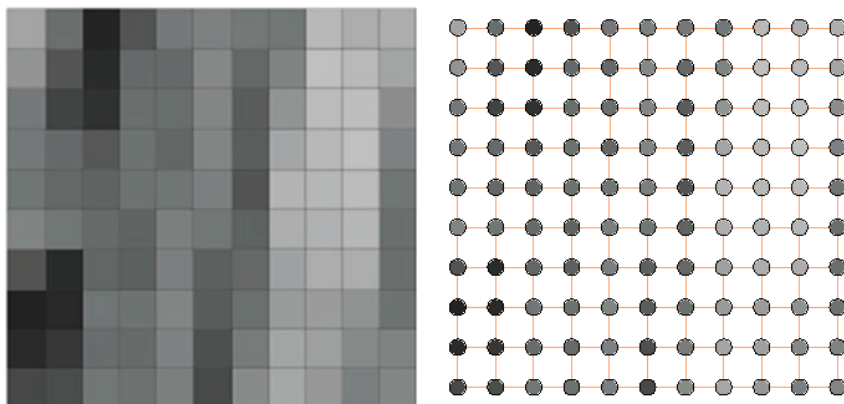


Рис. 1.2. Слева. На увеличенном изображении видны ячейки сетки; разными оттенками серого представлены значения пикселей изображения. Справа. Можно также считать, что значения изображения – это метки в узлах сетки, являющихся центрами ячеек сетки

Ячейки сетки, узлы сетки и смежность. Изображения, которые мы видим на экране, состоят из ячеек, покрашенных одним цветом. Исходя из такого представления, мы можем считать пиксель крохотным покрашенным квадратиком. Это так называемая *модель ячеек сетки*. Но можно также рассматривать пиксель как узел сетки, помеченный значением изображения. Эта *модель узлов сетки* уже была показана на рис. 1.1.

Врезка 1.1 (происхождение термина «пиксель»). Термин «пиксель» – сокращение от «picture element» (элемент изображения). Он был введен в оборот в конце 1960-х гг. группой исследователей из Лаборатории реактивного движения в Пасадене, штат Калифорния, которая занималась обработкой изображений, снятых космическими аппаратами. См. [R. B. Leighton, N. H. Horowitz, A. G. Herriman, A. T. Young, B. A. Smith, M. E. Davies and C. B. Leovy. Mariner 6 television pictures: First report. Science, 165:684–690, 1969].

Пиксели – «атомарные элементы» изображения. Сами по себе они не определяют никаких отношений смежности. В модели ячеек сетки мы можем предполагать, что пиксели являются смежными тогда и только тогда, когда соответствующие им закрашенные квадратики имеют общую сторону. Но можно также считать пиксели смежными, если они различны и их закрашенные квадратики имеют хотя бы одну общую точку (сторону или вершину).

Окна изображения. Окном $W_p^{m,n}(I)$ называется часть изображения I , имеющая размер $m \times n$ и позиционированная относительно *начальной точки* p (т. е. некоторого пикселя). По умолчанию подразумевается, что $m = n$ – нечетное число, а p – центр окна. На рис. 1.3 показано окно $W_{(453,134)}^{77,73}$ (SanMiguel).

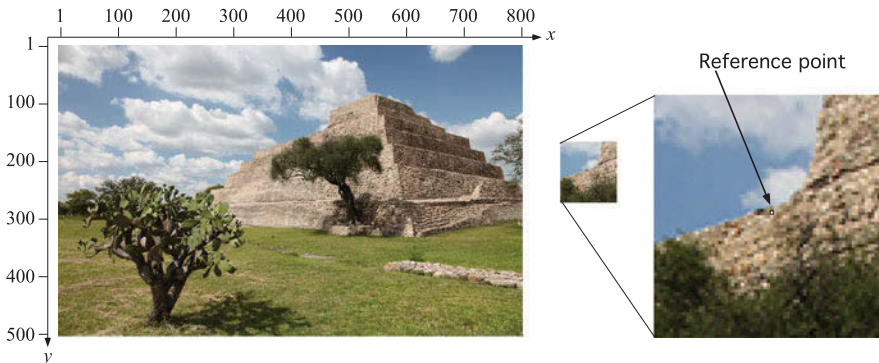


Рис. 1.3. Окно 73×77 изображения SanMiguel. На изображении главной пирамиды музея Каньяда-де-ла-Вирген в Мексике помечен начальный пиксель с координатами $p = (453, 134)$

Обычно мы упрощаем нотацию и пишем просто W_p , поскольку само изображение и размер окна известны из контекста.

1.1.2. Значения и основные статистики изображения

Значения изображения u берутся из дискретного множества возможных значений. В компьютерном зрении часто рассматривают отрезок вещественной оси $[0, 1] \subset \mathbb{R}$ как область значений скалярного изображения. Это особенно полезно, если значения изображения интерполируются выполняемыми процедурами и представляются типом данных REAL. В этой книге мы по умолчанию считаем, что значениями изображения являются целые числа.

Скалярные и бинарные изображения. Значениями скалярного изображения являются целые числа: $u \in \{0, 1, \dots, 2^a - 1\}$. Принято интерпретировать такие скалярные значения как уровни яркости, когда 0 соответствует черному цвету, $2^a - 1$ – белому, а остальные уровни линейно интерполируются между черным и белым. В таком случае мы говорим, что изображение *полутонное*. На протяжении многих лет было принято использовать $a = 8$, но недавно техническим стандартом стало $a = 16$. Чтобы не зависеть от таких деталей, мы будем обозначать $G_{\max} = 2^a - 1$.

В *бинарном изображении* пиксели могут принимать только два значения, традиционно обозначаемые 0 = белый и 1 = черный, т. е. черные объекты размещены на белом фоне.

Векторные и RGB-изображения. В векторном изображении, в отличие от скалярного, имеется более одного *канала* или *полосы*. Значениями изображения являются векторы $(u_1, \dots, u_{N_{\text{channels}}})$ длины N_{channels} . Например, цветные изображения в широко распространенной цветовой модели RGB имеют три канала, по одному для красной, зеленой и синей компоненты. Значения u_i в каждом канале принадлежат множеству $\{0, 1, \dots, G_{\max}\}$, т. е. каждый канал – это просто полутонное изображение. См. рис. 1.4.

Среднее. Пусть имеется скалярное изображение I размера $N_{\text{cols}} \times N_{\text{rows}}$. Вслед за математической статистикой мы называем *средним* (средним уровнем яркости) изображения I величину

$$\begin{aligned} \mu_I &= \frac{1}{N_{\text{cols}} \cdot N_{\text{rows}}} \sum_{x=1}^{N_{\text{cols}}} \sum_{y=1}^{N_{\text{rows}}} I(x, y) = \\ &= \frac{1}{|\Omega|} \sum_{(x,y) \in \Omega} I(x, y) \end{aligned} \quad (1.2)$$

где $|\Omega| = N_{\text{cols}} \cdot N_{\text{rows}}$ – мощность носителя Ω , содержащего все пиксели. Мы предпочитаем второй способ записи. В этой формуле мы используем I ,

а не u ; I – это отображение, определенное на Ω , а u мы оставляем для значений отдельных пикселей.

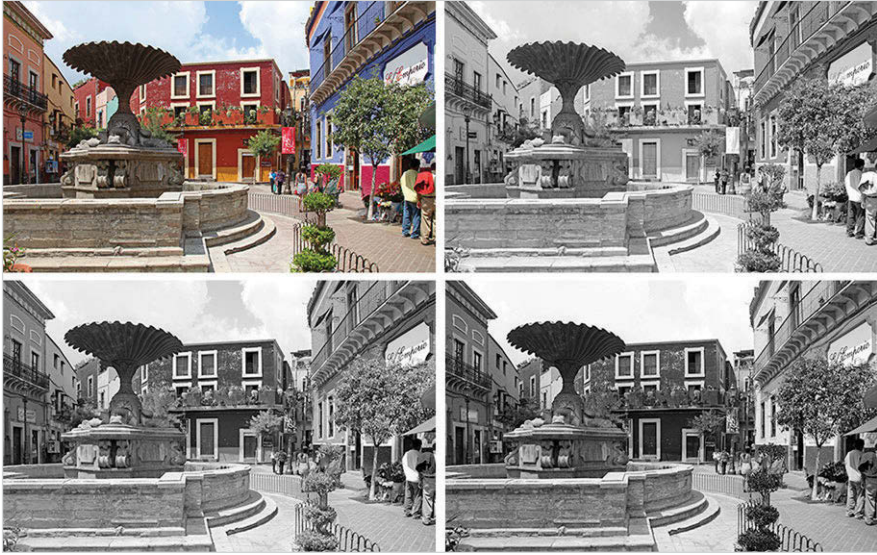


Рис. 1.4. Исходное цветное RGB-изображение Fountain (слева сверху), на котором запечатлена площадь в Гуанахуато, и его разложение на три канала: красный (справа сверху), зеленый (слева снизу) и синий (справа снизу).

Красному цвету соответствует высокая яркость красного канала и одновременно низкая яркость зеленого и синего каналов

Дисперсия и стандартное отклонение. Дисперсией изображения I называется величина

$$\sigma_i^2 = \frac{1}{|\Omega|} \sum_{(x,y) \in \Omega} [I(x,y) - \mu_i]^2. \quad (1.3)$$

Квадратный корень из нее σ_i называется *стандартным отклонением* изображения I .

Применимы такие хорошо известные формулы статистики, как

$$\sigma_i^2 = \left[\frac{1}{|\Omega|} \sum_{(x,y) \in \Omega} I(x,y)^2 \right] - \mu_i^2. \quad (1.4)$$

Формула (1.4) позволяет вычислять среднее и дисперсию путем однократного прохода по изображению. Если бы мы использовали определения (1.2) и (1.3), то понадобилось бы два прохода: один для вычисления среднего и второй для вычисления дисперсии.

Гистограммы. Гистограмма представляет таблицу частот, обычно в виде столбиков на графике. С помощью гистограмм представляются частоты значений скалярного изображения или одного из каналов векторного изображения.

Пусть имеется скалярное изображение I с пикселями (i, j, u) , где $0 \leq u \leq G_{\max}$. Мы определяем *абсолютные частоты*, подсчитывая, сколько раз значение u встречается в носителе Ω , содержащем все пиксели:

$$H_I(u) = |\{(x, y) \in \Omega : I(x, y) = u\}|, \quad (1.5)$$

где $|\cdot|$ обозначает мощность множества. *Относительные частоты* – значения между 0 и 1 – можно сравнить с функцией плотности вероятности (ФПВ) распределения дискретной случайной величины $I(p)$. Они вычисляются по формуле

$$h_I(u) = \frac{H_I(u)}{|\Omega|}. \quad (1.6)$$

Значения $H_I(0), H_I(1), \dots, H_I(G_{\max})$ определяют (абсолютную) гистограмму уровней яркости скалярного изображения I . На рис. 1.5 показаны гистограммы исходного изображения и трех его модифицированных вариантов.

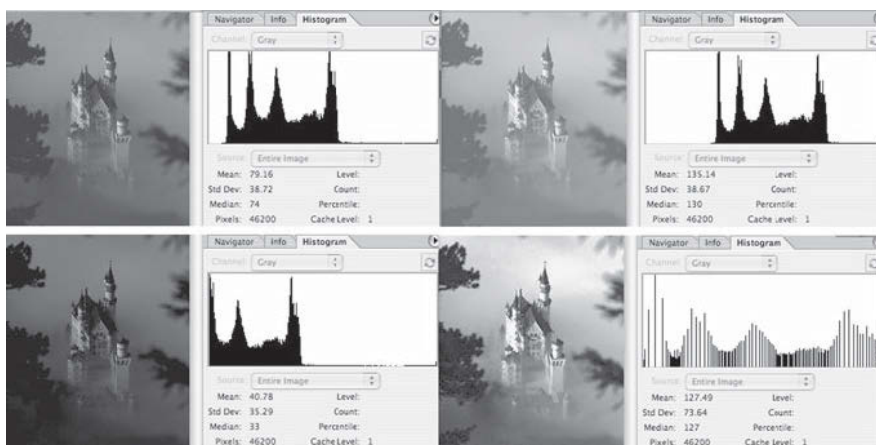


Рис. 1.5. Гистограммы для изображения Neuschwanstein размера 200×231 .

Слева вверху: исходное изображение. *Справа вверху:* более яркий вариант. *Слева внизу:* более темный вариант. *Справа внизу:* после выравнивания гистограммы (объяснение см. ниже)

Мы можем также вычислить среднее и дисперсию на основе относительных частот:

$$\mu_I = \sum_{u=0}^{G_{\max}} u \cdot h_I(u) \quad \text{или} \quad \sigma_I^2 = \sum_{u=0}^{G_{\max}} [u - \mu_I]^2 \cdot h_I(u). \quad (1.7)$$

Так получается быстрее, если гистограмма уже вычислена.

Абсолютная и относительная кумулятивная частота определяются следующим образом:

$$C_I(u) = \sum_{v=0}^u H_I(v) \quad \text{и} \quad c_I(u) = \sum_{v=0}^u h_I(v). \quad (1.8)$$

Эти значения показываются на кумулятивных гистограммах. Относительные частоты можно сравнить с функцией вероятности $\Pr[I(p) \leq u]$ дискретной случайной величины $I(p)$.

Статистики значений в окне. Рассмотрим окно по умолчанию $W = W_p^{n,n}(I)$, где $n = 2k + 1$ и $p = (x, y)$. Тогда в координатах окна имеем

$$\mu_W = \frac{1}{n^2} \sum_{i=-k}^{+k} \sum_{j=-k}^{+k} I(x+i, y+j). \quad (1.9)$$

См. рис. 1.6. Формулы для дисперсии и других показателей можно модифицировать аналогично.



Рис. 1.6. Пример двух окон размера 104×98 в изображении Yan и соответствующих им гистограмм.

Верхнее окно: $\mu_{w_1} = 133,7, \sigma_{w_1} = 55,4$. Нижнее окно: $\mu_{w_2} = 104,6, \sigma_{w_2} = 89,9$

Пример 1.1 (примеры окон и гистограмм). В изображении Yan размера 489×480 , показанном на рис. 1.6, мы видим два выделенных рамкой окна раз-

мера 104×98 : W_1 , содержащее лицо, и W_2 , содержащее части скамьи и платья. Там же приведены гистограммы для обоих окон.

Трехмерное представление уровней яркости (которые в таком случае интерпретируются как возвышения) иллюстрирует различные «степени однородности» изображения. Пример см. на рис. 1.7. Крутой подъем с нижнего плато на верхнее на левом рисунке – типичная иллюстрация «границы» внутри изображения.

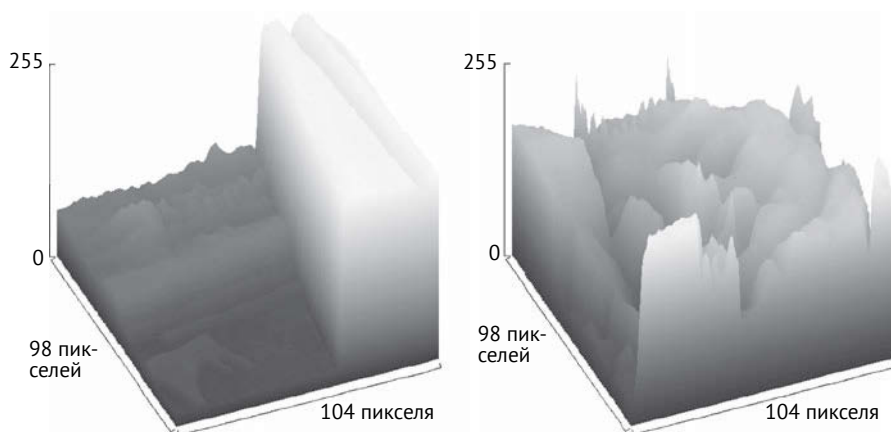


Рис. 1.7. Слева: «крутой переход от темного к яркому».

Справа: «незначительное» изменение. Обратите внимание на разные масштабы трехмерных представлений обоих окон на рис. 1.6

В процессе анализа изображения мы должны классифицировать окна по категориям, например: «внутри однородной области», «низкая контрастность», «содержит границу между двумя разными областями» или «высокая контрастность». Мы определяем *контрастность* $C(I)$ изображения I как среднюю абсолютную величину разности между значением пикселя и средним значением соседних пикселей.

$$C(I) = \frac{1}{|\Omega|} \sum_{(x,y) \in \Omega} |I(x,y) - \mu_{A(x,y)}|, \quad (1.10)$$

где $\mu_{A(x,y)}$ – среднее значение, вычисленное по пикселям, соседним с находящимся в позиции (x, y) .

Еще один пример использования низкоуровневых статистик для простой интерпретации изображений показан на рис. 1.4. Средние значения, посчитанные по красному, зеленому и синему каналам, показывают, что в изображении больше красного цвета (правый верхний рисунок, среднее значение равно 154) и меньше зеленого (левый нижний

рисунок, среднее значение равно 140) и синего (правый нижний рисунок, среднее значение равно 134). Это можно проверить более точно, взглянув на гистограммы всех трех каналов. На них видно, что «более ярким» является красный канал, особенно в области, занятой домом в центре, а зеленый и синий каналы в той же области «более темные».

1.1.3. Пространственные и временные меры данных

Уже приведенных выше базовых понятий из математической статистики достаточно, чтобы определить функции, описывающие изображение: построчно в случае одного изображения или покадрово в случае последовательности изображений.

Статистики значений в профиле яркости. При первоначальном анализе изображений в новой предметной области можно получить весьма полезную информацию, если визуализировать *профили яркости*, определяемые одномерными сечениями скалярных массивов данных.

На рис. 1.8 показаны два профиля яркости по оси x , построенных для заданного полутонового изображения. Как и раньше, мы можем использовать среднее, дисперсию и гистограммы «узких» окон размера $N_{cols} \times 1$, чтобы получить общее представление о распределении значений изображения.

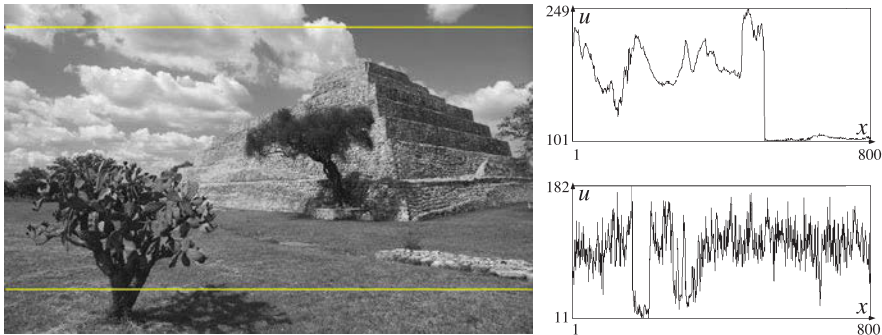


Рис. 1.8. Слева: выделены две строки изображения для представления в канале яркости (т. е. значения $(R + G + B)/3$) для изображения SanMiguel на рис. 1.3. Справа: профили яркости обеих выбранных строк

Пространственные и временные статистики значений. Гистограммы и профили яркости – примеры *пространственных статистик значений*. Так, профили яркости для строк с номерами от 1 до N_{rows} одного изображения I определяют последовательность дискретных функций,

которую можно сравнить с соответствующей последовательностью для другого изображения J .

Рассмотрим еще один пример – последовательность изображений, состоящую из кадров I_t , $t = 1, 2, \dots, T$, определенных на одном и том же носителе Ω . Для лучшего понимания распределения значений полезно определить скалярную меру данных (t), которая сопоставляет одному кадру I_t одно число, и сравнить различные меры данных для заданного дискретного временного интервала $[1, 2, \dots, T]$, получив тем самым *временные статистики данных*.

Например, определение контрастности (1.10) индуцирует меру данных $\mathcal{P}(t) = C(I_t)$, определение среднего (1.2) – меру данных $\mathcal{M}(t) = \mu_{I_t}$, а определение дисперсии (1.3) – меру данных $\mathcal{V}(t) = \sigma_{I_t}^2$.

На рис. 1.9 сверху показаны две меры данных на последовательности 400 изображений. (В данном контексте не важно, какие именно изображения и меры данных использовались.) Области значений разных мер различаются и могут быть охарактеризованы средним и дисперсией. Для более точного сравнения мы преобразуем обе меры в функции, имеющие одинаковые среднее и дисперсию.

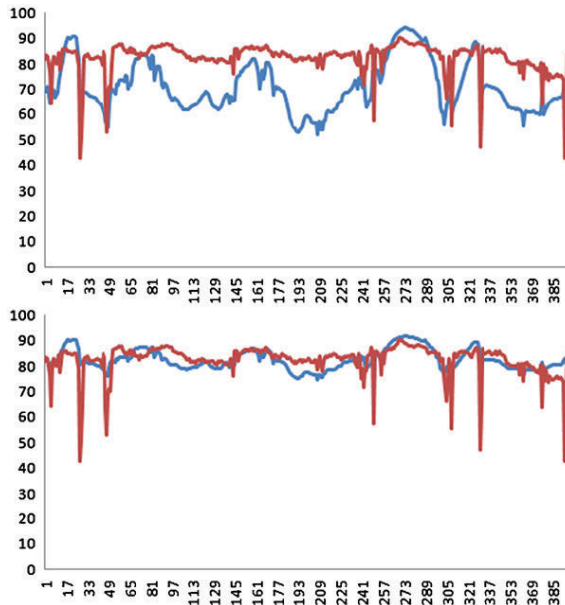


Рис. 1.9. Сверху: график двух мер данных для последовательности 400 кадров.
Снизу: те же меры после нормировки среднего и дисперсии

Нормировка двух функций. Обозначим μ_f и σ_f среднее и стандартное отклонение функции f . Пусть даны две вещественные функции f и g с одной и той

же дискретной областью определения, скажем $1, 2, \dots, T$, и ненулевыми дисперсиями. Положим

$$\begin{aligned} \alpha &= \frac{\sigma_g}{\sigma_f} \mu_f - \mu_g, \\ \beta &= \frac{\sigma_f}{\sigma_g} \end{aligned} \quad \text{и} \quad (1.11)$$

$$g_{new}(x) = \beta(g(x) + \alpha). \quad (1.12)$$

Тогда функция g_{new} будет иметь такие же среднее и дисперсию, как функция f .

Расстояние между двумя функциями. Теперь определим расстояние между двумя вещественными функциями с одной и той же областью определения, например $1, 2, \dots, T$:

$$d_1(f, g) = \frac{1}{T} \sum_{x=1}^T |f(x) - g(x)|, \quad (1.13)$$

$$d_2(f, g) = \frac{1}{T} \sqrt{\sum_{x=1}^T (f(x) - g(x))^2}. \quad (1.14)$$

Оба расстояния являются *метриками*, т. е. для них выполняются следующие аксиомы:

- 1) $f = g$ тогда и только тогда, когда $d(f, g) = 0$;
- 2) $d(f, g) = d(g, f)$ (симметричность);
- 3) $d(f, g) \leq d(f, h) + d(h, g)$ для произвольной третьей функции h (неравенство треугольника).

Структурное подобие мер данных. Пусть заданы две разные пространственные или временные меры данных \mathcal{F} и \mathcal{G} с одной и той же областью определения $1, 2, \dots, T$. Сначала отобразим \mathcal{G} в \mathcal{G}_{new} , так чтобы у обеих мер были одинаковые среднее и дисперсия, а затем вычислим расстояние между \mathcal{F} и \mathcal{G}_{new} , применяя любую из метрик L_1 или L_2 .

Две меры \mathcal{F} и \mathcal{G} называются *структурно подобными*, если расстояние между \mathcal{F} и \mathcal{G}_{new} близко к нулю. Структурно подобные меры принимают локальные максимумы и минимумы приблизительно в одних и тех же точках.

1.1.4. Ступенчато-граничная модель

Изломы в изображениях – это признаки, которые часто бывают полезны на начальной стадии процедуры анализа. Границы представляют собой важную информацию, позволяющую лучше понять изображение (например, устранить влияние переменного освещения). Удалив все данные, кроме границ, мы сможем упростить данные. На рис. 1.10 понятие «границы» иллюстрируется тремя примерами.



Рис. 1.10. Границы, или визуальные силуэты, уже много тысяч лет используются, чтобы показать «существенную информацию», например в пещерных рисунках. Слева: изображение Taroko, на котором показаны древние рисунки коренного населения Тайваня. В середине: фрагмент изображения Aussies с силуэтами, спроецированными на стену здания Q1 в городе Голд-Кост, Австралия. Справа: торговый центр в Шанхае, изображение OldStreet

Изломы в изображениях могут быть локализованы в небольших окнах (например, зашумленные пиксели) или определять границы между областями с разными характеристиками сигнала.

Что такое граница? На рис. 1.11 показано разнообразие границ в изображениях. Для этого мы двигаемся слева направо по профилю яркости в соответствии со ступенчато-граничной моделью (step-edge model). В этой модели предполагается, что границы определяются изменениями локальных производных. Альтернативную фазово-конгруэнтную модель мы обсудим в разделе 1.2.5.

После удаления шума предположим, что изображение представляет отсчеты непрерывной функции $I(x, y)$, определенной на евклидовой плоскости \mathbb{R}^2 , для которой можно взять первую и вторую частные производные по x и y . На рис. 1.12 иллюстрируются свойства таких производных.