

Автономная некоммерческая образовательная организация высшего
профессионального образования «Сколковский институт науки и технологий»

На правах рукописи



Кузьмин Андрей Игоревич

**МЕТОДЫ ОБУЧАЕМОЙ РЕГУЛЯРИЗАЦИИ В ЗАДАЧАХ
ПЛОТНОГО СОПОСТАВЛЕНИЯ ИЗОБРАЖЕНИЙ**

Специальность 05.13.17 —

«Теоретические основы информатики»

Диссертация на соискание учёной степени

кандидата технических наук

Научный руководитель:

к.ф.-м.н., доцент

Лемпицкий Виктор Сергеевич

Москва — 2018

Оглавление

1	Введение	1
2	Задача плотного сопоставления изображений	8
2.1	Постановка задачи	8
2.2	Бинокулярное стерео-сопоставление	10
2.3	Оптический поток	12
2.4	Многокадровая ультразвуковая эластография	13
2.5	Приложения	14
2.5.1	Обнаружение препятствий на дорожных сценах	14
2.5.2	Оценка скорости движения на дорожных сценах	15
2.5.3	Анализ механических свойств ткани человека для медицинской диагностики	16
3	Бинокулярное стерео-сопоставление	18
3.1	Введение в стерео-реконструкцию	18
3.1.1	Эпиполярная геометрия	18
3.1.2	Задача бинокулярного стерео-сопоставления	19
3.1.3	Особенности задачи стерео-сопоставления для реальных данных	20
3.2	Классификация методов стерео-сопоставления	21
3.2.1	Вычисление тензора энергий	21
3.2.2	Агрегирование тензора энергий	21
3.2.3	Вычисление диспаратета как задача оптимизации	22
3.2.3.1	Метод динамического программирования	22
3.2.3.2	Методы на основе случайных Марковских полей	23
3.2.4	Постобработка диспаратета	24
3.2.5	Локальные и глобальные методы стерео-сопоставления	24
3.3	Применение фильтров, учитывающих границы объектов на изображении	25
3.3.1	Билатеральный фильтр	25
3.3.2	Управляемый фильтр	26
3.3.3	Рекурсивный фильтр, учитывающий границы	27
3.3.4	Метод нелокальных средних	28
3.4	Обзор методов, основанных на глубоком машинном обучении	29
3.4.1	Метод обратного распространения ошибки	29
3.4.2	Обучение глубоких дескрипторов в задачах стерео-сопоставления	30
3.4.3	Обучаемая регуляризации на основе условных случайных полей	31

3.4.4	Функции потерь в обучаемых моделях стерео-сопоставления . . .	33
3.5	Обучаемое агрегирование тензора энергий сопоставления на основе сверточно-рекуррентной нейросети.	37
3.5.1	Архитектура нейросети	37
3.5.1.1	Вычисление энергий стерео-сопоставления	37
3.5.1.2	Детектор границ объектов на изображении	39
3.5.1.3	Сглаживание при помощи рекурсивного фильтра, учитывающего границы изображения	40
3.5.1.4	Агрегирование тензора энергий	42
3.5.1.5	Функция потерь	44
3.5.2	Численные эксперименты	45
3.5.2.1	Описание тренировочной выборки.	45
3.5.2.2	Методика оценки ошибки метода.	45
3.5.2.3	Процесс обучения.	46
3.5.2.4	Оценка количества арифметических операций.	52
3.5.2.5	Время исполнения на графическом ускорителе.	53
3.5.2.6	Обнаружение заслоненных объектов.	53
3.6	Выводы	54
4	Задача вычисления оптического потока	56
4.1	Постановка задачи	56
4.2	Обзор методов на основе минимизации функционалов энергии	57
4.2.1	Метод на основе инвариантной интенсивности и линеаризации изображения	57
4.2.2	Многомасштабная оценка оптического потока	59
4.2.3	Методы регуляризации.	60
4.2.3.1	Полная вариация	60
4.2.3.2	Двойственный градиентный метод для задачи шумоподавления	61
4.2.3.3	Обобщение полной вариации и условие гладкости второго порядка	62
4.2.3.4	Модель для обучаемой регуляризации на основе случайных марковских полей	63
4.2.4	Метод для оценки больших смещений	64
4.2.5	Интерполяция соответствий с учетом границ изображения	64
4.3	Обзор методов, основанных на глубоком машинном обучении	65
4.3.1	Сверточная нейросеть для задачи оптического потока и модель FlowNet	65
4.4	Обучаемая регуляризация для метода вычисления оптического потока в реальном времени	66
4.4.1	Двойственный метод оптимизации для вычисления оптического потока в реальном времени	67
4.5	Архитектура нейросети.	70
4.5.1	Численные эксперименты.	76
4.6	Многокадровые методы оценки оптического потока	77
4.6.1	Пространственно-временная фильтрация оптического потока	78
4.6.2	Метод, основанный на нахождении траекторий малого ранга	78
4.7	Выводы	80

5	Сопоставление медицинских ультразвуковых изображений	81
5.1	Задача ультразвуковой эластографии	81
5.2	Обзор методов эластографии, основанных на двух кадрах	82
5.3	Обзор многокадровых методов эластографии	84
5.3.1	Метод ElastMI	84
5.3.2	Видеопоследовательность ультразвуковых снимков и видео-эластограмма	84
5.4	Метод регуляризованной многокадровой эластографии	85
5.4.1	Описание метода	85
5.4.1.1	Оценка поля смещения	87
5.4.1.2	Оценка механической деформации	89
5.4.1.3	Функционал.	89
5.4.1.4	Регуляризация.	91
5.4.1.5	Минимизация.	92
5.4.1.6	Реализация с использованием графических ускорителей.	93
5.4.2	Численные эксперименты	94
5.5	Сопоставление сигналов в ультразвуковой томографии	106
5.5.1	Постановка задачи	107
5.5.2	Описание метода	108
5.5.2.1	Анализ углового распределения	108
5.5.2.2	Метод реконструкции	109
5.5.3	Численные эксперименты	110
5.5.4	Эксперименты на синтетических данных	110
5.5.5	Эксперименты на реальных данных	111
5.5.6	Оценка вычислительной сложности	111
5.6	Выводы	112
6	Заключение	114
	Литература	117

Глава 1

Введение

Актуальность темы. Задача сопоставления изображений является одной из наиболее важных задач в компьютерном зрении, которая возникает во многих практических приложениях, таких как бинокулярная стерео-реконструкция [1], детекция движения на видеопоследовательностях [2–4] и анализ медицинских ультразвуковых снимков [5–7].

В общем случае задача сопоставления изображений допускает несколько различных постановок, применимость которых зависит от конкретного приложения. Параметрическое сопоставление изображений представляет собой задачу поиска трансформации внутри выбранного семейства параметрических преобразований, таких как, например, аффинные преобразования, которые позволяют сопоставить изображения с учетом перспективных искажений. В данной работе рассмотрена задача непараметрического сопоставления изображений. Такой вариант задачи является наиболее общим: каждый пиксел изображения получает независимую трансформацию, при этом число степеней свободы пропорционально числу пикселов.

Другим важным аспектом постановки задачи является способ сопоставления изображений [1]. Первым важным случаем является разреженное сопоставление, при котором соотносятся отдельные визуально выделяющиеся элементы изображений. Вторым важным случаем, рассмотренным в данной работе, является плотное сопоставление. При этом соотносятся все пиксели изображения, и решением задачи является двумерное поле смещений. Такое поле определяет трансформацию для каждого из пикселов изображения.

Наиболее важными характеристиками методов сопоставления являются вычислительная сложность и качество сопоставления на реальных данных. В настоящее

время, наиболее перспективными методами сопоставления являются методы, основанные на глубоком машинном обучении [3, 8, 9]. При этом существенным недостатком большинства таких методов является высокая вычислительная сложность, что не позволяет применять их в задачах, требующих сопоставления в реальном времени [9, 10], таких как анализ дорожных сцен и медицинская диагностика в режиме реального времени (с частотой порядка 25 кадров в секунду и выше). В связи с этим, особый интерес представляет разработка методов машинного обучения, имеющих низкую вычислительную сложность на этапе исполнения [11, 12].

Большинство современных методов сопоставления изображений можно разделить на две категории. К первой категории относятся методы, основанные на глубоком машинном обучении [3, 9, 11, 13]. Для таких методов применяется обучение с учителем на большом количестве тренировочных данных. Методы второй категории основаны на формулировке задачи сопоставления изображений в виде оптимизационной задачи, при этом поле смещений получается в результате минимизации целевого функционала, зависящего от входных данных [7, 14]. Такой подход может быть применен в случае отсутствия тренировочных данных и является актуальным, например, для сопоставления медицинских ультразвуковых изображений - в этом случае трудно получить эталонные поля смещений.

В работе рассмотрена задача сопоставления изображений в трех различных приложениях. Первым является бинокулярная стерео-реконструкция, которая основана на оценке смещений для левого и правого изображений со стерео-камеры, возникающих за счет бинокулярного эффекта [15, 16]. Вторым является сопоставление изображений движущихся объектов на видео-последовательности, известная как задача вычисления оптического потока [1, 3, 17]. Третьим является задача ультразвуковой эластографии [5–7]. Она соответствует сопоставлению медицинских ультразвуковых снимков для тканей различной степени механического сжатия с целью оценки локальной деформации, которая является важной величиной для медицинской диагностики.

Сопоставление изображений в каждом из трех перечисленных выше приложений позволяет количественно оценивать различные свойства объектов на анализируемых изображениях. В задаче стерео-реконструкции, сопоставление позволяет оценить геометрию сцены, в задаче нахождения оптического потока – скорости движущихся объектов, а в задаче эластографии – механические свойства изучаемых тканей.

Целью данной работы является разработка методов машинного обучения для задачи сопоставления изображений, эффективных на этапе исполнения и позволяющих вычислять поля смещений в режиме реального времени (с частотой 25 кадров в

секунду и выше) для реальных данных с использованием параллельного программирования.

Для достижения поставленной цели необходимо было решить следующие задачи:

1. Аналитический обзор состояния задачи и систематизация методов сопоставления изображений.
2. Разработка новых методов машинного обучения для сопоставления изображений, имеющих низкую вычислительную сложность на этапе исполнения.
3. Экспериментальная проверка разработанных методов на реальных данных, сравнение результатов с предложенными в литературе методами с использованием количественных критериев качества сопоставления.
4. Программная реализация предложенных методов с использованием графических ускорителей, позволяющая вычислять поля смещений в реальном времени.

Основные положения, выносимые на защиту:

1. Предложена серия моделей для сопоставления изображений, имеющих низкую вычислительную сложность на этапе исполнения. В качестве основы для построения вычислительно эффективных моделей была выбрана обучаемая регуляризация. Этапы вычисления полей смещения были представлены как слои сверточной и рекуррентной нейросети, что позволило получить обучаемую модель.
2. Показаны результаты применения предложенных методов на реальных данных, включая дорожные сцены и медицинские ультразвуковые снимки. Рассмотрены такие приложения как стерео-реконструкция, оптический поток и ультразвуковая эластография. Проведен количественный анализ качества сопоставления.
3. Предложена эффективная параллелизация разработанных методов. Построен комплекс эффективных параллельных программ с использованием графических ускорителей, демонстрирующих применимость предложенных моделей в режиме онлайн для реальных данных.

Научная новизна:

1. Предложен новый метод сопоставления изображений, используемый в задаче бинокулярной стерео-реконструкции. В отличие от аналогичных подходов, основанных на глубоком машинном обучении и сверточных нейросетях, предложенный метод основан на комбинировании сверточной и рекуррентной нейросети, что позволяет получить алгоритм, эффективный на этапе исполнения, имеющий эффективную параллельную реализацию. Такой подход позволяет избежать трудоемкого сравнения визуальных дескрипторов большой размерности, являющегося ключевым этапом прочих методов стерео-сопоставления, основанных на глубоком машинном обучении.
2. Разработана новая архитектура нейросети для задачи сопоставления изображений, возникающей при вычислении оптического потока. Предложенный метод основан на обучении оператора регуляризации. Подход, основанный на представлении графа вычислений оптимизационного алгоритма в виде слоев сверточной нейросети, позволил получить сверточную архитектуру, имеющую более низкую вычислительную сложность по сравнению с методами, предложенными в литературе. При этом обучаемая регуляризация позволяет получить сопоставления более высокого качества по сравнению с оптимизационными алгоритмами низкой вычислительной сложности, предложенными в литературе.
3. Предложен новый оптимизационный метод для сопоставления ультразвуковых изображений, который позволяет улучшить качество сопоставления за счет использования серии из трех снимков. В отличие от предложенных в литературе методов, предложенный подход основан на применении адаптивной регуляризации, что позволило получить метод, устойчивый к участкам неверного сопоставления, при этом имеющий низкую вычислительную сложность. При этом вычислительная эффективность алгоритма достигается за счет обобщения функционала полной вариации. Предложенный функционал является выпуклым и позволяет применять эффективные двойственные методы минимизации.

Теоретическая значимость заключается в разработке новых моделей для задачи сопоставления изображений. Предложена модель для сопоставления изображений в применении к стерео-реконструкции, основанная на сверточно-рекуррентной нейросети. Такая модель является целиком обучаемой на эталонных данных и позволяет вычислять поля смещения в реальном времени на этапе исполнения. Также предложена модель на основе сверточной нейросети для задачи вычисления оптического

потока, которая позволяет обучать оператор регуляризации. Наконец, автором предложен метод сопоставления ультразвуковых снимков на основе выпуклой оптимизации, который позволяет эффективно вычислять смещения на основе нескольких ультразвуковых снимков.

Практическая значимость работы заключается в возможности решать задачу сопоставления изображений в режиме реального времени на данных соответствующих фотографиям дорожных сцен и медицинским ультразвуковым снимкам. Потенциальные приложения разработанных методов включают в себя системы беспилотного управления автомобилем, а также программное обеспечение, используемое в устройствах ультразвуковой медицинской диагностики.

Разработанный метод сопоставления серии ультразвуковых изображений был внедрен в программный продукт по анализу последовательности медицинских снимков ООО "СиВижинЛаб".

Достоверность полученных результатов обеспечивается серией численных экспериментов, проведенных с использованием открытых коллекций изображений.

Апробация работы. Основные результаты работы докладывались на:

1. Международная конференция “Machine Can See Summit”, 2017.
2. Международная конференция “IEEE Workshop on Machine Learning for Signal Processing”, 2017.
3. Семинар Вычислительного центра им. Дородницына ФИЦ ИУ РАН, 2017.
4. Международная конференция “IEEE 13th International Symposium on Biomedical Imaging (ISBI)”, 2016.
5. Международная конференция “IEEE 37th Annual International Conference on Medicine and Biology Society (EMBC)”, 2015.

Личный вклад. Все результаты получены автором лично.

Публикации. По тематике исследования опубликовано 5 научных работ в изданиях, рекомендованных ВАК:

1. Kuzmin Andrey, Mikushin Dmitry, Lempitsky Victor. End-to-end Learning of Cost-Volume Aggregation for Real-time Dense Stereo // Machine Learning for Signal Processing, 2017. MLSP 2017. IEEE Conference on / IEEE. 2017.

2. Fast low-cost single element ultrasound reflectivity tomography using angular distribution analysis / Andrey Kuzmin, Xiang Zhang, Jonathan Finche [и др.] // Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on / IEEE. 2016. С. 1021–1024.
3. A single element 3D ultrasound tomography system / Xiang Zhang, Jonathan Fincke, Andrey Kuzmin [и др.] // Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE / IEEE. 2015. С. 5541–5544.
4. Multi-frame elastography using a handheld force-controlled ultrasound probe / Andrey Kuzmin, Aaron M Zakrzewski, Brian W Anthony [и др.] // IEEE transactions on ultrasonics, ferroelectrics, and frequency control. 2015. Т. 62, № 8. С. 1486–1500.
5. Set2Model networks: Learning discriminatively to learn generative models / Alexander Vakhitov, Andrey Kuzmin, Victor Lempitsky // Computer Vision and Image Understanding. 2017, № 8.

Краткое содержание работы

Во введении обосновывается актуальность проводимых исследований, научная и практическая ценность работы, сформулированы цели и задачи, а также сформулированы основные положения, выносимые на защиту.

Глава 3 начинается с введения в методы стерео-сопоставления (раздел 3.1). Обзор основных этапов стерео-методов, предложенных в литературе содержится в подразделе 3.2. Главу продолжает описание методов фильтрации изображений, использующих нахождение границ объектов (раздел 3.3). Затем, в разделе 3.4 приводится описание методов глубокого машинного обучения для стерео-реконструкции. Главу завершает описание метода, основанное на комбинации сверточной и рекуррентной нейросети, предложенного автором (см. раздел 3.5). Подраздел так же содержит результаты численных экспериментов по сравнению различных методов.

В главе 4 рассматривается задача вычисления оптического потока и описывается метод, предложенный автором (см. раздел 4.4). В начале главы приведен обзор методов, основанных на постановке оптимизационной задачи (раздел 4.2), описывается основной набор практических трудностей, которые встречаются при решении данной задачи на реальных данных. Раздел 4.3 содержит обзор методов вычисления оптического потока, основанных на глубоком машинном обучении. В конце главы

приведен обзор многокадровых методов для оценки оптического потока, приведенных с целью введения в многокадровые методы эластографии, рассмотренные в главе 5.

Глава 5 содержит описание методов сопоставления ультразвуковых медицинских изображений. Далее приводится обзор методов эластографии по двум кадрам (раздел 5.2) и многим кадрам (раздел 5.3). Метод многокадровой эластографии, использующий ультразвуковую пробу с датчиком силы описан в разделе 5.4. Раздел также содержит численные эксперименты по сравнению различных методов эластографии на данных, полученных при помощи моделирования а также реальных снимков ткани человека. Завершает главу описание задачи сопоставления ультразвуковых сигналов, которая возникает в ультразвуковой отражательной томографии и результаты численных экспериментов с использованием метода реконструкции, предложенного автором.

Глава 2

Задача плотного сопоставления изображений

2.1 Постановка задачи

Задача сопоставления двух или серии изображений является одной из важных задач компьютерного зрения, которая возникает во многих приложениях.

Для сопоставления изображений требуется некоторое основное предположение, такое как, например, *инвариантность интенсивности*. Так, рассмотрим два черно-белых изображения с интенсивностями $I_0(x, y)$ и $I_1(x, y)$. Полем смещений называется двумерное векторное поле $\mathbf{u}(x, y) = (u_x(x, y), u_y(x, y))$, определенное в каждом пикселе изображения, $u_x(x, y)$ и $u_y(x, y)$ - горизонтальные и вертикальные компоненты смещений соответственно.

Решением задачи сопоставления изображений называется поле смещений, для которого следующая сумма по всем пикселям изображений минимальна:

$$E(\mathbf{u}) = \sum_{x,y} (I_0(x, y) - I_1(x + u_x(x, y), y + u_y(x, y)))^2. \quad (2.1)$$

Выбор прочих функций помимо квадратичного штрафа подробнее обсуждается в главах 3 и 4. Для реальных данных предположение об инвариантности интенсивности часто нарушено ввиду нескольких причин (см. рис. 2.1). Во-первых, некоторые соответствующие пиксели отсутствуют на втором изображении, что возникает, например, в областях, соответствующих заслоненным объектам. Во-вторых, такое



Рис. 2.1: Особенности сопоставления изображений, проиллюстрированные на примере стерео-сопоставления изображений коллекции KITTI 2015. Ряд (1) - яркие отражения источников света, ряд (2) - прозрачные объекты, ряд (3) - однотонные объекты, не содержащие текстуры, ряд (4) - повторяющиеся структуры, ряд (5) - заслоненные объекты: показаны изображения с левого и правого видов стереокамеры.

допущение не позволяет построить однозначное соответствие в однородных участках, не содержащих текстуры, а также в частях изображения, содержащих повторяющиеся элементы, что особенно актуально при сопоставлении ультразвуковых изображений. Инвариантность интенсивности может также быть нарушена ввиду изменений освещения, шума на изображениях, наличия прозрачных или отражающих объектов.

Для того, чтобы построить плотное сопоставление изображений, основываясь на неполной информации, требуется ввести регуляризацию, которая накладывает штраф на пространственные изменения полей смещения. Как правило, в таком случае задача сопоставления записывается в виде задачи оптимизации.

Другой возможностью является построение сопоставления для серии кадров изображения, при этом основой для регуляризации является пространственно-временная непрерывность поля смещений.

В продолжении данной главы описаны варианты задачи сопоставления в разных приложениях.

2.2 Биноккулярное стерео-сопоставление

Примером задачи, решение которой требует сопоставления изображений, является трехмерная стерео-реконструкция на основе пары изображений со стерео-камеры. Данный подраздел описывает идеи геометрических построений, используемых в методах стерео-сопоставления.

Одной из особенностей визуального восприятия человека является разница между изображениями, получаемыми для левого и правого глаза. Положение объекта, наблюдаемого левым и правым глазом смещено, при этом смещение обратно пропорционально расстоянию от глаза до объекта. Такое свойство может быть использовано для стерео-реконструкции. Решение этой задачи позволяет восстановить трехмерную геометрию сцены посредством оценки глубины для каждого из пикселей изображения.

Величина горизонтального смещения объектов между левым и правым видом стерео-камеры обратно пропорционально расстоянию до наблюдаемого объекта. Для того, чтобы использовать данное свойство для стерео-реконструкции, следует описать модель камеры. Для точечной модели камеры геометрические свойства сцены для

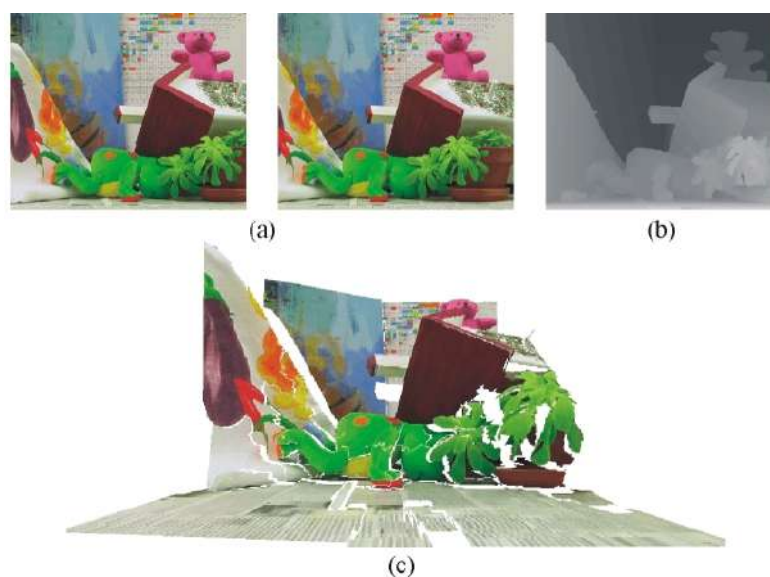


Рис. 2.2: Пример стерео-реконструкции (изображение из работы [18]). Первый ряд изображений (a) показывает левое и правое изображение, второй ряд (b) - поле смещений, третий ряд (c) - результат трехмерной реконструкции.

двух видов описаны в работе [19] (см. рис 2.3). Наиболее сложным этапом стерео-реконструкции является оценка одномерного поля смещения.

Одним из первых приложений, послуживших причиной развития методов стерео-сопоставления, является фотограмметрия и восстановление карт высот на основе серии аэроснимков. Стерео-реконструкция также широко применяется в робототехнике, включая современные системы беспилотного управления автомобилем.

Пример стерео-реконструкции показан на рис. 2.2. После спецификации модели камеры и соответствующих параметров, основным этапом реконструкции является сопоставление пары изображений. После ректификации изображений, которая заключается в применении аффинного преобразования, поле смещений, сопоставляющее левое и правое изображений, становится одномерным. Более подробно процесс стерео-сопоставления рассмотрен в главе 3 и работе [11], опубликованной автором.

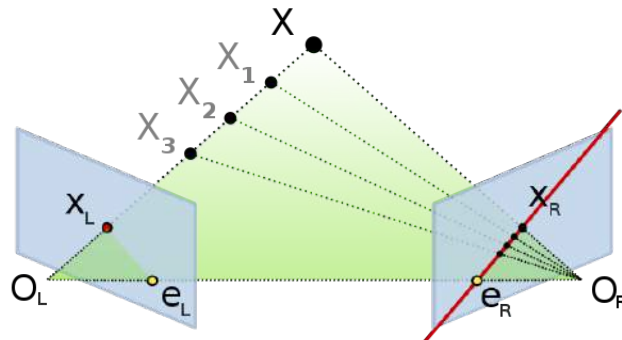


Рис. 2.3: Геометрия двух видов стерео-камеры. Рисунок показывает формирование изображения для левого и правого вида для серии точек с разным расстоянием от камеры. Как видно на рисунке, увеличение расстояния ведет к уменьшению диспаратета.



Рис. 2.4: Пример вычисления оптического потока для кадра видеопоследовательности из публичной коллекции изображений [20]. Эталонный кадр показан на рисунке слева, поле смещений представлено на рисунке посередине, цветовая схема для кодирования двумерных векторов смещения представлена на рисунке справа.

2.3 Оптический поток

Наблюдаемое движение объектов на кадрах видеопоследовательности является важным источником информации для последующего анализа. Задача обнаружения движения движения на кадрах видеопоследовательности называется вычислением *оптического потока*. Так, для изображений $I_0(x, y)$ и $I_1(x, y)$ требуется найти такое двумерное поле смещений $\mathbf{u}(x, y) = (u_x(x, y), u_y(x, y))$, что интенсивности $I_0(x, y)$ и $I_1(x + u_x(x, y), y + u_y(x, y))$ совпадают. Компоненты $u_x(x, y)$ и $u_y(x, y)$ соответствуют горизонтальным и вертикальным смещениям. Для визуализации двумерного поля смещений обычно используется цветовое кодирование, при котором направление движение описывается оттенком цвета, а величина - его интенсивностью. Пример вычисления оптического потока показан на рис. 2.4.

Оптический поток используется во многих приложениях по анализу видео, включая

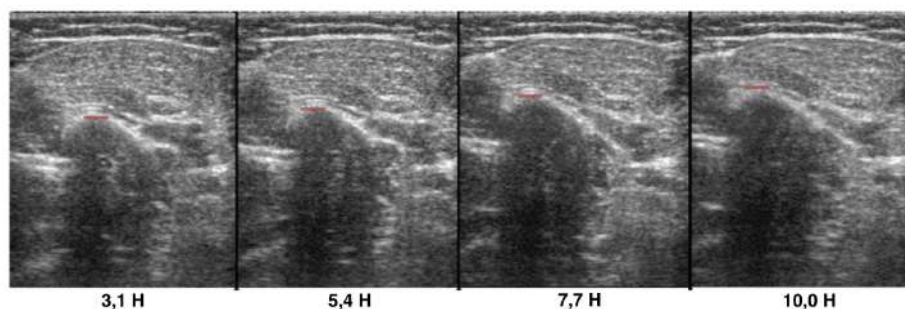


Рис. 2.5: Серия ультразвуковых снимков ткани с разной степенью сжатия. На рисунке также показаны значения силы. Движение одной из точек ткани отмечено красным маркером. Маркер приближается к ультразвуковой пробе по мере увеличения силы.

шумоподавление, стабилизацию камеры, многокадровую семантическую сегментацию и оценку трехмерных векторов движения на сцене. Задача более подробно рассмотрена в главе 4 и работе [12], представленной автором.

2.4 Многокадровая ультразвуковая эластография

Эластографией называется метод оценки механических свойств ткани человека посредством сопоставления изображений ткани с различной степени сжатия. Часто, применение эластографии позволяет избежать биопсии при диагностике опухолей, а также избежать использования невоспроизводимых процедур, таких как пальпация. Эластография основана на сопоставлении ультразвуковых или МРТ снимков. В данной работе рассмотрена так называемая квазистатическая эластография, в которой для получения изображений к ультразвуковой пробе применяется постоянная сила.

Оценка механической деформации в эластографии основана на сопоставлении двух или серии ультразвуковых изображений. При этом за эталонное принимается изображение, полученное без применения механического сжатия. Это изображение сопоставляется с одним или серией изображений, полученных при применении контактной силы (см. рис. 2.5). В данной работе рассмотрена многокадровая эластография, при этом значение силы нажатия измеряется напрямую при помощи ультразвуковой пробы, которая снабжена специальным датчиком.

Протокол использования ультразвуковой эластографии состоит из следующих этапов (см. рис. 2.6). На первом этапе получается серия двух или более изображений, полученных с применением разной контактной силы, начиная с нулевой. При этом

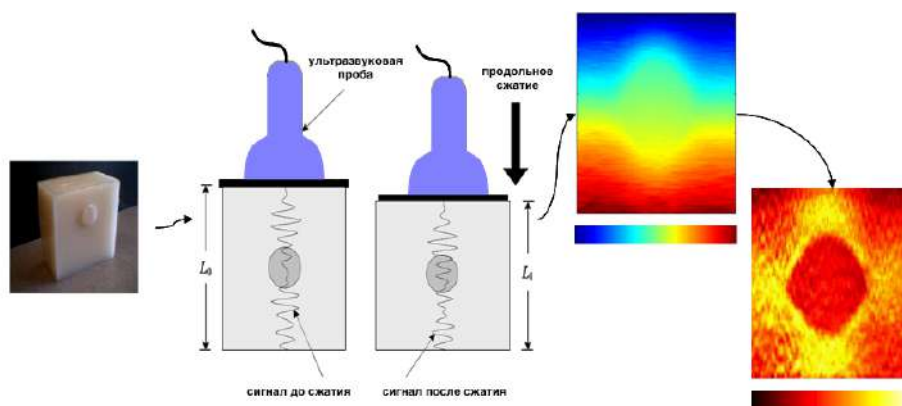


Рис. 2.6: Схематическое изображение этапов работы метода эластографии, изображение из работы [21]. Столбцы слева-направо: (1) объект, содержащий жесткое сферическое включение, (2) ультразвуковое изображение без применения силы, (3) изображение, полученное применением контактной силы продольного направления, (4) продольное поле смещений, (5) продольная механическая деформация, вычисленная из продольного поля смещений.

контактная сила применяется в продольном направлении, т.е. в направлении распространения ультразвукового пучка. Продольное смещение вычисляется путем сопоставления изображений. Затем поле механической деформации оценивается как локальная степень сжатия. Поле механической деформации представляет интерес для медицинской диагностики. Более подробно методы ультразвуковой эластографии рассмотрены в главе 3 и работе [14], представленной автором.

2.5 Приложения

Данный подраздел описывает приложения плотного сопоставления изображений, в которых могут быть использованы методы сопоставления, предложенные автором.

2.5.1 Обнаружение препятствий на дорожных сценах

Одной из важных задач при разработке беспилотного автомобиля является определение геометрии окружающих объектов для того, чтобы прокладывать дальнейший маршрут. Некоторые системы используют лазерный сканер (LIDAR), в то время как более экономически целесообразным представляется использование стерео-камеры. При этом вместо использования аппаратного решения, задача реконструкции геометрии окружающей сцены ложится на программную реализацию метода стереосопоставления.

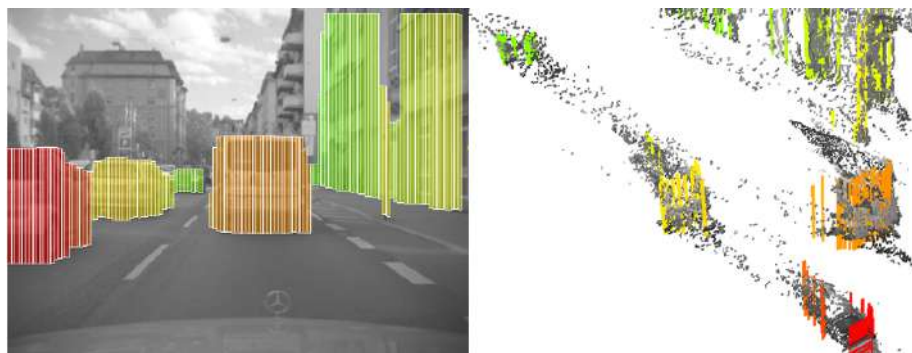


Рис. 2.7: Визуализация препятствий на сцене, полученная путем стерео-реконструкции, изображение из работы [22]. Столбец (1) - изображение с левой камеры. Столбец (2) показывает трехмерное облако точек, полученное в результате стерео-реконструкции (вид сверху). Сходными цветами обозначены точки близкой глубины. Более близкие препятствия помечены красным, препятствия средней глубины помечены оранжевым и наиболее далекие препятствия, соответствующие зданиям, помечены зеленым.

Целью стерео-реконструкции является качественное сопоставление изображений в режиме реального времени с использованием ограниченных вычислительных ресурсов. Задачей таких методов является преодоление практических трудностей, таких как изменяющееся освещение, заслоненные объекты, блики и однородные области, не содержащие текстуры.

Пример трехмерного облака точек, полученного методом стерео-реконструкции показан на рис. 2.7 справа. Каждый из наборов областей постоянного диспаратета соответствует некоторому объекту, такому как, например, пешеход, автомобиль или стена здания [22]. Остальная часть облака, соответствующая дорожной поверхности, не включена в визуализацию.

2.5.2 Оценка скорости движения на дорожных сценах

Оптический поток - это плотное двумерное поле смещений, которое описывает движение объектов на сцене относительно камеры. Приложения такой задачи включают в себя навигацию в робототехнике, ориентирование в динамически меняющихся окружениях, механическую манипуляцию объектами и любые другие задачи, требующие оценки поля скоростей для окружающих объектов. Вычисление оптического потока также используется в системах дополненной реальности.

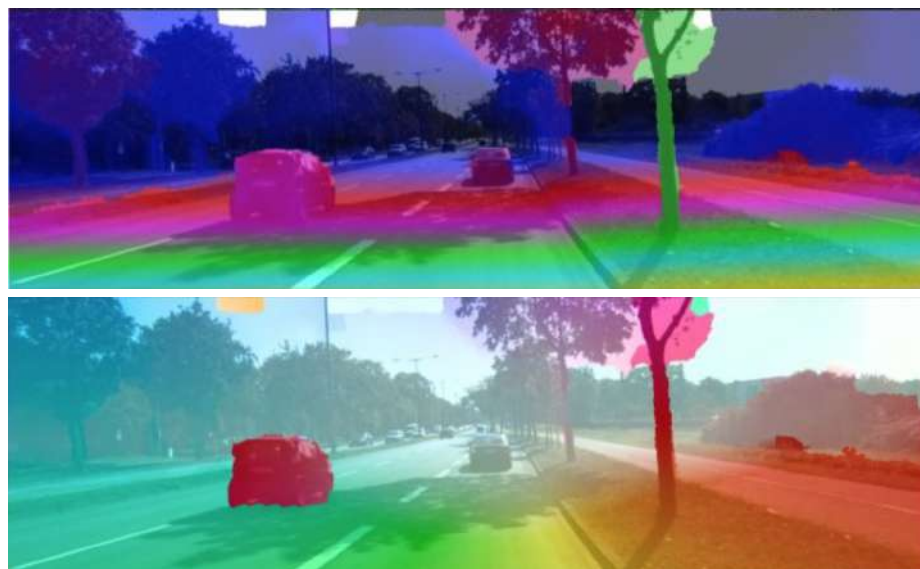


Рис. 2.8: Пример поля оптического потока, изображение из работы [23]. Верхний ряд показывает поле диспаратета поверх эталонного изображения, нижний ряд показывает поле оптического потока поверх эталонного изображения. Комбинация полей диспаратета и оптического потока позволяет восстановить трехмерные поля скоростей окружающих объектов.

Пример вычисления оптического потока для дорожной сцены показан на рис. 2.8. В данном случае вычисление оптического потока произведено одновременно со стереореконструкцией, что позволяет улучшить качество при решении обеих задач. Полученные в результате трехмерные вектора, описывающие скорости окружающих объектов относительно камеры, предоставляют исчерпывающую информацию о движении окружающих объектов, включая пешеходов. Такая информация является ключевой для планирования дальнейшего маршрута в системах автоматического управления автомобилями.

2.5.3 Анализ механических свойств ткани человека для медицинской диагностики

Эластография - это метод оценки механических свойств ткани человека. Данные, полученные при применении эластографии, сходны с результатами пальпации, однако являются более воспроизводимыми и менее подвержены ошибкам ввиду человеческого фактора. При этом, в отличие от пальпации, изображения, полученные методом эластографии, могут использоваться для дальнейшего количественного анализа с применением алгоритмов сегментации. Локальная деформация, получаемая методом эластографии, обратно пропорциональна жесткости ткани, что позволяет эффективно диагностировать злокачественные опухоли в ткани человека

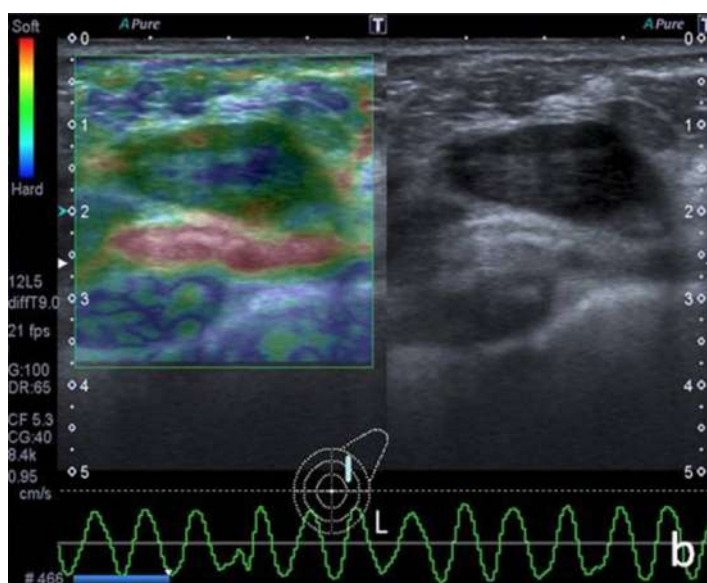


Рис. 2.9: Пример поля механической деформации, полученного методом ультразвуковой эластографии. Поле деформации соответствует злокачественной опухоли, обнаруженной в груди человека, изображение из работы [24]. Поле механической деформации показано поверх ультразвукового снимка на рисунке слева. Ультразвуковой снимок показан на рисунке справа. Твердое включение, соответствующее синему цвету, представляет собой злокачественную фибroadену, обнаруженную у пациента.

без хирургического вмешательства (рис. 2.9). Прочие заболевания включают в себя, например, фиброз печени, тиронд простаты, а также патологии почек и лимфоузлов. Более подробное описание различных клинических применений эластографии приведено в работе [25].

Глава 3

Биноккулярное стерео-сопоставление

Данная глава начинается с описания трехмерной геометрии двух видов и постановки задачи стерео-сопоставления. Подраздел 3.2 содержит обзор методов плотного сопоставления, основанных на минимизации функционалов энергии. Также описаны понятия стоимостей сопоставления и агрегирования тензора энергий. Подраздел 3.3 содержит обзор методов фильтрации изображений, учитывающих границы объектов, которые используются в методах стерео-реконструкции. Главу продолжает обзор современных методов для задачи, основанных на сверточных нейросетях (подраздел 3.4). Завершает главу описание метода агрегирования тензора энергий, предложенного автором и результаты численных экспериментов на реальных данных.

3.1 Введение в стерео-реконструкцию

3.1.1 Эпиполярная геометрия

Пусть дан пиксел p на левом изображении стерео-камеры I_l . Задачей стерео-сопоставления является нахождение соответствующего пиксела на правом изображении I_r . Данный подраздел описывает последовательность аффинных преобразований пары изображений, которые сводят задачу стерео-сопоставления к вычислению одномерного поля смещений, называемого *диспаратетом*.

На рис. 3.1 показано, как точка трехмерного пространства отображается на *эпиполярные прямые*. Эпиполярная плоскость (см. рис. 3.1) есть плоскость, проходящая

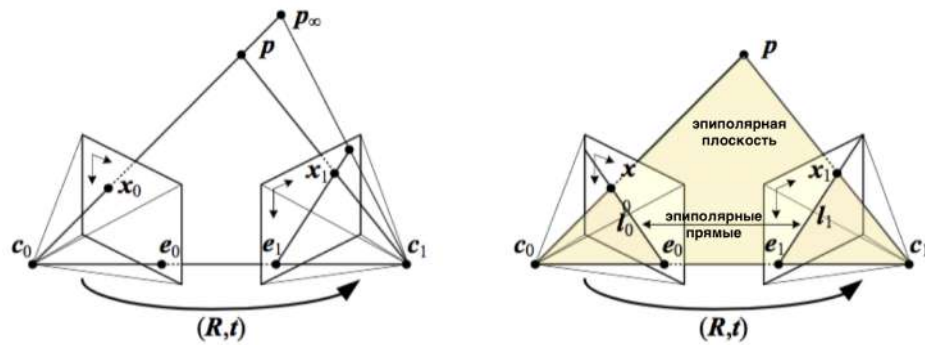


Рис. 3.1: Геометрия двух видов для простейшей точечной модели камеры (изображение из работы [1]). Эпилярная плоскость, определенная для точки p окрашена желтым цветом. Эпилярные прямые получаются пересечением эпилярной плоскости с двумя плоскостями изображения.

через центры левой и правой камер. Эпилярные прямые получаются при пересечении эпилярной плоскости и плоскости изображения.

Процедура, позволяющая совместить эпилярные прямые на двух изображениях, называется *ректификацией* и состоит из следующих этапов, каждый из которых включает в себя применение аффинного преобразования к исходным изображениям: на первом этапе изображения применяется такая аффинная трансформация, что эпилярные прямые становятся параллельными. На втором этапе плоскости изображений поворачиваются в трехмерном пространстве так, что эпилярные прямые совпадают [1].

После выполнения процедуры ректификации, задача стерео-реконструкции состоит в том, чтобы найти поле диспаратета, т.е. одномерное поле горизонтальных смещений d , которое позволяют сопоставить пиксели изображений $I_l(x, y)$ и $I_r(x - d, y)$. При этом каждая сопоставляемая пара пикселей соответствует лучам, исходящим из центров камер, которые пересекаются в ближайшей точке трехмерной поверхности.

3.1.2 Задача бинокулярного стерео-сопоставления

Задачей стерео-сопоставления является оценка значения глубины для каждого пиксела на основе двух изображений с откалиброванной стерео-камеры. Реконструкция основана на вычислении поля диспаратета, т.е. горизонтального поля смещения, вычисленного для каждого пиксела левого изображения $I_L(x, y)$. Как было показано

выше, поле диспаратета является одномерным. При использовании точечной модели камеры, глубина z вычисляется из диспаратета d , используя соотношение:

$$z = \frac{fB}{d}, \quad (3.1)$$

где f - фокусное расстояние камеры, а B - расстояние между центрами камер.

3.1.3 Особенности задачи стерео-сопоставления для реальных данных

Задача стерео-сопоставления на реальных изображениях встречает ряд практических трудностей. Изображения, полученные с левого и правого вида стерео-камер претерпевают ряд случайных изменений разной природы. Во-первых, это шум и фотометрические изменения. Во-вторых, это разница в освещении при наблюдении с разных точек пространства. Некоторые участки изображения содержат яркие отражения источников света, прозрачные части, а также отражающие свет объекты. В третьих, фотографии дорожных сцен содержат значительное количество однородных участков, которые не содержат текстуры: в таких случаях, исходя из визуальной информации, можно сделать вывод о большом количестве возможных совпадений, при этом только одно из них соответствует реальному трехмерному объекту.

Дополнительным препятствием к установлению соответствия является присутствие заслоненных объектов. В этом случае ожидаемым результатом работы стерео-реконструкции является реконструкция глубины для объектов, которые представлены только на одном из видов. При этом использование только визуальной информации недостаточно для установления соответствия, требуется введение дополнительных ограничений на ожидаемую форму объектов в подобных случаях. Дополнительную сложность также создают перспективные преобразования - различные объекты меняют форму при съемке с разных углов зрения.

Список наиболее часто встречаемых особенностей стерео-сопоставления на примере коллекции изображений KITTI 2015 представлен на рис. 2.1. Соответствующие фрагменты изображений выделены синими прямоугольниками.

3.2 Классификация методов стерео-сопоставления

Данный подраздел описывает различные этапы работы методов стерео-сопоставления в соответствии с работой [15], в завершении приведена краткая классификация таких методов.

3.2.1 Вычисление тензора энергий

Классические методы стерео-сопоставления используют различные метрики для сопоставления изображений, включая абсолютное значение попиксельных разностей (SAD), квадрат разностей интенсивности (SSD), нормированные кросс-корреляции (NCC), взимная информация и многие другие. Например, используя абсолютное значение попиксельных разностей для фрагментов размера 1×1 , тензор энергии может быть вычислен как:

$$C_{SAD}(x, y, d) = |I_l(x, y) - I_r(x - d, y)|. \quad (3.2)$$

На практике, вычисление тензора энергий, основанное на локальном сравнении прямоугольных фрагментов изображения, ведет к очень шумным полям диспаритета для реальных данных, если не применяются дополнительных методов обработки. Один из способов уменьшить количество выбросов поля диспаритета - увеличить размер используемых фрагментов. Однако, это ведет к утолщению границ объектов. Так, способность уменьшить ошибку поля диспаритета путем агрегирования визуальной информации для соседних пикселей имеет фундаментальный предел. Каждый метод данного семейства достигает баланса между количеством выбросов вычисленного поля диспаритета и способности аккуратно восстановить границы объектов на сцене. Этот предел послужил причиной развития методов регуляризации, учитывающих границы объектов. Такие методы описаны в разделе 3.3.

3.2.2 Агрегирование тензора энергий

Одним из общих подходов для построения методов стерео-сопоставления является агрегирование тензора энергии по некоторой окрестности выбранного пикселя изображения. При этом усреднение по двумерной окрестности соответствует поверхностям, параллельным плоскости изображения камеры, а использование трехмерных окрестностей - наклонным поверхностям.

Обозначим трехмерный тензор энергий как $C(x, y, d)$. Тогда агрегирование с использованием трехмерной окрестности может быть записано как:

$$C_{smooth}(x, y, d) = C(x, y, d) * \omega(x, y, d).$$

Использование более совершенных методов агрегирования позволяет уменьшить ошибку оцениваемого поля диспаритета. Примеры используемых в данном контексте фильтров включают в себя управляемый фильтр, предложенный в работе [26], рекурсивный фильтр, учитывающий границы объектов [27], а так же нелокальные подходы к регуляризации [28].

3.2.3 Вычисление диспаритета как задача оптимизации

3.2.3.1 Метод динамического программирования

Одним из широко используемых на практике методов является алгоритм semi-global matching, предложенный в работе [29]. Метод имеет низкую вычислительную сложность благодаря использованию следующих двух упрощений. Во-первых, функция штрафа, накладываемая на поле диспаритета, рассматривает всего три различных возможности для каждого пиксела p : первый - диспаритет постоянен в окрестности данного пиксела, второй - диспаритет изменяется на единицу (например, в участке наклонной поверхности), третий - диспаритет имеет разрыв. При этом вводятся две константы, P_1 и P_2 , которые соответствуют штрафу за единичное изменение и разрыв диспаритета соответственно. Особенностью такой функции штрафа является независимость накладываемого штрафа от величины разрыва диспаритета, что позволяет построить метод низкой вычислительной сложности с использованием динамического программирования.

Итак, будем вычислять поле диспаритета как решение следующей оптимизационной задачи:

$$E_{SGM}(D) = \sum_{\mathbf{p}} \left(C(\mathbf{p}, D(\mathbf{p})) + \sum_{\mathbf{q} \in \mathcal{N}_p} P_1 \times \{|D(\mathbf{p}) - D(\mathbf{q})| = 1\} + \sum_{\mathbf{q} \in \mathcal{N}_p} P_2 \times \{|D(\mathbf{p}) - D(\mathbf{q})| > 1\} \right), \quad (3.3)$$

Где $\mathbf{p} = (x, y)$ - это пиксел изображения, а \mathcal{N}_p - множество пикселов окрестности. Данный метод решает описанную двумерную задачу минимизации приближенно,

используя серию аналогичных одномерных задач. При этом используется окрестность из четырех или восьми пикселей изображения. Для того чтобы уменьшить количество выбросов, которые возникают ввиду приближенного решения, применяется несколько итераций, каждый из которых выполняет четыре направленных прохода - слева направо, справа налево, сверху вниз, снизу вверх (вектор r соответствует различным направлениям прохода):

$$C_r(x, y, d) = C(x, y, d) - \min_k C_r(x-r, d, k) + \min_k \{C_r(x-r, d, d), C_r(x-r, d, d-1) + P_1, C_r(x-r, d, d+1) + P_1, \min_k \{C_r(x-r, d, k) + P_2\}\}. \quad (3.4)$$

Финальный тензор энергий оценивается путем усреднения результатов прохода по четырем направлениям:

$$C_{SGM}(x, y, d) = \frac{1}{4} \sum_r C_r(x, y, d). \quad (3.5)$$

Для вычисления диспаратета используется выбор минимальной энергии по третьему измерению тензора:

$$D(x, y) = \arg \min_d C_{SGM}(x, y, d). \quad (3.6)$$

Таким образом, решение двумерной задачи 3.3 было сведено к серии одномерных задач. Описанный метод имеет низкую вычислительную сложность, что позволяет быстро решать задачу стерео-сопоставления на реальных данных. Для того, чтобы гарантировать достаточную степень гладкости получаемых полей смещений, на практике как правило используется несколько итераций алгоритма. При этом сглаженный тензор энергий 3.5 снова используется для решения оптимизационной задачи 3.3.

3.2.3.2 Методы на основе случайных Марковских полей

Рассмотрим следующую оптимизационную задачу:

$$E_{MRF}(D) = \sum_{x,y} \left(C(x, y, D(x, y)) + \rho(d(x, y) - d(x+1, y)) + \rho(d(x, y) - d(x, y+1)) \right), \quad (3.7)$$

где вместо фиксированных штрафов P_1 и P_2 рассмотрим монотонно растущую функцию ρ , так что целевая функция будет расти по мере роста величины перепада поля смещений между соседними пикселями.

В литературе предложено большое количество методов стерео-сопоставления, основанных на минимизации функционала 3.7 и его модификациях. При этом минимизация может быть как сведена к серии одномерных задач, либо решена напрямую приближенно - обзор таких методов приведен в работах [30, 31].

3.2.4 Постобработка диспаратета

Поле диспаратета, полученное в качестве решения задачи дискретной оптимизации, является целочисленным, поэтому требуется вычислить субпиксельное уточнение. Одним из применяемых на практике подходов является вписывание параболы в соседние три значения энергии [8]. Рассмотрим тензор энергий $C(x, y, d)$ в некотором пикселе $\mathbf{p} = (x, y)$, которому соответствует поле смещений $d_0 = d(x, y)$. Пусть смещению d_0 соответствует энергия $C = C(x, y, d_0)$, а смещениям $(d_0 + 1)$ и $(d_0 - 1)$ - энергии $C_+(x, y, d_0 + 1)$ и $C_-(x, y, d_0 - 1)$ соответственно. Рассмотрим три вышеперечисленных значения одномерной функции энергии и в эти точки впишем параболу. Ее минимум будет находиться в точке d_{sub} [8]:

$$d_{sub} = d_0 - \frac{C_+ - C_-}{2(C_+ - 2C + C_-)}. \quad (3.8)$$

Таким образом, используя три соседних значения энергии, для каждой пары (x, y) можно вычислить субпиксельное уточнение к первоначальной оценке поля диспаратета.

Помимо субпиксельного уточнения, используются дополнительные шаги постобработки, такие как применение билатеральной фильтрации. На практике используются быстрые приближения билатерального фильтра [32], а также его обучаемые версии [33].

3.2.5 Локальные и глобальные методы стерео-сопоставления

Большинство стерео методов могут отнесены в одну из следующих категорий, в соответствии с этапами стерео-сопоставления, описанными в работе [15]. Каждый алгоритм содержит подмножество следующих шагов:

- Вычисление тензора энергий,

- Агрегирование тензора энергий,
- Вычисление или оптимизация диспаритета,
- Постобработка диспаритета.

Например, семейство *локальных* методов основано на следующих этапах: первый - вычисление тензора энергий, используя сопоставление интенсивностей фрагментов изображения, второй - агрегирование тензора энергий по квадратным окнам для соседних пикселей, третий - вычисление диспаритета путем выбора минимального значения энергии. В *глобальных* методах как правило используется функция штрафа, накладываемая на пространственное изменение диспаритета. При этом финальное поле смещений получается как решение соответствующей оптимизационной задачи.

3.3 Применение фильтров, учитывающих границы объектов на изображении

Одним из основных подходов к улучшению качества стерео-сопоставления, является использование информации о границах объектов на изображении. При этом на практике представляется целесообразным использовать предположение о том, что набор перепадов диспаритета соответствует подмножеству границ объектов. Это предположение может быть использовано как на этапе агрегирования энергий, так и на этапе оптимизации диспаритета. В последнем случае значение функции штрафа на изменение диспаритета уменьшается на границах объектов. Данный подраздел содержит обзор применяемых в данном контексте подходов к фильтрации изображений.

3.3.1 Билатеральный фильтр

Билатеральный фильтр - это нелинейный фильтр, сохраняющий границы объектов, который вычисляет интенсивность каждого пикселя как взвешенную сумму интенсивностей его окрестности. Идея подхода состоит в том, чтобы вычислять значения весов в зависимости от двух факторов: Евклидово расстояние до центрального пикселя и расстояние в пространстве цветов. Например, для двумерного черно-белого изображения с интенсивностью $I(i, j)$, выход такого фильтра может быть вычислен следующим образом:

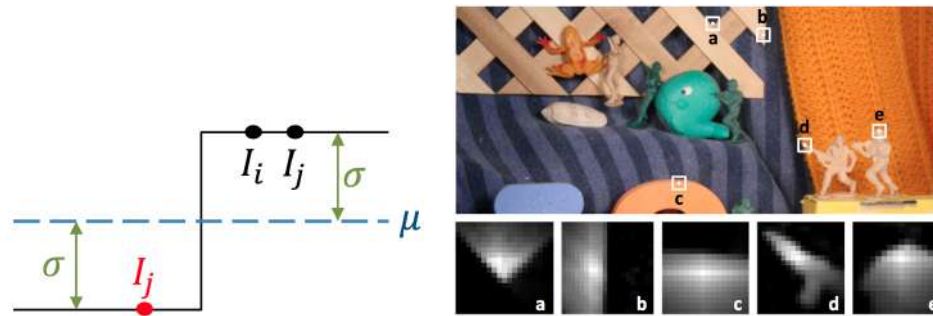


Рис. 3.2: Работа управляемого фильтра (изображения из работы [26]). Столбец (1) - пример одномерного перепада интенсивностей, иллюстрирующего работу соотношения 3.11 на границе объектов. Столбец (2) показывает вклад пикселей в результат фильтрации для квадратного окна.

$$I_{filt}(i, j) \sim \sum_{k,l} \exp\left(-\frac{(i-k)^2 + (j-l)^2}{2\sigma_s}\right) \cdot \exp\left(-\frac{\|I(i, j) - I(k, l)\|}{2\sigma_r}\right) I(k, l) \quad (3.9)$$

Первый множитель характеризует пространственное расстояние, в то время как второй множитель характеризует цветовое расстояние. При этом, каждый из множителей представляет собой нормальное распределение с дисперсией σ_s и σ_r соответственно.

3.3.2 Управляемый фильтр

Другим подходом к нелинейной фильтрации является *управляемый фильтр* [34]. Его выход зависит от управляющего изображения, которое может совпадать или не совпадать с фильтруемым изображением. Рассмотрим применение этого метода для построения фильтра, учитывающего границы объектов.

Выход управляемого фильтра для входного изображения I_0 может быть вычислен как следующее взвешенное среднее:

$$I_{filt}(x, y) = \sum_{p,q} W(x, y, p, q) I_0(p, q), \quad (3.10)$$

где суммирование производится по всему изображению. При этом веса W зависят управляющего изображения \hat{I} согласно соотношению:

$$W(x, y) = \frac{1}{|\omega|^2} \sum_{k:(i,j) \in \omega_k} \left(1 + \frac{(\hat{I}_i - \mu_k)(\hat{I}_j - \mu_k)}{\sigma_k^2 + \epsilon} \right), \quad (3.11)$$

где μ_k и σ_k - среднее и дисперсия изображения \hat{I} по квадратному окну ω_k размера $r \times r$ с центром в пикселе k , $|\omega|$ - количество пикселей в этом окне, а $\epsilon > 0$ есть параметр сглаживания.

Для того, чтобы показать действие такого фильтра в окрестности границы объекта, рассмотрим одномерный срез интенсивности, изображенный на рис. 3.2 слева. Числитель в соотношении 3.11, равный $(\hat{I}_i - \mu_k)(\hat{I}_j - \mu_k)$ имеет положительный знак, если пиксели I_j и I_i расположены по одну сторону от перепада интенсивности, иначе знак отрицательный. Таким образом, интенсивности пикселей не усредняются, если они разделены границей объекта. Степень сглаживания контролируется параметром ϵ : если оно велико, то действие фильтра приближается к фильтру низких частот с равными весами $W(x, y) = \frac{1}{|\omega|^2} \sum_{k:(i,j) \in \omega_k} 1$.

Таким образом, управляемый фильтр является еще одним способ построения сглаженных изображений с учетом границ объектов [26, 34]. Пример вклада соседних пикселей при показан на рис. 3.2 справа.

3.3.3 Рекурсивный фильтр, учитывающий границы

Рекурсивный фильтр является одним из подходов к фильтрации изображений, имеющим низкую вычислительную сложность и возможность эффективной параллелизации. Данный метод имеет связь с приближенным билатеральным фильтром [35]. Практические преимущества метода включают в себя небольшой объем используемой памяти и независимость вычислительной сложности от параметров фильтра.

Алгоритм вычисления фильтра принимает на вход сигнал x и вектор коэффициентов $w_i \in [0, 1]$. Результатом применения фильтра является сглаженный сигнал y . Для одномерных сигналов, вычисление фильтра происходит в соответствии со следующей рекуррентной последовательностью. Начиная с $y_1 = x_1$, для последующих $i = 2, \dots, N$ имеем:

$$y_i = (1 - \omega_i)x_i + \omega_i y_{i-1}. \quad (3.12)$$

Варьирование весов ω_i используется для контроля степени сглаживания, что дает возможность сохранять пространственные границы объектов на изображении: в

самом деле фрагменты изображения с величиной ω_i , близкой к единице, аккумулируются между соседними пикселями y_{i-1} и y_i , тогда как величина w_i , близкая к нулю (например, в пикселях, соответствующих границам объектов), заставляет фильтр подавать на выход сигнал, равный входу, т.е. $y_i = x_i$.

Для двумерных изображений фильтр применяется в сепарабельном виде, т.е. в виде четырех направленных одномерных проходов: слева направо, справа налево, сверху вниз, снизу вверх. Подробное описание алгоритма приводится в разделе 3.5.1.3.

3.3.4 Метод нелокальных средних

Одним из важных наблюдений, применяемых при решении задач шумоподавления, а также при решении недоопределенных обратных задач является наличие похожих фрагментов, встречаемых на реальных изображениях. Поэтому использование всего изображения при решении задачи существенно уменьшить ошибку получаемых решений.

Одним из таких подходов является использование оператора нелокальных средних. Пусть дано изображение v , определенное в области Ω , тогда выходное изображение может быть вычислено следующим образом:

$$u(p) = \frac{1}{C(p)} \int_{\Omega} v(q) f(p, q) dq, \quad (3.13)$$

где $f(p, q)$ функция весов и интеграл вычисляется по всей области Ω . Нормировочная функция $C(p)$ в соответствии с соотношением:

$$C(p) = \int_{\Omega} f(p, q) dq. \quad (3.14)$$

В качестве функции весов будем использовать следующее нормальное распределение:

$$f(p, q) = \exp\left(-\frac{|B(q) - B(p)|^2}{\sigma^2}\right), \quad (3.15)$$

где $B(p)$ и $B(q)$ - локальные средние интенсивности изображения, вычисляемые по некоторому квадратному окну, например $B(p)$ может быть вычислено следующим образом:

$$B(p) = \frac{1}{|R(p)|} \sum_{i \in R(p)} v(i), \quad (3.16)$$

где $R(p) \subseteq \Omega$ - квадратное окно с центром в пикселе p и $|R(p)|$ - число пикселей в окне.

Метод нелокальных средних является в частности, одним из подходов к агрегированию тензора энергий в задаче стерео-сопоставления [28].

3.4 Обзор методов, основанных на глубоком машинном обучении

Данный подраздел содержит обзор основных методов стерео-сопоставления, основанных на глубоком машинном обучении. Обзор начинается с описания метода обратного распространения ошибки, который является общим методом, используемым для обучения нейросетевых моделей.

3.4.1 Метод обратного распространения ошибки

Рассмотрим метод обратного распространения ошибки, который применяется для обучения фильтров в сверточных и рекуррентных нейросетях, используемых для задачи стерео-сопоставления.

Опишем метод на примере одного слоя нейросети, который описывается функцией $f(x; \theta)$, где θ - набор обучаемых параметров. Обозначим выход слоя как y :

$$y = f(x; \theta). \quad (3.17)$$

Пусть выход слоя подается на вход последующего слоя, описываемого функцией z :

$$z(x) = z(f(x; \theta)). \quad (3.18)$$

Градиент выхода слоя z по входу x выражается следующим образом:

$$\frac{\partial z}{\partial x} = \frac{\partial y^T}{\partial x} \frac{\partial z}{\partial y}, \quad (3.19)$$

а градиент z по параметрам θ вычисляется согласно выражению:

$$\frac{\partial z}{\partial \theta} = \frac{\partial y^T}{\partial \theta} \frac{\partial z}{\partial y}. \quad (3.20)$$

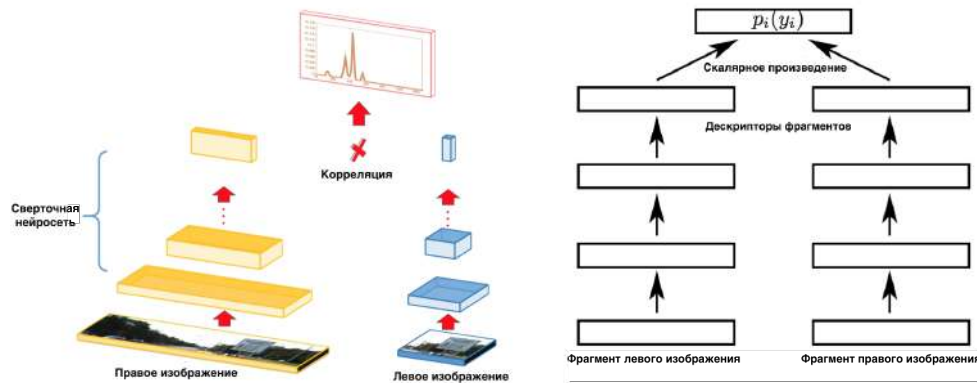


Рис. 3.3: Архитектура сверточной нейросети, используемой для задачи стереосопоставления (изображение из работы [9]). Операция корреляции используется для сравнения фрагмента левого изображения и массива соответствующих фрагментов-гипотез на правом изображении. При этом каждый элемент тензора энергий вычисляется как скалярное произведение соответствующих дескрипторов размерности 64.

Формула 3.20 используется для вычисления градиента выхода слоя z по параметрам θ , при этом требуется вычислять градиенты $\frac{\partial z}{\partial y}$, затем $\frac{\partial y^T}{\partial \theta}$. Такой порядок вычислений является обратным по отношению к порядку вычисления выхода слоя z согласно выражению 3.18, поэтому описанный алгоритм вычисления градиента называется методом *обратного распространения ошибки*.

3.4.2 Обучение глубоких дескрипторов в задачах стереосопоставления

Данный раздел содержит обзор работ, которые основаны на обучении глубоких дескрипторов, используемых для стереосопоставления.

Идея данного семейства методов состоит в том, чтобы использовать сверточную нейросеть для обучения метрики, используемой для сравнения фрагментов изображения:

$$C_{\text{CNN}}(x, y, d) = -f(\langle P^L(x, y), P^R(x - d, y) \rangle). \quad (3.21)$$

Где f - выход сверточной нейросети, $P^L(x, y)$ и $P^R(x - d, y)$ - фрагменты левого и правого изображения с центрами в пикселах (x, y) и $(x - d, y)$ соответственно. При этом, как правило, используются разновидности так называемой Сиамской архитектуры [8, 9, 13]

Первый метод подобного рода, предложенный в литературе, описан в работе [13]. При этом задача сопоставления изображений формулируется в виде задачи классификации на похожие и непохожие пары фрагментов. В качестве тренировочной выборки используются сэмплы в виде набора пар фрагментов из коллекций изображений, для которых задан эталонный диспаритет. На этапе тестирования энергия сопоставления двух фрагментов вычисляется как вероятность принадлежности к похожей паре для данных двух фрагментов. В работе предложено два варианта метода, один из которых позволяет вычислить диспаритет с меньшей ошибкой, а другой имеет меньшую вычислительную сложность.

Более эффективная модель с вычислительной точки зрения предложена в работе [9]. Метод основан на использовании дифференцируемой операции корреляции (рис. 3.3 слева), которая используется в качестве слоя нейросети и используется для сравнения патча с изображением левого вида и линейного массива патчей-гипотез с правого изображения. При этом для моделирования эталонных совпадений используется дельта-функция с пиком, положение которого соответствует верному значению диспаритета. В качестве функции потерь используется функция кросс-энтропии.

Несмотря на сравнительно низкую ошибку диспаритета, которую имеют на практике подобные методы, полученный результат требует постобработки для получения оценки максимально возможного качества. Такая постобработка состоит из нескольких этапов и включает в себя базовый метод агрегирования тензора энергий и оптимизацию поля диспаритета. Последняя представляется наиболее важной для уменьшения ошибки метода на практике. В методе также используется проверка на наличие заслоненных объектов, основанная на сравнении диспаритета для левого и правого вида с последующей интерполяцией выявленных несоответствий.

Несмотря на использование эффективных векторизованных операций, реализация метода [9] требует 1 сек. для вычисления диспаритета с использованием современного графического ускорителя и является неприменимой для практических приложений, требующих вычисления в реальном времени. При этом наиболее трудоемким этапом является вычисление скалярного произведения для большого количества дескрипторов высокой размерности.

3.4.3 Обучаемая регуляризации на основе условных случайных полей

Один из подходов, сочетающих достоинства методов, основанных на обучении глубоких дескрипторов и методов, основанных на дискретной оптимизации предложен в работе [36]. При этом сверточная нейросеть используется как для вычисления

глубоких дескрипторов, так и для предсказания оптимальных потенциалов для последующего метода оптимизации диспаритета, основанного на использовании Марковских случайных полей. Вывод основан на применении двойственного блочного градиентного метода, на практике требующего небольшого количества итераций. В качестве функции потерь используется структурный метод опорных векторов. Такой подход позволяет вычислить диспаритет без необходимости дальнейшей постобработки.

Поле диспаритета получается в виде решения следующей оптимизационной задачи:

$$\min_{x \in X} \left(\sum_{i \in \mathcal{V}} f_i(x_i) + \sum_{i, j \in \mathcal{E}} f_{ij}(x_i, x_j) \right), \quad (3.22)$$

Где \mathcal{V} - множество всех вершин графа (пикселей изображения), \mathcal{E} - множество всех ребер, а $X \in L^\nu$ - набор возможных меток для диспаритета. Первое слагаемое соответствует унарным потенциалам, а второе - бинарным потенциалам, $f_{ij}: L \times L \rightarrow R$:

$$f_{ij}(x_i, x_j) = \omega_{ij} \rho(|x_i - x_j|, P_1, P_2). \quad (3.23)$$

Веса w_{ij} предсказываются для каждого пикселя изображения I_l с использованием сверточной нейросети, ρ - функция сравнения, устойчивая к выбросам, аналогичная функции, используемой в работе [29]:

$$\rho(|x_i - x_j|) = \begin{cases} 0, & \text{если } |x_i - x_j| = 0, \\ P_1, & \text{если } |x_i - x_j| = 1, \\ \text{иначе } P_2. \end{cases} \quad (3.24)$$

Унарные потенциалы получаются сравнением признаков, которые предсказываются последовательностью сверточных слоев. При этом для нормировки используется функция гиперболический тангенс.

Для того чтобы вычислить тензор энергий $C(x, y, d)$, для полученных дескрипторов $\phi^0(x, y)$ и $\phi^1(x - d, y)$, используется функция кросс-корреляции:

$$C(x, y, d) = \frac{\exp \langle \phi^0(x, y), \phi^1(x - d, y) \rangle}{\sum_{j \in D} \exp \langle \phi^0(x, y), \phi^1(x - j, y) \rangle}. \quad (3.25)$$



Рис. 3.4: Иллюстрация алгоритма поиска одномерного поля смещений. На рисунке сверху представлено левое изображение I_l , фрагмент для которого производится поиск соответствия выделен фиолетовым цветом. Правое изображение I_r показано на рисунке снизу: линия, по которой ведется поиск соответствия, выделена зеленым.

Для обучения такой нейросети используется функция потерь, используемая в структурном методе опорных векторов. Параметры нейросети обучаются методом обратного распространения ошибки.

3.4.4 Функции потерь в обучаемых моделях стере-сопоставления

Одним из ключевых аспектов построения моделей машинного обучения для задачи стере-сопоставления является выбор функции потерь, естественным образом описывающей задачу выбора значения энергии, соответствующего эталонному диспаритету. Рассмотрим операцию *argmin*, которая используется в классических методах сопоставления:

$$D(x, y) = \arg \min_d C(x, y, d). \quad (3.26)$$

В самом деле, такая операция является дискретной, что не позволяет вычислить производную от выхода $D(x, y)$ по элементам тензора энергий $C(x, y, d)$. Для того, чтобы построить нейросеть для задачи стере-сопоставления, требуется сконструировать дифференцируемую функцию, которая приближает такую операцию. На

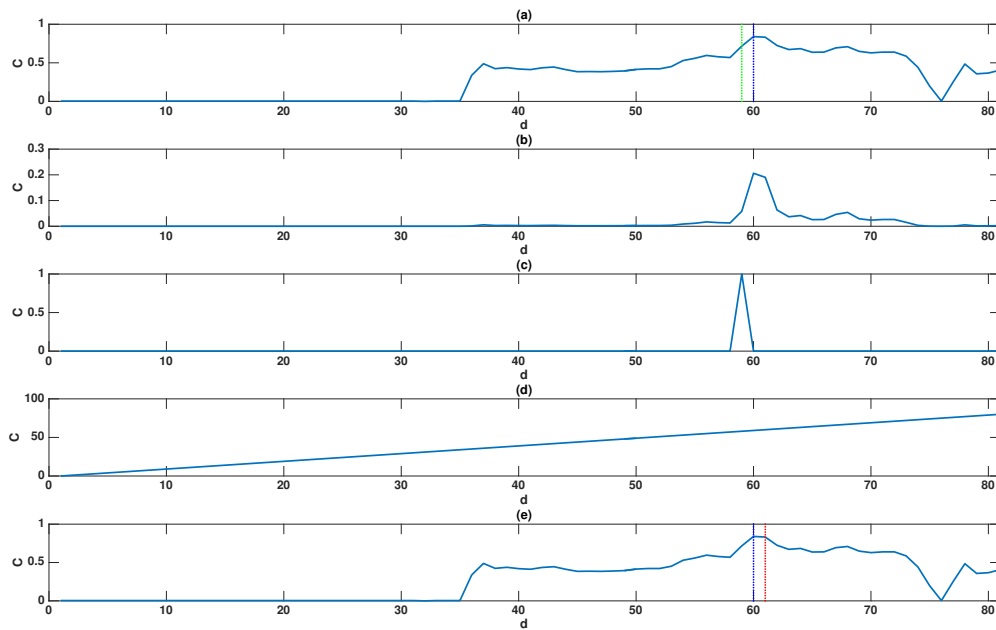


Рис. 3.5: Иллюстрация методов построения функции потерь для задачи стереосопоставления. (a) Пример одномерной функции энергии. Синим цветом показан максимум функции, зеленым цветом выделено эталонное значение диспаратета. (b) Результат применения операции *softmax* к функции энергии. (c) Дельта функция с пиком, соответствующим эталонному значению диспаратета. (d) Вектор значений диспаратета, соответствующих оси абсцисс. (e) Одномерная функция энергии. Синим цветом выделен результат дискретной операции *argmax*, красным цветом выделен результат дифференцируемой операции *argmax*, заданной выражением 3.34.

практике часто удобнее приближать функцию выбора максимума, поэтому далее будет рассмотрен такой эквивалентный вариант задачи.

Для того, чтобы наглядно проиллюстрировать изложенные соображения, рассмотрим одномерную функцию энергии сопоставления, соответствующую фрагменту на левом изображении с центром в точке (x, y) :

$$E_{x,y}(d) = C(x, y, d). \quad (3.27)$$

Пример такой функции, соответствующий изображениям на рис. 3.4, показан на рис. 3.5 (a). Функция имеет глобальный максимум, который соответствует значению диспаратета $d_{argmax} = 60$ (синяя пунктирная линия), при этом эталонное значение диспаратета $d_{gt} = 59$ (зеленая пунктирная линия). Также имеется несколько локальных максимумов. В таком случае, функция потерь должна быть построена таким образом, чтобы усилить главный максимум посредством изменения параметров модели.

Функция *softmax* и кросс-энтропийная функция потерь

Для того, чтобы выделить максимум функции энергии сопоставления, требуется применить функцию предобработки, которая усилит имеющиеся максимумы на фоне остальных ненулевых значений функции. В качестве такой функции может быть использована операция *softmax*, заданная отображением $\sigma : \mathbf{R}^K \rightarrow [0, 1]^K$. Такое отображение для некоторого одномерного вектора энергий E выражается следующим образом:

$$\sigma_j(E) = \frac{e^{E_j}}{\sum_{i=1}^K e^{E_i}}, \quad (3.28)$$

Подобная функция предобработки применяется независимо ко всем одномерным функциям энергии, соответствующим различным пикселям изображения:

$$C_{sm}(x_0, y_0, d) = \sigma(C(x_0, y_0, d)). \quad (3.29)$$

Результат применения функции *softmax* для энергии на рис. 3.5 (а) показан на рис. 3.5 (b).

Для того, чтобы смоделировать функцию, которая имеет максимум в эталонном значении диспаратета, используется дельта функция (см. рис. 3.5 (c)):

$$C_{gt}(x, y, d) = \begin{cases} 1, & \text{если } d = D_{gt}(x, y) \\ 0, & \text{иначе.} \end{cases} \quad (3.30)$$

При этом представляется целесообразным использовать кросс-энтропийную функцию потерь:

$$L(C, C_{gt}) = - \sum_{x,y} \sum_d C_{sm}(x, y, d) \log C_{gt}(x, y, d). \quad (3.31)$$

Таким образом, используя операцию *softmax*, была построена серия дифференцируемых операций, приближающая дискретную функцию *argmax*. Такое приближение позволяет обучать нейросеть для задачи стерео-сопоставления методом обратного распространения ошибки.

Пороговая функция потерь

Другим способом построения функции потерь для задачи стерео-сопоставления является наложение штрафа в соответствии с пороговой функцией. Для описания

этого метода, введем следующее обозначение для значений тензора энергии с индексами, соответствующими эталонному диспаритету:

$$C_D(x, y) = C(x, y, D(x, y)). \quad (3.32)$$

Используя двумерное поле C_D и некоторую положительную константу τ , пороговая функция потерь может быть записана следующим образом:

$$L(C, D_{gt}) = \sum_{x,y} \sum_d \max[0, C_D(x, y) + \tau - C(x, y, d)]. \quad (3.33)$$

Главным недостатком такого подхода является необходимость использовать одно значение параметра τ для всего изображения. В самом деле, интенсивность изображения значительно меняется, что, вообще говоря, требует соответствующих изменений значений порога.

Построение дифференцируемой операции выбора максимума

Альтернативным способом построения дифференцируемой операции *argmax* является использование взвешенной суммы, в которой значение энергии имеет вес, равный значению соответствующего диспаритета [37]:

$$D(x, y) = \sum_{d=0}^{d_{max}} d \cdot C_{sm}(x, y, d). \quad (3.34)$$

Если применить к энергии предобработку в виде операции *softmax*, можно ожидать, что значения диспаритета, не соответствующие минимальному значению, будут иметь пренебрежимо малый вес (3.5 (b)) и значение взвешенной суммы будет соответствовать верному диспаритету. Пример применения такого метода для нахождения максимума показан на рис 3.5. Значение векторов энергии после применения операции *softmax* показано на рис. рис 3.5(b), вектор индексов, который соответствует первому множителю в скалярном произведении изображен на рис. 3.5(d). Максимум, выбранный в соответствии с выражением 3.34 показан на рис. 3.5(d) красной пунктирной линией, истинное максимальное значение показано синей пунктирной линией. Так, описанный подход позволяет приблизить операцию выбора максимума с точностью, приемлемой для решения задачи стерео-сопоставления на практике [37].

3.5 Обучаемое агрегирование тензора энергий сопоставления на основе сверточно-рекуррентной нейросети.

Данный раздел описывает сверточно-рекуррентную нейросеть для быстрого стереосопоставления, предложенную автором.

Метод имеет этапы вычислений, сходные с серией методов для быстрого стереосопоставления, предложенных в литературе. Во-первых, это метод динамического программирования, применяемый для агрегирования тензора энергий в виде серии одномерных направленных проходов [29]. Во-вторых, это методы агрегирования тензора энергий, использующих фильтры изображений, учитывающих границы, такие как билатеральный фильтр [38], управляемый фильтр [26], и рекурсивный фильтр [39].

Предложенный метод основан на использовании дифференцируемого рекурсивного фильтра [40], который основан на аналогии между используемым графом вычисления и прямым проходом рекуррентной нейросети [41]. Такой подход был впервые использован в задаче семантической сегментации [40]. Серия подходов для семантической сегментации, показывающих связь между условными случайными полями и рекуррентными нейросетями, таких как [42] и [43], также имеет аналогии с предложенным подходом в части построения обучаемых сглаживающих фильтров, использующихся для сегментации изображений.

Подобно работам [8, 9, 13], предложенный метод использует сверточную архитектуру нейросети, однако, машинное обучение применяется на этапе агрегирования тензора энергий, что позволяет построить модель, имеющую более низкую вычислительную сложность на этапе исполнения.

3.5.1 Архитектура нейросети

3.5.1.1 Вычисление энергий стереосопоставления

Предложенный метод в явном виде хранит в памяти трехмерный тензор энергий стереосопоставления в виде трехмерного массива размера (h, w, d_{max}) , где h and w - размеры изображения, а d_{max} - максимальный разрешенный диспаратет.

Вычисление тензора энергий производится в соответствии с локальными методами стереосопоставления. При этом, энергия равна сумме двух членов:

$$E(x, y, d) = \alpha E_{SAD}(x, y, d) + (1 - \alpha) E_{census}(x, y, d), \quad (3.35)$$

где коэффициент $\alpha \in (0, 1)$ используется для того, чтобы изменять вклад каждого из слагаемых.

Первый член есть абсолютное значение разностей интенсивностей соответствующих пикселей (SAD):

$$E_{SAD}(x, y, d) = \sum_{r,g,b} |I^L(x, y) - I^R(x - d, y)|, \quad (3.36)$$

при этом используются фрагменты размера 1×1 , т.е. индивидуальные пиксели, для того чтобы сохранить информацию о текстуре изображения. Сглаживание же происходит только на стадии агрегирования тензора энергий.

Второй член $E_{census}(x, y, d)$ основан на сопоставлении локальных дескрипторов, предложенных в работе [44]. Вычисление такого дескриптора происходит следующим образом. Для черно-белого изображения I , определим функцию ξ , которая будет принимать значение 0 или 1 в зависимости от результата сравнения интенсивностей в пикселях \mathbf{p} и \mathbf{q} :

$$\xi(\mathbf{p}, \mathbf{q}) = \begin{cases} 1, & \text{если } I(\mathbf{q}) < I(\mathbf{p}), \\ \text{иначе } 0. \end{cases} \quad (3.37)$$

Используя такую функцию, определим ценсус-преобразование, которое ставит в соответствии каждому пикселу изображения следующий многомерный вектор из нулей и единиц:

$$R_\tau(\mathbf{p}) = \bigotimes_{[i,j] \in D_w} \xi(\mathbf{p}, \mathbf{p} + [i, j]), \quad (3.38)$$

где \otimes - операция конкатенации, а D_w - набор возможных двумерных смещений внутри квадратного окна размера $n \times n$ с центром в пикселе \mathbf{p} . Таким образом локальная структура каждого фрагмента описывается последовательностью $(n^2 - 1)$ бит. Каждый бит задается сравнением интенсивности центрального пиксела фрагмента с интенсивностью остальных пикселей фрагмента.

Полученные дескрипторы, заданные битовыми последовательностями, сравниваются, используя расстояние Хэмминга. Например, расстояние для пары бинарных векторов \mathbf{a} и \mathbf{b} определим, используя индикаторную функцию \mathbb{I} :

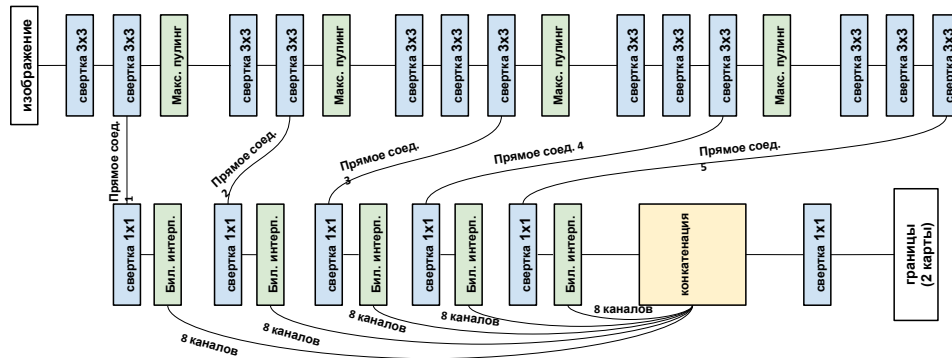


Рис. 3.6: Архитектура сверточного детектора границ [45]. Нейросеть последовательно извлекает признаки с 5 масштабов изображения, используя операцию свертки. Переход к более крупному масштабу осуществляется при помощи операции пулинга максимального элемента в окрестности размера 2×2 . Последний сверточный слой вычисляет линейную комбинацию предсказаний нейросети для всех пяти масштабов изображения. Для извлечения признаков используются прямые соединения, показанные на рисунке. Таким образом, в предсказании участвует информация со всех пяти масштабов.

$$H(\mathbf{a}, \mathbf{b}) = \sum_i \mathbb{I}(\mathbf{a}_i \neq \mathbf{b}_i). \quad (3.39)$$

Такой алгоритм вычисления локального дескриптора естественным образом параллелизуется, и поэтому на практике эффективно реализуется на графическом ускорителе. Так как при этом каждый поток вычисляет дескриптор для единичного пиксела, такая операция выполняется за константное время.

Вычисление тензора энергий $E(x, y, d)$ не участвует в обучении нейросети методом обратного распространения ошибки, поэтому может содержать произвольный набор недифференцируемых операций.

3.5.1.2 Детектор границ объектов на изображении

Аналогично работе [26] в предложенном методе используется агрегирование энергий, которое зависит от входного изображения, описанное далее в подразделе 3.5.1.4. При этом схема, основанная на машинном обучении гарантирует, что сглаживание производится в соответствии со спецификой задачи стерео-сопоставления и учетом особенностей обучающей выборки. Так, задача метода состоит в определении границ объектов, релевантных для перепада поля диспаритета и игнорирование остальных границ. В результате ошибка метода существенно снижается в результате обучения алгоритма.

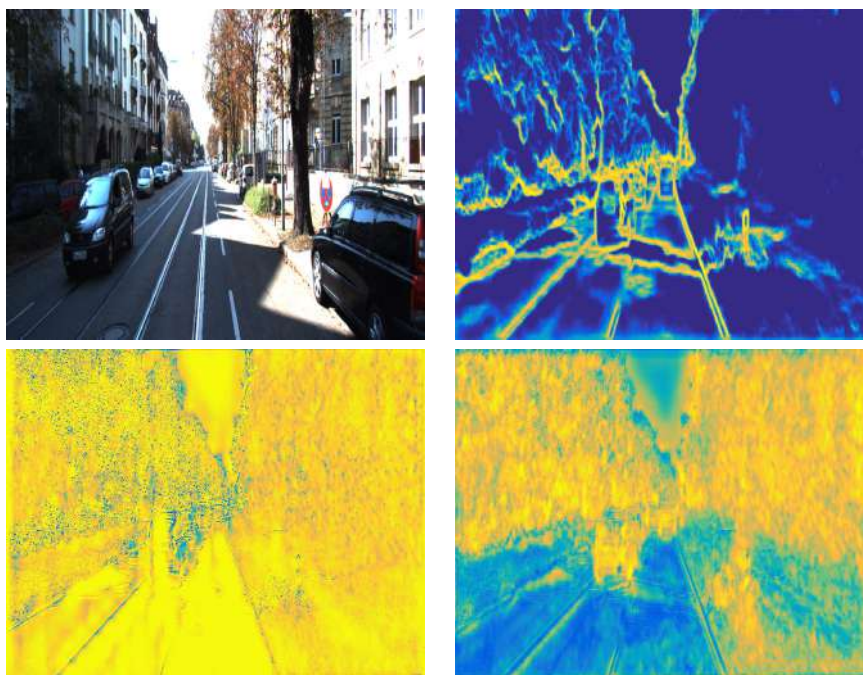


Рис. 3.7: Пример предсказанных весов рекуррентного фильтра. Исходное изображение (1), исходная карта весов до применения обучения (2), веса горизонтального прохода (3), веса вертикального прохода (4). Карта весов значительно меняется в процессе обучения.

В качестве сверточной нейросети, используемой для предсказания границ объектов, была использована многомасштабная архитектура, описанная в работе [45]. Ее описание приводится на рис. 3.6.

Предложенный метод стерео-сопоставления оценивает две карты границ (рис. 3.7), используемых для вертикального и горизонтального проходов нейросети, описанных далее. В соответствии с этим, архитектура сверточной нейросети была модифицирована следующим образом: количество карт на выходе каждого из масштабов было увеличено до восьми, так чтобы последний сверточный слой вычислял значение двух выходных каналов в виде линейной комбинации из независимых восьми слагаемых.

3.5.1.3 Сглаживание при помощи рекурсивного фильтра, учитывающего границы изображения

Предложенный автором метод агрегирования энергий сопоставления основан на рекурсивном фильтре, учитывающем границы, предложенным в работе [35] и его обучаемой версии, использованной для построения сверточной архитектуры для задачи

семантической сегментации в работе [40]. Применение такого метода для одномерных описано в разделе 3.3.3.

Рекурсивный фильтр, учитывающий границы объектов применяется к двумерным изображениям в сепарабельном виде, т.е. вычисление рекурсии осуществляется в виде серии одномерных направленных проходов - горизонтального, слева направо и справа налево и вертикального, сверху вниз и снизу вверх. Ввиду причин, описанных в подразделе 3.5.1.4, представляется целесообразным использовать отдельные карты весов W_h и W_v для горизонтальных и вертикальных проходов соответственно. Действие двумерного фильтра, который принимает на вход изображение I , две карты весов и вычисляет выходное изображение I_{filt} обозначим следующим образом:

$$I_{filt} = F(I, W_h, W_v). \quad (3.40)$$

Алгоритм вычисления описанных выше четырех рекуррентных проходов задается соотношениями:

$$I^L(x, y, d) = (1 - W_h(x, y)) I(x, y) + W_h(x, y) I(x - 1, y), \quad (3.41)$$

$$I^R(x, y, d) = (1 - W_h(x, y)) I^L(x, y) + W_h(x, y) I^L(x + 1, y), \quad (3.42)$$

$$I^T(x, y, d) = (1 - W_v(x, y)) I^R(x, y) + W_v(x, y) I^R(x, y - 1), \quad (3.43)$$

$$I^B(x, y, d) = (1 - W_v(x, y)) I^T(x, y) + W_v(x, y) I^T(x, y + 1), \quad (3.44)$$

где выражения для I^L, I^R, I^T, I^B соответствуют направленным подходам слева направо, справа налево, снизу вверх и сверху вниз соответственно, результат каждого следующего прохода подается на вход предыдущего. На каждом из этапов вычисления осуществляются независимо для строк или столбцов изображения.

Для того, чтобы построить обучаемую модель фильтрации, веса предсказываются на основе входного изображения. Для того, чтобы описать работу алгоритма обратного распространения ошибки для вычисления одномерной фильтрации (3.12), описанной выше, положим, что выход операции передается на вход последующему слою L . Так, в процессе обратного распространения ошибки, градиент каждого сэмпла выходного сигнала y_i получает вклад от производной $\frac{\partial L}{\partial y_i}$. Для того, чтобы вычислить градиент фильтра по входу x , требуется развернуть данный рекуррентный проход в обратном направлении, т.е. для $i = N, N - 1, \dots, 2$ имеем:

$$\frac{\partial L}{\partial x_i} = (1 - \omega_i) \frac{\partial L}{\partial y_i}. \quad (3.45)$$

Для того, чтобы вычислить градиент выхода по набору весов w_i , используется соотношение:

$$\frac{\partial L}{\partial w_i} = \frac{\partial L}{\partial w_i} + (y_{i-1} - x_i) \frac{\partial L}{\partial y_i}, \quad (3.46)$$

вычисление градиента выхода по y производится в соответствии с выражением:

$$\frac{\partial L}{\partial y_{i-1}} = \frac{\partial L}{\partial y_{i-1}} + \omega_i \frac{\partial L}{\partial y_i}. \quad (3.47)$$

Так, четыре прохода описанного выше фильтра могут быть скомбинированы в обучаемую модель, где соотношения 3.12 может быть рассмотрено в качестве элемента рекуррентной нейросети. В самом деле, в последовательности вычислений 3.41, 3.42, 3.43, 3.44, каждый следующий направленный проход принимает на вход результат предыдущего.

Каждое из рекуррентных соотношений аналогично пороговому элементу (gated recurrent unit), предложенному в работе [46] для обучения на данных-последовательностях. Такой элемент используется, например, в задачах анализа текстов. При этом значение $(1 - w_i)$ управляет механизмом обнуления накопленной ранее информации, а значение y_{i-1} является кандидатом на активацию текущего элемента [40].

3.5.1.4 Агрегирование тензора энергий

Автором работы был выбран подход, в котором тензор энергий сглаживается с учетом границ объектов на изображении. Сходная стратегия используется в работе [26]. Предложенный автором подход может также рассматриваться как обобщение работы [27, 39] на случай обучаемого метода. Так параметры фильтра предсказываются на основе входного изображения с использованием сверточной нейросети.

Схема алгоритма агрегирования энергий представлена на рис 3.8. Процесс фильтрации выполняется используя четыре направленных прохода, при этом для того чтобы использовать двумерные карты весов для $W_{h,v}(x, y)$ для фильтрации трехмерного массива $E_d = E(x, y, d)$, $x = 0, \dots, w$, $y = 0, \dots, h$, применяется простое дублирование весов по третьему недостающему измерению:

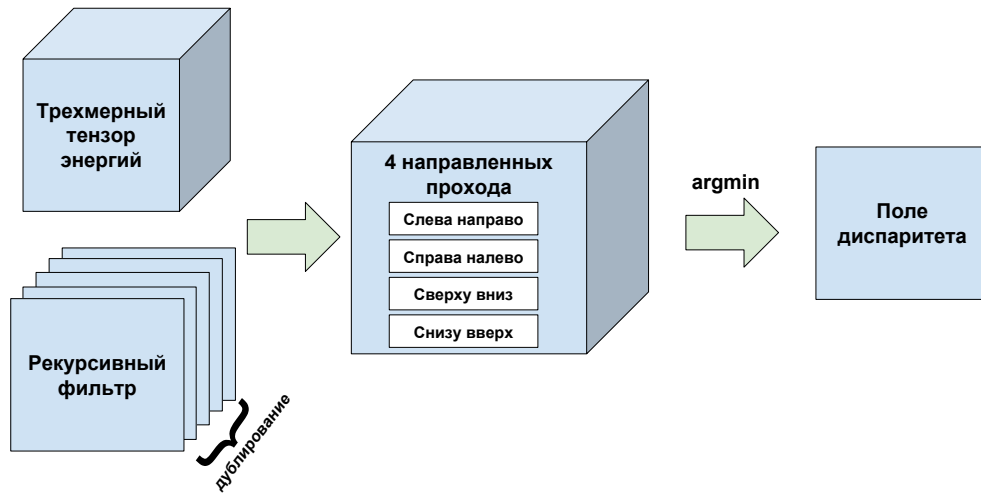


Рис. 3.8: Схематическое описание алгоритма агрегирования тензора энергий. Веса рекуррентного фильтра используются для фильтрации двумерных срезов трехмерного тензора энергий посредством четырех направленных проходов (слева направо, справа налево, сверху вниз, снизу вверх). Фильтрация срезов трехмерного тензора энергий выполняется параллельно. Выбор минимального элемента энергии используется для вычисления диспаратета.

$$E_d^{filt} = F(E_d, W_h, W_v), \quad (3.48)$$

$$d = \{0, 1, \dots, d_{max}\}.$$

Вычисление каждого из двумерного среза тензора E_d осуществляется независимо, поэтому может быть использовано для параллелизации.

Для вычисления поля диспаратета, используется выбор минимального элемента по третьему измерению.

Авторы [40] используют одну карту весов W для каждого из четырех направленных проходов. Подход, описанный в данной работе, однако, использует отдельные карты для вертикального и горизонтального прохода - такой выбор основан на следующем практическом наблюдении: частота и величина изменения диспаратета в горизонтальном и вертикальном направлении существенно различается. Использование двух карт не увеличивает число арифметических операций на этапе исполнения, при этом позволяет уменьшить ошибку сопоставления.

Для того, чтобы далее понизить вычислительную сложность метода, детектор границ применяется к изображениям, уменьшенным до половины размера оригинальных изображений, при этом для применения весов к исходному масштабу используется билинейная интерполяция. Тензор энергий при этом вычисляется для изображений в исходном масштабе. Ввиду того, что изображения уменьшенного размера содержат достаточно информации для обнаружения релевантных границ, данный подход представляется целесообразным, т.к. ведет к уменьшению времени исполнения алгоритма с незначительными или вовсе без издержек со стороны ошибки метода.

Для того, чтобы использовать выходы сверточной нейросети E_h и E_v в качестве входа рекуррентной нейросети, используется следующее нелинейное преобразование:

$$W_h = \exp(-\sigma E_h), \quad (3.49)$$

$$W_v = \exp(-\sigma E_v), \quad (3.50)$$

где σ - настраиваемый параметр.

3.5.1.5 Функция потерь

Для того, чтобы сопоставить фильтруемый тензор энергий с эталонным полем диспаритета, каждая из эталонных меток представляется в виде дельта-функции с пиком, соответствующим эталонному целочисленному значению. При этом используется операция *softmax* и кросс-энтропийная функция потерь, заданная соотношением 3.31.

Модель обучается методом обратного распространения ошибки. При этом разница между вычисленным и эталонным полями диспаритета учитывается функцией потерь и ее градиент вносит вклад в обучение фильтров сверточной нейросети (см. рис. 3.9). Данный подход не требует постобработки.

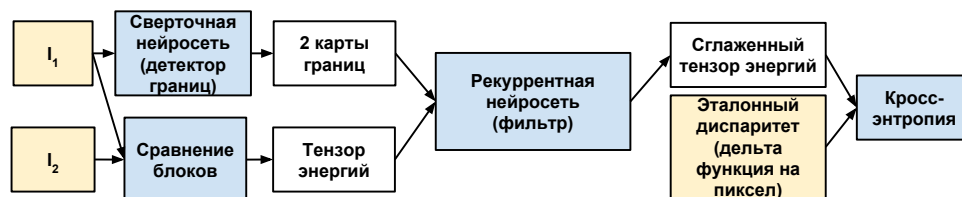


Рис. 3.9: Схема предложенной сверточно-рекуррентной нейросети. Левое и правое изображения стерео-камеры подаются на вход процедуры вычисления тензора энергий. Левое изображение, выбранное в качестве эталонного вида, используется сверточной нейросетью для предсказания релевантных границ. Рекурсивный фильтр, обучаемый в качестве рекуррентной нейросети, принимает на вход тензор энергий и две карты весов, предсказанных сверточной нейросетью. Кросс-энтропийная функция потерь используется в качестве штрафа за отклонение вычисленного диспаритета от эталонных значений. В процессе обучения методом обратного распространения ошибки, градиент функции потерь проходит через рекуррентную нейросеть и вносит вклад в процесс обучения фильтров сверточной нейросети.

3.5.2 Численные эксперименты

3.5.2.1 Описание тренировочной выборки.

Предложенный алгоритм стерео-сопоставления протестирован с использованием открытой коллекции изображений, описанной в работе [15, 23, 47]. Коллекция состоит из ректифицированных пар фотографий со стерео-камеры. Для каждой пары предоставлены эталонные значения диспаритета, которые вычислены используя поля глубины, полученные с использованием лазерного сканера. При этом сканер монтируется на крыше автомобиля, оборудованного стерео-камерой. Ввиду построчного сканирования, эталонный диспаритет представляется набором горизонтальных полос (рис. 3.10), что в результате ведет к уменьшению количества обучаемых данных. При этом количество данных может быть увеличено, используя эвристические подходы, например, с использованием дилатации. Однако, применение подобных техник ограничено в виду невозможности аккуратно представить перепады диспаритета на границах объектов сцены.

3.5.2.2 Методика оценки ошибки метода.

В задачах стерео-сопоставления используется несколько подходов для оценки ошибки. Первый подход состоит в том, что вычисляется средняя абсолютная ошибка поля

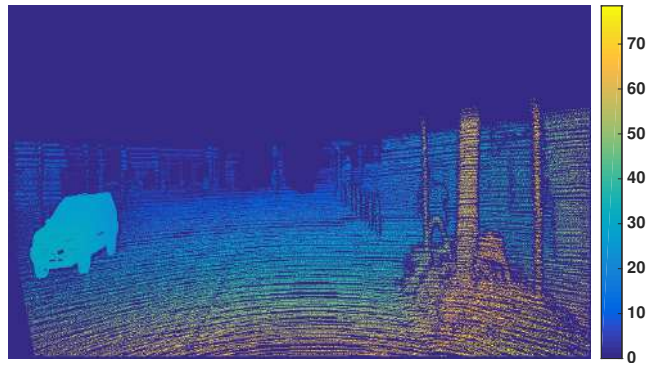


Рис. 3.10: Пример эталонного поля диспаратета, полученного при помощи лазерного сканера. Построчный процесс лазерного сканирования ведет к разреженной структуре обучающих данных, что принимается в расчет при построении методов машинного обучения для данной задачи.

диспаратета по изображению. Подобная стратегия чаще применяется при построении алгоритмов вычисления оптического подхода. Другой подход состоит в подсчете доли пикселей, в которых абсолютная ошибка диспаратета превосходит некоторое пороговое значение:

$$e(D, D_{gt}) = \frac{1}{N} \sum_{(x,y)} (\mathbb{I}[|D(x, y) - D_{gt}(x, y)| > t]), \quad (3.51)$$

где N - количество пикселей, для которых определено эталонное поле смещений D_{gt} . Такая методика оценки ошибки применяется в данной работе, значение порога $t = 3$.

Принципиально другим подходом представляется оценка ошибки посредством использования полученного поля диспаратета для деформации второго изображения. При этом деформированное изображение сравнивается с эталонным в некоторой норме и вычисляется энергия полученного сопоставления. В таком случае важно исключить из рассмотрения пиксели, соответствующие заслоненным объектам. При наличии эталонных полей смещений, данный подход не дает дополнительной информации по сравнению с ошибкой 3.51 и поэтому не используется в данной работе.

3.5.2.3 Процесс обучения.

Использованная коллекция изображений была разбита на тренировочную выборку (160 пар изображений) и валидационную выборку (40 пар изображений). На вход сверточной нейросети подавались трехканальные цветные изображения.

Метод	левый вид	оба вида	мед. фильтр
без агрегирования	58,67	39,37	28,58
без обучения	35,33	26,53	24,03
предложенный метод	8,33	6,12	6,05
метод [26]	35,14	25,03	32,67
метод [26] + тензор энергий (3.35)	13,42	10,44	11,79

Таблица 3.1: Количественное сравнение различных методов агрегирования энергий с использованием 40 изображений валидационной выборки из коллекции KITTI 2015.

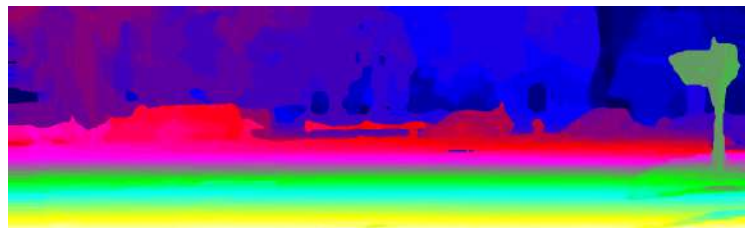
Тензор энергий соответствовал линейной комбинации абсолютного значения попиксельной разности и ценсус-преобразования 3.36 на основе фрагментов размера 7×7 . Оптимальное значение коэффициента линейной комбинации $\alpha = 0,43$ выбрано экспериментально методом линейного поиска. Для параметра σ выбрано значение 4, экспериментальная проверка показала, что значение этого параметра не оказывает заметного влияния на ошибку метода.

Исходная сверточная архитектура детектора границ была незначительно изменена. Во-первых, количество карт сверточных слоев было уменьшено наполовину для того, чтобы уменьшить вычислительную сложность. Во-вторых количество сверточных карт, используемых для финальной линейной комбинации было увеличено с 1 до 8 (рис. 3.6).

Предложенная сверточно-рекуррентная нейросеть была обучена с использованием кросс-энтропийной функции ошибки, используя адаптивную разновидность метода стохастического градиентного спуска, описанную в работе [48]. Использовалась величина шага градиентного метода, равная $2,5 \cdot 10^{-5}$. Используемый детектор границ был предобучен, используя коллекцию изображений BSDS [49]. При этом использованы эталонные границы, полученные методом ручной разметки.

Каждое изображение имеет размер 1242×375 , интервал допустимых значений диспаритета соответствует значениям $[0..256]$ пикселей. Вычисление стерео-сопоставления представляется наиболее трудным в участках изображений, не содержащих текстуры, включая однотонные автомобили, а так же в участках отражающий поверхностей, таких как стекла автомобилей. Визуализация вклада соседних пикселей при использовании обученного рекурсивного фильтра приведена на рис. 3.15.

Ошибка	Объекты	Фон	Все изображение
Все пиксели	4,97 %	1,38 %	4,47 %
Незаслоненные объекты	4,91 %	1,38 %	4,41 %



Ошибка	Объекты	Фон	Все изображение
Все пиксели	4,30 %	6,64 %	4,56 %
Незаслоненные объекты	4,01 %	6,64 %	4,30 %

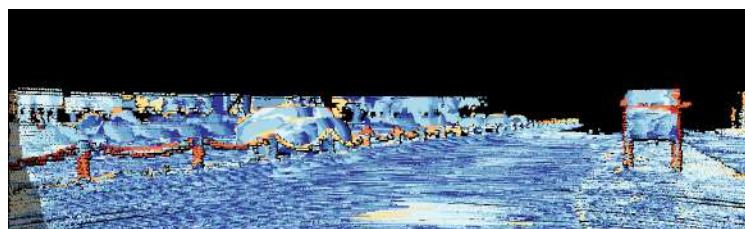
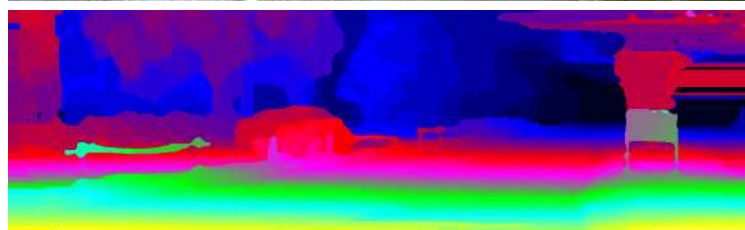
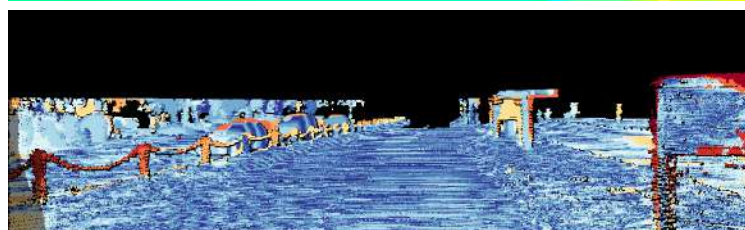
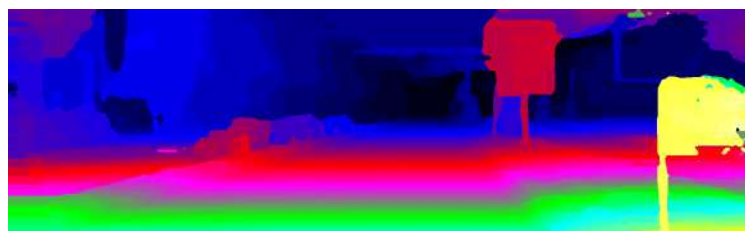


Рис. 3.11: Результаты предложенного метода сопоставления на коллекции изображений КИТТИ 2015 (тестовые изображения №0,1). Каждый рисунок содержит три изображения: ряд (1) изображение с левой камеры, (2) вычисленное поле диспаратета, (3) карта ошибки для поля диспаратета. Таблица содержит долю ошибочных пикселей с учетом и без учета заслоненных областей для областей соответствующих фону, передним объектам сцены, а так же для всего изображения.

Ошибка	Объекты	Фон	Все изображение
Все пикселы	5,68 %	9,09 %	5,84 %
Незаслоненные объекты	5,00 %	9,09 %	5,20 %



Ошибка	Объекты	Фон	Все изображение
Все пикселы	5,92 %	9,14 %	6,21 %
Незаслоненные объекты	5,45 %	9,14 %	5,80 %

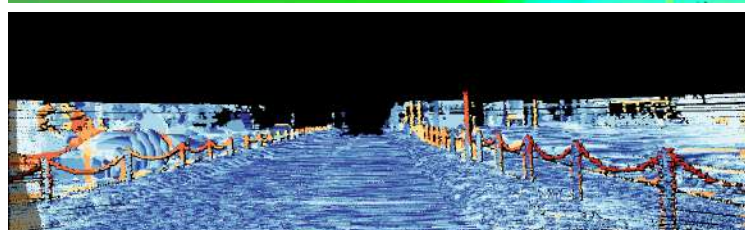
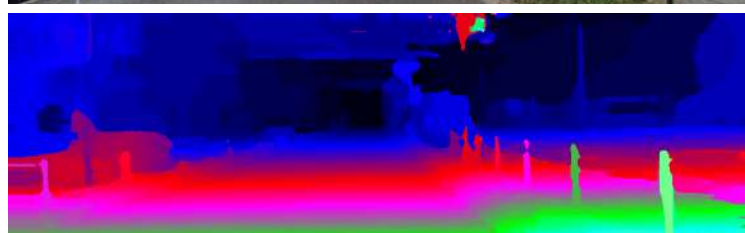
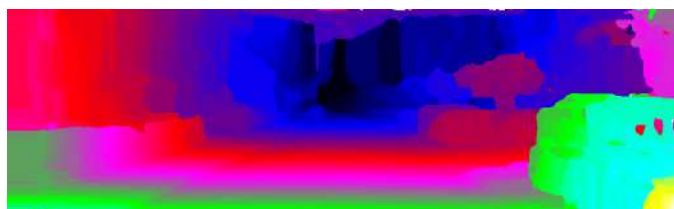


Рис. 3.12: Результаты предложенного метода сопоставления на коллекции изображений КИТТИ 2015 (тестовые изображения №2,3). Каждый рисунок содержит три изображения: ряд (1) изображение с левой камеры, (2) вычисленное поле диспаратета, (3) карта ошибки для поля диспаратета. Таблица содержит долю ошибочных пикселей с учетом и без учета заслоненных областей для областей соответствующих фону, передним объектам сцены, а так же для всего изображения.

Порог ошибки t	Незаслоненные объекты	Все объекты
3 пиксела	3,42 %	5,34 %
4 пиксела	2,98 %	4,35 %
5 пикселей	2,36 %	3,18 %



Порог ошибки t	Незаслоненные объекты	Все объекты
3 пиксела	3,08 %	5,22 %
4 пиксела	2,38 %	3,81 %
5 пикселей	2,06 %	2,87 %

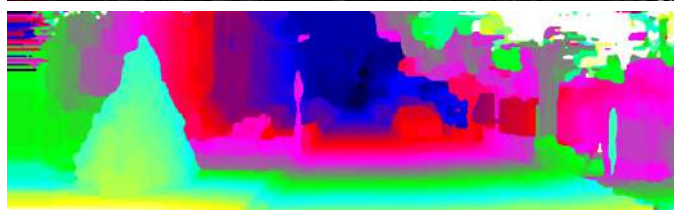
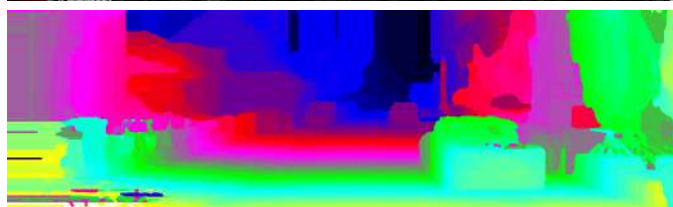


Рис. 3.13: Результаты предложенного метода сопоставления на коллекции изображений KITTI 2015 (тестовые изображения №0,1). Каждый рисунок содержит три изображения: ряд (1) изображение с левой камеры, (2) вычисленное поле диспаратета, (3) карта ошибки для поля диспаратета. Таблица содержит долю ошибочных пикселей с учетом и без учета заслоненных областей для областей для разных пороговых значений.

Порог ошибки t	Незаслоненные объекты	Все объекты
3 пиксела	11,37 %	12,19 %
4 пиксела	10,00 %	10,54 %
5 пикселей	9,05 %	9,52 %



Порог ошибки t	Незаслоненные объекты	Все объекты
3 пиксела	3,99 %	6,35 %
4 пиксела	2,81 %	4,76 %
5 пикселей	1,72 %	3,28 %

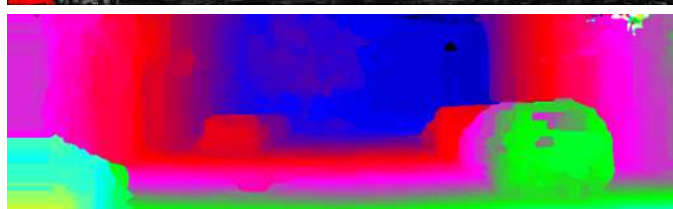


Рис. 3.14: Результаты предложенного метода сопоставления на коллекции изображений KITTI 2015 (тестовые изображения №2,3). Каждый рисунок содержит три изображения: ряд (1) изображение с левой камеры, (2) вычисленное поле диспаратета, (3) карта ошибки для поля диспаратета. Таблица содержит долю ошибочных пикселей с учетом и без учета заслоненных областей для областей для разных пороговых значений.



Рис. 3.15: Пример сглаживания с применением рекурсивного фильтра. Каждый из трех фрагментов справа, обозначенных разными цветами, показывает относительный вклад центрального пиксела в интенсивности остальных пикселей фрагмента. Центральный пиксел первого фрагмента соответствует дорожному столбу (процесс сглаживания ограничен границами столба ввиду перепада диспаритета). Фрагменты, обозначенные зеленым и красным цветом, являются примерами отток или границ дорожной поверхности, релевантных (красный цвет) или нерелевантных (зеленый цвет) перепадам диспаритета.

Метод	Объекты	Фон	Все изображение
Предложенный метод	5,34	11,35	6,34
метод [27]	8,43	18,51	10,11

Таблица 3.2: Сравнение предложенного метода и работы [27] на полном тестовом наборе коллекции KITTI 2015.

Таблица 3.1 содержит количественное сравнение методов агрегирования энергий. Была вычислена ошибка для трех случаев: ошибка метода на основе только левого вида, ошибка после постобработки на основе сопоставления диспаритета для левого и правого вида и ошибка после применения медианного фильтра размера 5×5 . Ошибка для метода [27] была вычислена для всей тестовой коллекции KITTI 2015, сравнение с предложенным методом приведено в таблице 3.2 (оба метода основаны на аналогичных тензорах энергии). Также приведена ошибка методов для случая описанного тензора энергии без применения агрегирования и ошибка метода до начала процесса обучения.

3.5.2.4 Оценка количества арифметических операций.

Наиболее трудоемким этапом алгоритма является агрегирование тензора энергий, его трудоемкость оценивается как $\mathcal{O}(nd_{max})$, где n - количество пикселей в изображении, а d_{max} - максимальный диспаритет.

Трудоемкость наиболее быстрого метода, основанного на машинном обучении, предложенного в литературе [9] оценивается как $\mathcal{O}(nd_{max}k)$, где k - размерность глубоких дескрипторов.

Этап	Кол-во вызовов	Время, мс
вычисление тензора энергий	2	3
сверточный детектор границ	1	9
рекурсивный фильтр	2	19
операция <i>argmin</i>	2	1
обнаружение заслоненных объектов	1	2
полное время		34

ТАБЛИЦА 3.3: Время исполнения параллельной реализации метода с использованием графического ускорителя NVIDIA GeForce GTX Titan X на изображения коллекции KITTI 2015. Процедура стерео-сопоставления была использована как для левого, так и для правого вида стерео-камеры с последующим обнаружением и интерполяцией заслоненных объектов. Выход сверточной нейросети вычисляется сразу для пары изображений. Наибольшая доля времени исполнения приходится на рекурсивный фильтр и сверточный детектор границ.

3.5.2.5 Время исполнения на графическом ускорителе.

Метод был реализован, используя комбинацию фреймворка Theano [50] и эффективной реализации процедур для вычисления тензора энергий на основе платформы CUDA. Реализация, применяемая на этапе исполнения использует библиотеку CuDNN для вычисления границ. Программная реализация процедуры рекурсивной фильтрации предложена автором. Обучение производилось на рабочей станции с использованием ускорителя NVIDIA Titan X, процесс обучения занимает 4-5 часов.

Время исполнения различных стадий предложенного метода показано в таблице 3.3. Полное время исполнения составляет 34 мсек, что соответствует частоте 29 кадров в секунду для пары изображений, включая издержки на сопоставление диспаратетов, полученных для правого и левого видов. Наибольшая доля времени исполнения соответствует сверточному детектору границ и рекурсивному фильтру.

3.5.2.6 Обнаружение заслоненных объектов.

Т.к. ошибка диспаратета считается для полных изображений, включая заслоненные объекты, представляется целесообразным вычислять маски заслоненных объектов на основе сравнения полей диспаратетов, вычисленных для правого и левого видов стерео-камеры [13]. Каждый из пикселей левого изображения получает одну из трех возможных меток $l(x, y) \in \{0, 1, 2\}$, которые соответствуют совпадению, ошибочному пикселу и заслоненному объекту, согласно следующему критерию, где $D_l(x, y)$ и $D_r(x, y)$ - оценки диспаратетов для левого и правого видов:

$$l(x, y) = \begin{cases} 0, & \text{если } |d - D^R(x - d, y)| \leq 1 \text{ для } d = D^L(x, y), \\ 1, & \text{если } |d - D^R(x - d, y)| \leq 1 \text{ для некоторого } d \in [0, d_{max}], \\ \text{иначе } 2. \end{cases} \quad (3.52)$$

Затем поле диспаратета интерполируется согласно следующему алгоритму. Диспаратет в пикселах с меткой 0 остается без изменений. В пикселах с меткой 2 - значение диспаратета берется из ближайшего пиксела с меткой 0, находящегося в той же строке изображения слева. Пикселы с меткой 1 получают значения ближайшего пиксела с меткой 0. Подобный алгоритм на практике позволяет существенно снизить ошибку сопоставления (см. табл. 3.16).

3.6 Выводы

Наиболее точные методы оценки полей диспаратета основаны на методах глубокого машинного обучения. Одним из факторов развития данного семейства методов является доступность больших коллекций обучающих данных, достаточных для обучения моделей с большим количеством параметров.

Большинство современных методов содержит следующие этапы сопоставления: вычисление тензора энергий, агрегирование тензора энергий и оптимизация поля диспаратета. При этом каждый из перечисленных этапов может быть использован в составе обучаемой модели. Сравнение дескрипторов большой размерности, вычисленные с использованием разновидностей Сиамской архитектуры сверточной нейросети, позволяют вычислять поля диспаратета с низкой ошибкой, однако, использование подобных подходов имеет высокую вычислительную сложность. Напротив, применение обучаемого агрегирования тензора энергий, предложенного автором, является более вычислительно эффективным подходом к построению обучаемой модели. Полная совокупность стерео-методов, предложенных в литературе, находится на кривой, соответствующей критериям: качество и время исполнения. Метод, предложенный автором в статье [11], является Парето-оптимальным по совокупности этих двух критериев, т.е. предложенные в литературе методы либо имеют большее время исполнения, либо имеют большую ошибку. В качестве направления для будущего исследования, наиболее перспективным представляется разработка метода, способного совместить обучение глубоких дескрипторов и предложенный метод агрегирования тензора энергий.

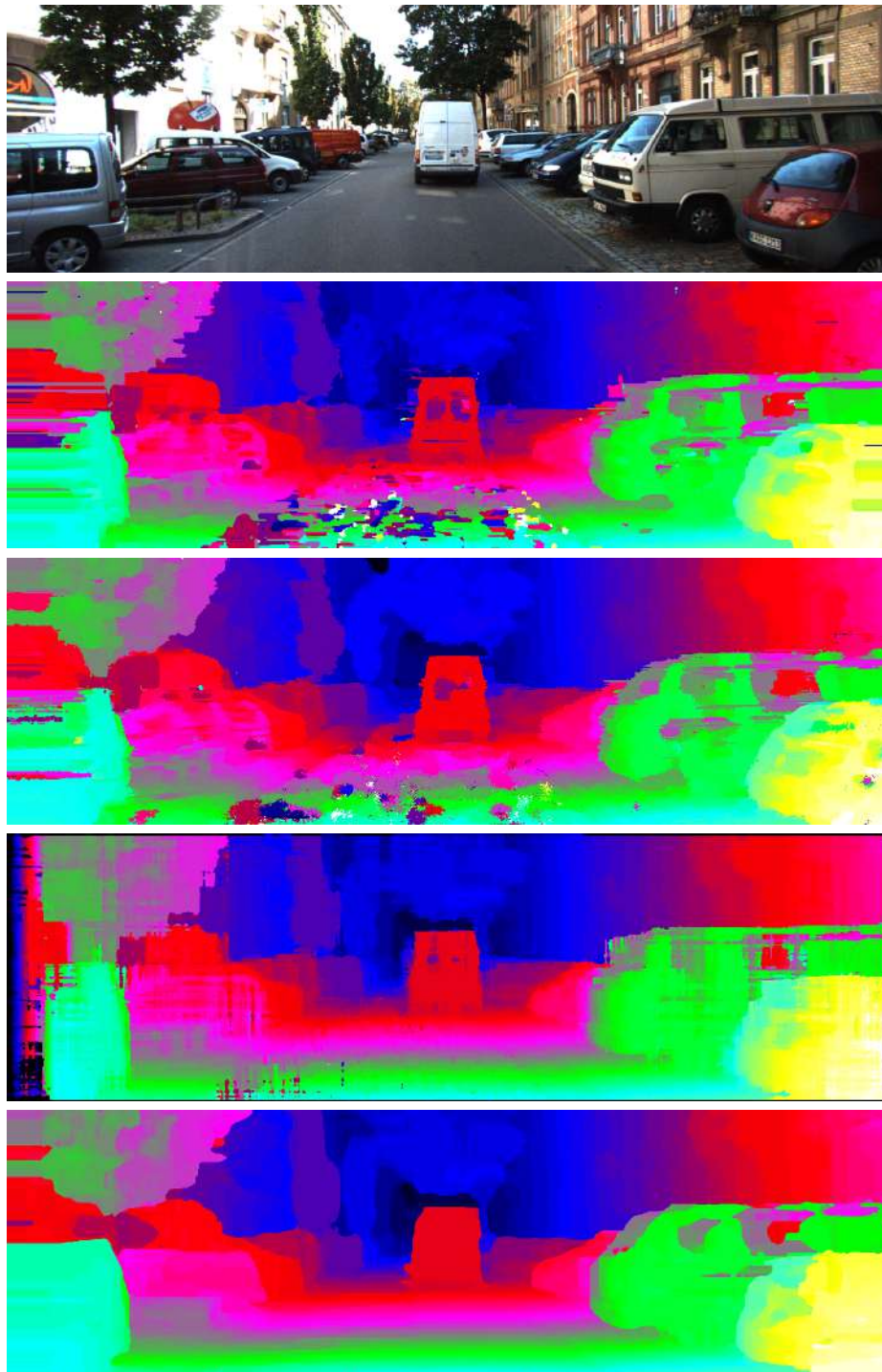


Рис. 3.16: Сравнение различных методов агрегирования тензора энергии с использованием тестового изображения из коллекции KITTI 2015. Изображения сверху-вниз: (1) – изображение с левого вида стерео-камеры, (2) – поле диспаратета, полученное методом [27], (3) – результат метода [26] при того же тензора энергий, что и в данной работе, (4) результат метода SGM [29] с использованием тензора энергий из данной работы, (5) – результат предложенного метода.

Глава 4

Задача вычисления оптического потока

Данная глава содержит обзор методов вычисления методов оптического потока, а также подробное описание подхода, предложенного автором в работе [12]. Глава начинается с описания алгоритмов на основе оптимизации функционалов, предложенных для задачи вычисления оптического потока (раздел 4.2). Целью данного описания является иллюстрация практических сложностей, которые встречаются при вычисления оптического потока. Раздел 4.3 содержит описание методов вычисления оптического потока на основе глубокого машинного обучения. Последующий раздел 4.6 описывает методы, использующие несколько кадров. Такие методы во многом сходны с методами, используемыми для многокадровой ультразвуковой эластографии, представленными в главе 5.

4.1 Постановка задачи

В задаче вычисления оптического потока требуется вычислить двумерное поле смещений для пары изображений $I_{0,1}$, определенных в области Ω . Оптический поток $\mathbf{u} = (u_x, u_y)$ имеет две компоненты u_x и u_y , соответствующих горизонтальным и вертикальным смещениям соответственно. Искомое поле смещений должно сопоставлять изображения пары таким образом, что для всех пикселей, $I_0(x, y)$ и $I_1(x + u_x, y + u_y)$ соответствуют одним и тем же точкам сцены. В отличие от задачи стерео-сопоставления, которая сводится к нахождению одномерного поля смещений, движение объектов на сцене может быть произвольным, поэтому поле смещений в задаче оптического потока является двумерным.

4.2 Обзор методов на основе минимизации функционалов энергии

Начиная с первой работы, в которой была рассмотрена задача вычисления оптического потока [17], методы для этой задачи оптического потока существенно эволюционировали, однако набор практических сложностей при работе с реальными данными остался прежним. Данный подраздел содержит описание работы [17] с иллюстрацией соответствующих недостатков метода, послуживших мотивацией к дальнейшему усовершенствованию алгоритмов для данной задачи.

4.2.1 Метод на основе инвариантной интенсивности и линеаризации изображения

Решением задачи оптического потока является двумерное поле смещений (u_x, u_y) , которое сопоставляет деформированную версию изображения I_1 и изображение I_2 . Одним из допущений, позволяющих решить данную задачу, является предположение об инвариантной интенсивности движущихся объектов, которое может быть записано следующим образом:

$$I_0(x, y, t) = I_1(x + u_x(x, y), y + u_y(x, y), t + \Delta t). \quad (4.1)$$

Для того, чтобы сформулировать задачу оценки оптического потока в виде оптимизационной задачи, введем следующий функционал, отвечающий за соответствие оценки оптического потока входным изображениям (особенности квадратичной целевой функции обсуждаются далее):

$$E_{data}(\mathbf{u}) = \sum_{x,y} (I_1(x + u_x(x, y), y + u_y(x, y)) - I_0(x, y))^2. \quad (4.2)$$

Такой функционал, зависящий от переменных (u_x, u_y) , является невыпуклым, что не позволяет построить эффективный метод минимизации без использования дополнительных ограничений, таких как пространственная гладкость.

Еще одним важным свойством такого функционала является то, что он может быть минимизирован для каждого пиксела изображения независимо. Такое свойство позволяет использовать метод полного перебора при ограничении абсолютной величины смещений. Однако, решения полученные независимо для каждого пиксела,

содержат большое количество выбросов, что не позволяет построить метод с низкой ошибкой без использования пространственной регуляризации.

Вычислительная сложность полного перебора вектором смещений может быть существенно снижена при использовании линейной аппроксимации изображения [17] в малой окрестности текущей оценки оптического потока:

$$I(x + u_x(x, y), y + u_y(x, y), t + \Delta t) \approx I(x, y, t) + \frac{\partial I}{\partial x} u_x(x, y) + \frac{\partial I}{\partial y} u_y(x, y) + \frac{\partial I}{\partial t} \Delta t. \quad (4.3)$$

Используя такое линейное приближение, исходный функционал может быть записан следующим образом:

$$E_{data}(u, v) = \int_{\Omega} \left(\frac{\partial I}{\partial x} u_x(x, y) + \frac{\partial I}{\partial y} u_y(x, y) + I_t \right)^2 d\mathbf{x}, \quad (4.4)$$

или используя вектор \mathbf{u} :

$$E_{data}(\mathbf{u}) = \int_{\Omega} (\nabla I \cdot \mathbf{u} + I_t)^2 d\mathbf{x}. \quad (4.5)$$

Рассмотрим скалярное произведение $\nabla I \cdot \mathbf{u}$ и заметим, что значение функционала не изменится, если прибавить к вектору \mathbf{u} произвольный вектор, ортогональный градиенту изображения ∇I . Как следствие, оптический поток может быть оценен с точностью до вектора, ортогонального градиенту изображения. Таким образом, использование такого функционала недостаточно для оценки оптического потока в участках малой величины градиента изображения. Однако, возможно избежать подобной трудности при использовании пространственной регуляризации.

Одним из подходов к регуляризации поля движений является наложение штрафа на квадрат пространственного изменения поля оптического потока [17]:

$$E_{reg}(\mathbf{u}) = \int_{\Omega} \alpha \left(\|\nabla u_x\|_2^2 + \|\nabla u_y\|_2^2 \right) d\mathbf{x}, \quad (4.6)$$

будем называть такой подход $L2$ регуляризацией. Такой способ аналогичен регуляризации Тихонова, накладываемой на пространственные производные полей смещений.

Так, оценка оптического потока $\mathbf{u} = (u_x, u_y)$ может быть получена путем минимизации следующего функционала:

$$J(u, v) = \int_{\Omega} (I_0(x, y) - I_1(x + u_x(x, y), y + u_y(x, y)))^2 + \alpha (\|\nabla u_x\|_2^2 + \|\nabla u_y\|_2^2) \, d\mathbf{x}. \quad (4.7)$$

где $\mathbf{x} = (x, y)$. Необходимое условие на минимум такого функционала записывается при помощи следующих уравнений Эйлера-Лагранжа:

$$\begin{cases} 0 = -(I_0(\mathbf{x}) - I_1(\mathbf{x} + \mathbf{u})) \frac{\partial I_1}{\partial x}(\mathbf{x} + \mathbf{u}) - \alpha^2 \operatorname{div}(\nabla u_x) \\ 0 = -(I_0(\mathbf{x}) - I_1(\mathbf{x} + \mathbf{u})) \frac{\partial I_1}{\partial y}(\mathbf{x} + \mathbf{u}) - \alpha^2 \operatorname{div}(\nabla u_y) \end{cases}. \quad (4.8)$$

При использовании описанной выше линейризации изображения, система уравнений может быть записана следующим образом:

$$\begin{cases} 0 = -(I_0(\mathbf{x}) - I_1(\mathbf{x} + \mathbf{u}^n) - \nabla I_1(\mathbf{x} + \mathbf{u}^n)(\mathbf{u}^{n+1} - \mathbf{u}^n)) \frac{\partial I_1}{\partial x}(\mathbf{x} + \mathbf{u}^n) - \alpha^2 \operatorname{div}(\nabla u_x^n) \\ 0 = -(I_0(\mathbf{x}) - I_1(\mathbf{x} + \mathbf{u}^n) - \nabla I_1(\mathbf{x} + \mathbf{u}^n)(\mathbf{u}^{n+1} - \mathbf{u}^n)) \frac{\partial I_1}{\partial y}(\mathbf{x} + \mathbf{u}^n) - \alpha^2 \operatorname{div}(\nabla u_y^n) \end{cases}. \quad (4.9)$$

После дискретизации, такая система уравнений получает вид разреженной линейной системы, которая может быть решена итерационным методом, используя приближение \mathbf{h}^n , полученное на предыдущем шаге.

4.2.2 Многомасштабная оценка оптического потока

Для того, чтобы оценить большие смещения, используются методы оценки оптического потока на основе нескольких масштабов. При этом оптический поток на более крупных масштабах оценивается для пары уменьшенных изображений $I_{1,2}$, а полученный результат используется на более мелких масштабах после билинейной интерполяции. В подобном подходе пирамида изображений соответствует набору используемых масштабов $s = 0, 1, \dots, N_s - 1$. Полное число масштабов обозначим как N_s . Рассмотрим пару последовательных масштабов s и $(s - 1)$: интенсивности соответствующих изображений I_s и I_{s-1} связаны следующим соотношением:

$$I^s(\eta x, \eta y) = G_{\sigma} * I^{s-1}(\eta x, \eta y), \quad (4.10)$$

где G_σ - Гауссовский фильтр, используемый для того, чтобы получить сглаженное изображение на более крупном масштабе, а $\eta \in (0, 1)$ - пропорция между последовательными масштабами. При этом значение параметра σ выбирается в зависимости от значения η . Начиная с нулевого начального значения на первом масштабе, оценка оптического потока используется для следующего масштаба после интерполяции и шкалирования, которое может быть записано следующим образом:

$$u_x^{s-1}(x, y) = \frac{1}{\eta} u_x^s(\eta x, \eta y), \quad (4.11)$$

$$u_y^{s-1}(x, y) = \frac{1}{\eta} u_y^s(\eta x, \eta y). \quad (4.12)$$

Многомасштабная схема позволяет оценивать поля смещений большей величины с меньшей ошибкой. На практике, в зависимости от ожидаемой скорости движения объектов, требуется ввести достаточное количество масштабов N_s , так что величина смещений на масштабе $s = 0$ ограничена. Однако, подобный подход обладает фундаментальным недостатком: более мелкие объекты, претерпевающие значительные смещения не представлены на крупных масштабах - как следствие оптический поток не может быть оценен корректно для таких объектов.

4.2.3 Методы регуляризации.

Выбор регуляризационного функционала представляется одним из наиболее важных аспектов при построении методов вычисления оптического потока. При этом метод регуляризации выбирается в соответствии с ожидаемыми свойствами оцениваемых полей смещений.

Часто, используемые методы регуляризации в вариационных методах оценки оптического потока повторяют набор методов регуляризации, используемых для задачи шумоподавления. В рассмотренной выше работе [17] была использована регуляризация Тихонова, заданная соотношением 4.6. Однако, использование такого функционала ведет к переглаженным решениям ввиду того, что единичные выбросы ведут к значительным увеличениям значения целевого функционала [51].

4.2.3.1 Полная вариация

Одним из способов, позволяющими избежать пересглаженных решений в задаче [17] является использованием в качестве регуляризации функционала полной вариации

(total variation), предложенного в работе [51]. Подобный подход позволяет корректно производить реконструкцию пространственных разрывов поля смещений.

Полная вариация для изображения $u(x, y)$ может быть определена следующим образом:

$$\text{TV}(u) = \int_{\Omega} \sqrt{\frac{\partial u}{\partial x}^2 + \frac{\partial u}{\partial y}^2} d\mathbf{x}. \quad (4.13)$$

Такая функция изотропна и не является гладкой. Анизотропная версия полной вариации может быть определена в следующем виде:

$$\text{TV}_{aniso}(u) = \int_{\Omega} \left(\left| \frac{\partial u}{\partial x} \right| + \left| \frac{\partial u}{\partial y} \right| \right) d\mathbf{x}. \quad (4.14)$$

4.2.3.2 Двойственный градиентный метод для задачи шумоподавления

Эта секция описывает метод минимизации функционала полной вариации, которая используется для построения предложенного автором метода [12], описанного в разделе 4.4.

Сформулируем задачу шумоподавления в виде следующей оптимизационной задачи, где в качестве регуляризации используется функционал полной вариации:

$$E_p(x) = \min_x \frac{1}{2} \|y - x\|_2^2 + \lambda \text{TV}(x), \quad (4.15)$$

Где y - шумное изображение, x - сглаженная версия и $\|\cdot\|_{1,2}$ - смешанная $l_{1,2}$ норма:

$$\|x\|_{1,2} = \sum_{k=1}^K \left\{ \sum_{l=1}^L \|x_{k,l}\|^2 \right\}^{1/2} = \sup_{\|y\|_{\infty,2} \leq 1} \langle x, y \rangle. \quad (4.16)$$

Используя двойственное определение нормы ($\omega = (w_1, w_2)$ - двойственная переменная), целевая функция может быть записана в виде:

$$\min_x \max_{\|\omega\|_{\infty,2} \leq 1} \frac{1}{2} \|y - x\|_2^2 + \lambda \langle x, \nabla^* \omega \rangle, \quad (4.17)$$

либо в следующей форме:

$$\max_{\|\omega\|_{\infty,2} \leq 1} \min_x \frac{1}{2} x^T x - x^T (y - \lambda \nabla^* \omega) + \frac{1}{2} y^T y. \quad (4.18)$$

Итак, целевая функция может быть записана в виде:

$$\max_{\|\omega\|_{\infty,2} \leq 1} \min_x \frac{1}{2} \|x - (y - \lambda \nabla^* \omega)\|^2 + \frac{1}{2} \|y\|_2^2 - \frac{1}{2} \|y - \lambda \nabla^* \omega\|_2^2. \quad (4.19)$$

Оптимальным значением x^* для переменной x является выражение $x^* = y - \lambda \nabla^* \omega$. После такой подстановки, двойственная задача может быть записана в следующем виде:

$$E_d(\omega) = \max_{\|\omega\|_{\infty,2} \leq 1} \frac{1}{2} \|y\|_2^2 - \frac{1}{2} \|y - \lambda \nabla^* \omega\|_2^2. \quad (4.20)$$

Эта задача может быть решена методом градиентного подъема с последующим перепроектированием на единичный шар:

$$\nabla_{\omega} f(\omega) = \lambda \nabla(\lambda \nabla^* \omega - y), \quad (4.21)$$

$$\omega_{k+1} = P_{\infty,2}(\omega_k - \lambda \nabla(\lambda \nabla^* \omega - y)), \quad (4.22)$$

где $P_{\infty,2}$ - попиксельная проекция на единичный шар в L_2 норме:

$$P_{\infty,2}(\omega) = (P_2(\omega_1), \dots, P_2(\omega_N)). \quad (4.23)$$

Итак, задача шумоподавления, где в качестве регуляризации выступает полная вариация, может быть решена с использованием двойственного градиентного метода. В качестве критерия сходимости используется разница между значением целевых функций для прямой и двойственной задачи.

4.2.3.3 Обобщение полной вариации и условие гладкости второго порядка

Функционал полной вариации, описанный в предыдущем подразделе, ведет к кусочно-постоянным решениям, что не всегда является желательным на практике. Для случаев, когда такое условие на решения не выполнено, имеет смысл рассмотреть функционал второго порядка. В этом случае, минимизация функционала порождает кусочно-линейные решения. Такие решения являются более естественными с точки зрения визуального анализа, производимого человеком. Подобное обобщение полной вариации на гладкость второго порядка было предложено в работах [52, 53]:

$$\text{TGV}_\alpha^2(u, v) = \min_v \alpha_1 \int_\Omega |\nabla u - v| d\mathbf{x} + \alpha_0 \int_\Omega |\varepsilon(v)| d\mathbf{x}. \quad (4.24)$$

Минимум находится среди всех векторных полей v , определенных в области Ω . При этом оператор второй производной ε определяется следующим образом:

$$\varepsilon(v) = \frac{1}{2}(\nabla v + \nabla v^T). \quad (4.25)$$

Вводится дискретизация операторов div_2^h , ε^h , ∇^h , div_1^h с использованием метода конечных разностей:

$$(\text{div}_2^h)^* = -\varepsilon^h, \quad (4.26)$$

$$(\text{div}_1^h)^* = -\nabla^h. \quad (4.27)$$

Такой функционал регуляризации применяется для решения обратных задач, таких как, например, МРТ реконструкция [54].

4.2.3.4 Модель для обучаемой регуляризации на основе случайных марковских полей

Важным шагом в развитии методов регуляризации был подход, предложенный в работе [55]. Авторами была разработана общая конструкция для построения обучаемой регуляризации для задач восстановления изображений. Метод совмещает идеи из метода разреженного кодирования [56] и методов на основе случайных Марковских полей. Подход позволяет выявлять закономерности, возникающие между интенсивностями соседних пикселей реальных изображений и использовать при решении ряда задач компьютерного зрения. Метод позволяет обобщить традиционные модели на основе случайных Марковских полей путем обучения функций потенциалов для расширенных окрестностей. Потенциалы смоделированы с использованием модели [57], которая использует нелинейное преобразование большого количества откликов на различные фильтры. Метод применяется в задачах шумоподавления и прочим задачам восстановления изображений с использованием приближенной схемы вывода.

4.2.4 Метод для оценки больших смещений

Общая для многих методов вычисления оптического потока многомасштабная схема упрощает исходную задачу, т.к. поля смещений оцениваются на более крупном масштабе. Полученный результат используется для инициализации алгоритма поиска смещений на более мелком масштабе, что позволяет получить обновленное решение для более мелких элементов на изображении. Следствием такого подхода является неспособность корректно оценить поля смещений в тех случаях, когда более мелкие объекты претерпевают большие смещения. Причиной такого недостатка является отсутствие мелких деталей на более крупных масштабах, что ведет к некорректной инициализации алгоритма на последующих масштабах и попаданию в локальный минимум при оптимизации функционала энергии.

Метод оценки больших смещений, описанный в работе [58] использует разреженное сопоставление на основе визуальных дескрипторов для того, чтобы более точно оценивать движение более мелких фрагментов изображения. В данной работе был предложен способ совместить вариационные методы с методами на основе сопоставления визуальных дескрипторов за счет введения в функционал энергии специального члена [58].

4.2.5 Интерполяция соответствий с учетом границ изображения

Важным шагом в разработке алгоритмов вычисления оптического потока стал метод, описанный в работе [59]. Подход использует разреженную оценку оптического потока, описанную в работе [60]. Плотное поле смещений получено из разреженной оценки путем интерполяции с учетом границ изображения. Полученный результат служит инициализацией для последующего вариационного метода минимизации.

Наиболее важным аспектом предложенного метода является использование метода интерполяции, который основан на использовании геодезического расстояния. Такой подход позволяет аккуратно отражать границы движущихся объектов и эффективно решать проблему заслоненных объектов.

Для оценки геодезического расстояния используется вычислительно эффективное приближение [59]. Такое расстояние может быть определено как кратчайшее расстояние между двумя пикселями изображения по отношению к некоторой карте стоимостей, которая для данной задачи вычисляется из границ изображения. Так, пиксел, принадлежащий непрерывно движущемуся сегменту является близким ко всем пикселям сегмента, но при этом является достаточно далеким по отношению к

пикселям за границей движущегося сегмента. Метод основан на допущении о том, что границы движущихся объектов являются подмножеством границ изображения.

4.3 Обзор методов, основанных на глубоком машинном обучении

Методы вычисления оптического потока на основе глубокого машинного обучения, предложенные в литературе, можно условно разделить на следующие три категории. В первую категорию входят методы, основанные на сопоставлении глубоких дескрипторов изображений, аналогичные методам стерео-сопоставления. При этом результат сопоставления используется для построения задачи оптимизации, решением которой является искомое поле оптического потока [10]. Другая категория методов основана на подходе PatchMatch и его модификациях [61]. К третьей категории методов относится модель [3] и ее модификация [4], которые основаны на вариантах многомасштабной сверточной архитектуры для попиксельного предсказания оптического потока.

4.3.1 Сверточная нейросеть для задачи оптического потока и модель FlowNet

Модель FlowNet [3] является общей архитектурой, которая использует сверточную нейросеть для попиксельного предсказания оптического потока. Прочие применения подобного подхода включают в себя семантическую сегментацию, предсказание глубины на основе одного кадра и детекцию границ.

Итак, нейросеть принимает на вход пару изображений и вычисляет значение оптического потока $\mathbf{u} = (u_x, u_y)$. При этом в работе предложено два варианта архитектуры [3]. Первый вариант (*FlowNetSimple*) является общей сверточной архитектурой, не содержащей операций, специфических для данной задачи. Такая нейросеть способна выучить функцию вычисления оптического потока на основе большого количества тренировочных данных. Однако при этом нет гарантий, что метод локальной оптимизации, основанный на градиентном спуске позволит методу сойтись к желаемому решению. Поэтому, был предложен второй вариант модели (*FlowNetCorr*), который содержит корреляционный слой, введенный специально для задачи сопоставления изображений. Идея метода состоит в том, чтобы добавить в нейросеть операцию корреляции, широко используемую при сопоставлении изображений. Выход такого слоя для пары фрагментов изображения с центрами в (x_1, y_1) и (x_2, y_2) вычисляется следующим образом:

$$c(x_1, y_1, x_2, y_2) = \sum_{(o_x, o_y) \in [-k, k] \times [-k, k]} \langle \mathbf{f}_1(x_1 + o_x, y_1 + o_y), \mathbf{f}_2(x_2 + o_x, y_2 + o_y) \rangle. \quad (4.28)$$

Предшествующая и последующая корреляционному слою часть нейросети используется для обучения параметров. В качестве функции потерь используется Евклидова ошибка для двумерного поля смещений, определенного на всем изображении. Важным аспектом архитектуры является использование операций пулинга, которые позволяют оценивать поля смещений для изображений меньшего размера. При этом используются операции, аналогичные декодирующей части сверточных автокодировщиков, которые позволяют привести полученные поля смещений к оригинальному разрешению. Такой подход позволяет получить сглаженные поля оптического потока без необходимости постобработки.

4.4 Обучаемая регуляризация для метода вычисления оптического потока в реальном времени

Одним из наиболее важных аспектов построения алгоритмов вычисления оптического потока представляется выбор функционала регуляризации. Начиная с классической работы [17], использующей L^2 норму в качестве регуляризации, методы регуляризации претерпели значительную эволюцию. Важным шагом в разработке таких методов было использование полной вариации (total variation [51]), при этом были разработаны эффективные методы минимизации для такого функционала [2, 62]. Позднее были предложены методы регуляризации второго порядка [52, 54], также применяемые в задаче стерео-сопоставления [63, 64]. Важным этапом развития методов сопоставления изображений является использование нелокальных методов регуляризации на основе полной вариации первого и второго порядка [65, 66].

Развитие вариационных методов для вычисления оптического потока связано с разработкой эффективных методов оптимизации. При этом набор функционалов, используемых в задаче, ограничен возможностью представить задачу в виде минимизации одного или последовательности выпуклых функционалов [67]. Метод основанный на применении сверточной нейросети, предложенный в данной статье позволяет обучать оператор регуляризации с использованием тренировочной выборки из пар изображений, для которых предоставлены эталонные значения оптического потока. Идея построения обучаемой модели состоит в представлении итераций оптимизационного алгоритма в виде слоев сверточной нейросети. Впервые подобные

методы были предложены для задачи шумоподавления в работах [68, 69]. Также обучаемая регуляризация на основе случайных марковских полей была предложена в работе [55]. Наиболее схожая работа из представленных в литературе [70] предлагает обучение проксимальной нейросети в качестве постобработки для метода вычисления оптического потока, однако не позволяет получить целиком обучаемую модель.

Метод, предложенный автором, основывается на представлении итераций TV-L1 алгоритма [2, 62] в качестве слоев сверточной нейросети, что позволяет получить обучаемый оператор регуляризации, который является частью быстрого метода вычисления оптического потока.

Вклад данной работы состоит из следующих частей. Была предложена новая архитектура сверточной нейросети для задачи оптического потока, которая имеет низкую вычислительную сложность и потребление памяти на этапе исполнения. Также был предложен способ обучения оператора регуляризации для задачи оптического потока. При этом оператор регуляризации выражается как серия операций, содержащая свертки с фильтрами, обучаемыми методом обратного распространения ошибки. В статье приводятся результаты численных экспериментов, показывающих уменьшение ошибки, достигаемое за счет обучения оператора регуляризации.

4.4.1 Двойственный метод оптимизации для вычисления оптического потока в реальном времени

Данный раздел описывает метод вычисления оптического потока, основанный на двойственном методе оптимизации, предложенный в работах [2, 62]. Следующий раздел описывает обобщение этого метода, предложенное в данной работе.

Постановка задачи. Рассмотрим два черно-белых изображения, заданных интенсивностями $I_0(x, y)$ и $I_1(x, y)$, определенных в области Ω . При этом значение интенсивностей каждого из пикселей с координатами (x_0, y_0) обозначено как $I_{0,1}(x_0, y_0)$ и задано целыми числами от 0 до 255. Обозначим как $\mathbf{x} = [x, y]$ - вектор координат. Целью является оценка двумерного поля смещения $\mathbf{u}: \Omega \rightarrow \mathbb{R}^2$, такого что пиксели $I_0(x, y)$ и $I_1(x + u_x, y + u_y)$ соответствуют одним и тем же точкам сцены. Двумерный вектор смещений обозначим как $\mathbf{u} = (u_x, u_y)$.

Функционал энергии. Рассмотрим формулировку задачи вычисления оптического потока в виде оптимизационной задачи. Целью исследования является построение сверточной архитектуры нейросети, обладающей низкой вычислительной сложностью на этапе исполнения. Выбор оптимизационного метода продиктован целью

представить его итерации в виде слоев нейросети, так что переиспользованы эффективные реализации операций для построения сверточных нейросетей. Метод, наиболее соответствующий выбранной цели, это двойственный метод, предложенный в работе [2] и использующий функционал полной вариации в качестве регуляризации. Рассмотрим данный метод более подробно.

Целевой функционал состоит из следующих трех членов:

$$E(I_0, I_1, \mathbf{u}, \mathbf{v}) = \int_{\Omega} \{\lambda |\rho(\mathbf{v})| + \sum_{d=x,y} \frac{1}{2\theta} (u_d - v_d)^2 + \|\nabla \mathbf{u}\|_{1,2}\} d\mathbf{x}. \quad (4.29)$$

Первый член $\lambda |\rho(\mathbf{v})|$ отвечает за соответствие оптического потока входным данным, на практике рассматривается следующее линейное приближение (\mathbf{u}_0 - некоторая первоначальная оценка поля смещений):

$$\rho(\mathbf{u}, \mathbf{u}_0, x, y) = \|I_1(x + u_x, y + u_y) - I_0(x, y)\|_1 \approx I_1(x + u_{0x}, y + u_{0y}) + \langle \nabla I_1, \mathbf{u} - \mathbf{u}_0 \rangle - I_0(x, y). \quad (4.30)$$

При этом для изображения I_1 используется линейное приближение в окрестности первоначального приближения $(x + u_{0x}, y + u_{0y})$, ∇I_1 - градиент изображения.

Параметр λ - константа, которая позволяет варьировать соотношение между входными данными и регуляризацией, изменяя таким образом степень сглаживания полученных решений. Третий член $\|\nabla \mathbf{u}\|_{1,2}$ является функционалом регуляризации, второй член вводится для разделения первого и третьего члена. Целью такого разделения является минимизация исходного функционала посредством разбиения на две подзадачи. Первая подзадача состоит в минимизации суммы первого и второго члена исходного функционала, а вторая подзадача - второго и третьего члена.

При этом, вводится вспомогательное поле смещений \mathbf{v} , которое соответствует в процессе оптимизации более шумной версии поля смещений \mathbf{u} . Параметр θ является константой, которая регулирует степень взаимного влияния полей \mathbf{u} и \mathbf{v} в процессе решения задачи. Итак, минимизация выполняется в виде решения двух подзадач. Далее описаны оба шага минимизации.

Подзадача 1. Для каждого d при фиксированном u_d , решается следующая подзадача:

$$\min_{\mathbf{v}} \int_{\Omega} \{\lambda |\rho(\mathbf{v})| + \sum_{d=x,y} \frac{1}{2\theta} (u_d - v_d)^2\} d\mathbf{x}. \quad (4.31)$$

Описанная подзадача решается независимо для каждого пиксела, при этом решение записывается в следующем виде [2]:

$$\mathbf{v} = \mathbf{u} + \begin{cases} \lambda\theta\nabla I_1, & \text{если } \rho(\mathbf{u}) < -\lambda\theta(\nabla I_1)^2 \\ -\lambda\theta\nabla I_1, & \text{если } \rho(\mathbf{u}) > \lambda\theta(\nabla I_1)^2 \\ -\rho(\mathbf{u})/\nabla I_1, & \text{если } |\rho(\mathbf{u})| \leq \lambda\theta(\nabla I_1)^2. \end{cases} \quad (4.32)$$

Такое решение получается попиксельно путем раскрытия модуля в выражении $|\rho(\mathbf{v})|$ и последующей подстановкой линейного приближения 4.30 в функционал 4.31. Для краткости введем для выражения 4.32 следующую запись:

$$\mathbf{v} = T(\mathbf{u}, \nabla I_1, I_t). \quad (4.33)$$

Такая операция является пороговым преобразованием (thresholding) входной переменной \mathbf{u} . Временная производная изображения I_t вычисляется как $I_t = I_1(x + u_{0x}, y + u_{0y}) - I_0(x, y)$ в соответствии с первым и третьим членом в выражении (4.30).

Градиент изображения $\nabla I_1(x + u_{0x}, y + u_{0y}) = \left[\frac{\partial I_1}{\partial x}(x + u_{0x}, y + u_{0y}), \frac{\partial I_1}{\partial y}(x + u_{0x}, y + u_{0y}) \right]$ вычисляется с использованием операции деформации (warping) следующим образом:

$$\frac{\partial I_1}{\partial x} = \frac{W(I_1, u_{0x} + \Delta x, u_{0y}) - W(I_1, u_{0x} - \Delta x, u_{0y})}{2\Delta x}, \quad (4.34)$$

$$\frac{\partial I_1}{\partial y} = \frac{W(I_1, u_{0x}, u_{0y} + \Delta y) - W(I_1, u_{0x}, u_{0y} - \Delta y)}{2\Delta y}, \quad (4.35)$$

где $W(I, u_x, u_y)$ - результат деформации изображения I с использованием двумерного поля смещений \mathbf{u} , а Δx и Δy - малые возмущения горизонтальной и вертикальной компоненты оптического потока u_{x0} и u_{y0} , используемые для вычисления градиента с использованием центральной разностной схемы.

Подзадача 2. Для фиксированного \mathbf{v} решается следующая подзадача:

$$\min_{\mathbf{u}} \int_{\Omega} \left\{ \sum_{d=x,y} \frac{1}{2\theta} (u_d - v_d)^2 + \|\nabla \mathbf{u}\|_{1,2} \right\} dx. \quad (4.36)$$

Решение может быть получено с помощью двойственного метода градиентного спуска с перепроектированием на единичный шар, описанного в работе [62]. Итерация метода записывается следующим образом:

$$\mathbf{u} = \mathbf{v} - \theta \nabla^* \mathbf{p}, \quad (4.37)$$

где $\mathbf{p} = (p_1, p_2)$ есть двойственная переменная (переменные $p_{1,2}$ соответствуют двумерным полям смещений, определенное в области Ω) и ∇^* есть оператор, сопряженный к оператору градиента ∇ . Шаг двойственного метода производится с использованием градиентного подъема с последующим перепроектированием на шар, соответствующий единичной норме:

$$\mathbf{p}^{n+1} = \mathbf{P}(\mathbf{p}^n + (1/\theta)\nabla \mathbf{v}), \quad (4.38)$$

где операция проекции определена следующим образом:

$$\mathbf{P}(\mathbf{p}) = \frac{\mathbf{p}}{\max(1, \|\mathbf{p}\|_2)}. \quad (4.39)$$

4.5 Архитектура нейросети.

В соответствии с методом оптимизации, описанным в предыдущем подразделе, определим архитектуру нейросети, которая будет состоять из модуля деформации и модуля итерации двойственного метода (рис. 4.1). Нейросеть получает на вход пару изображений $I_{0,1}$ и вычисляет оптический поток, представленный в виде двухканального изображения. Нейросеть обучается на основе размеченных пар изображений, для которых определены эталонные значения оптического потока. При этом в качестве функции потерь используется Евклидова ошибка для двумерных векторов смещений, усредненная по изображению. Результат, полученный нейросетью, не требует дополнительной постобработки.

Оптический поток вычисляется с использованием многомасштабного подхода. Начиная с наиболее крупного масштаба, для которого нейросеть получает на вход нулевое приближение, полученная оценка оптического потока масштабируется с использованием билинейной интерполяции и шкалируется. Результат вычисления используется на более и более мелком масштабе, вплоть до исходного разрешения.

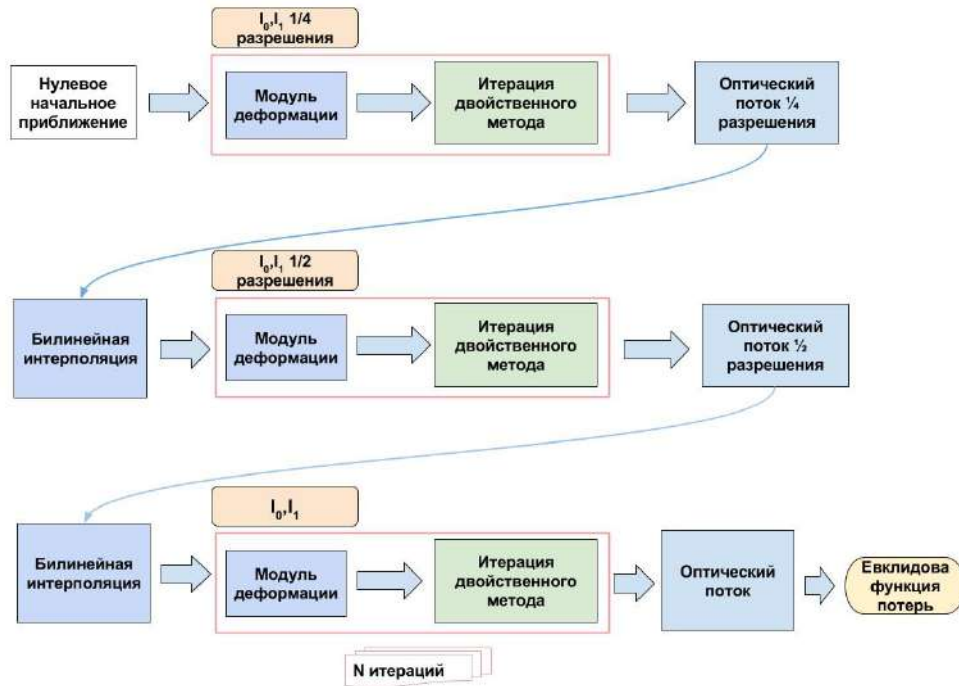


Рис. 4.1: Архитектура нейросети для задачи оптического потока. Сверточная нейросеть вычисляет оптический поток последовательно, начиная с нулевого приближения. При этом используется многомасштабный подход: поле смещений оценивается на изображениях уменьшенного масштаба, и затем оценка используется в качестве начального приближения для вычисления на следующем масштабе. Для получения поля смещений на большем масштабе используется билинейная интерполяция. Нейросеть для каждого из масштабов состоит из повторяющихся комбинаций из пары модулей: модуля деформации изображения и модуля итерации двойственного метода. С целью эффективной визуализации, на рисунке показана пирамида из 3 масштабов с коэффициентами масштабирования 2. В численных экспериментах используется большее количество масштабов, а также меньшие значения коэффициентов масштабирования с целью увеличения качества метода.

Модуль деформации. Модуль деформации использует операцию порогового преобразования (4.33) как слой нейросети. Модуль принимает на вход пару изображений $I_{0,1}$ на соответствующем масштабе и начальную оценку поля смещений u_0 (см рис. 4.2). Для вычисления градиента используется процедура деформации изображения и центральная разностная схема, описанная выше. Для того, чтобы обучать нейросеть методом обратного распространения ошибки, для вычисления деформированного изображения используется слой пространственной деформации (spatial transformer layer), предложенный в работе [71]:

$$\frac{\partial I_1}{\partial x} = \frac{S(I_1, u_{0x} + \Delta x, u_{0y}) - S(I_1, u_{0x} - \Delta x, u_{0y})}{2\Delta x}, \quad (4.40)$$

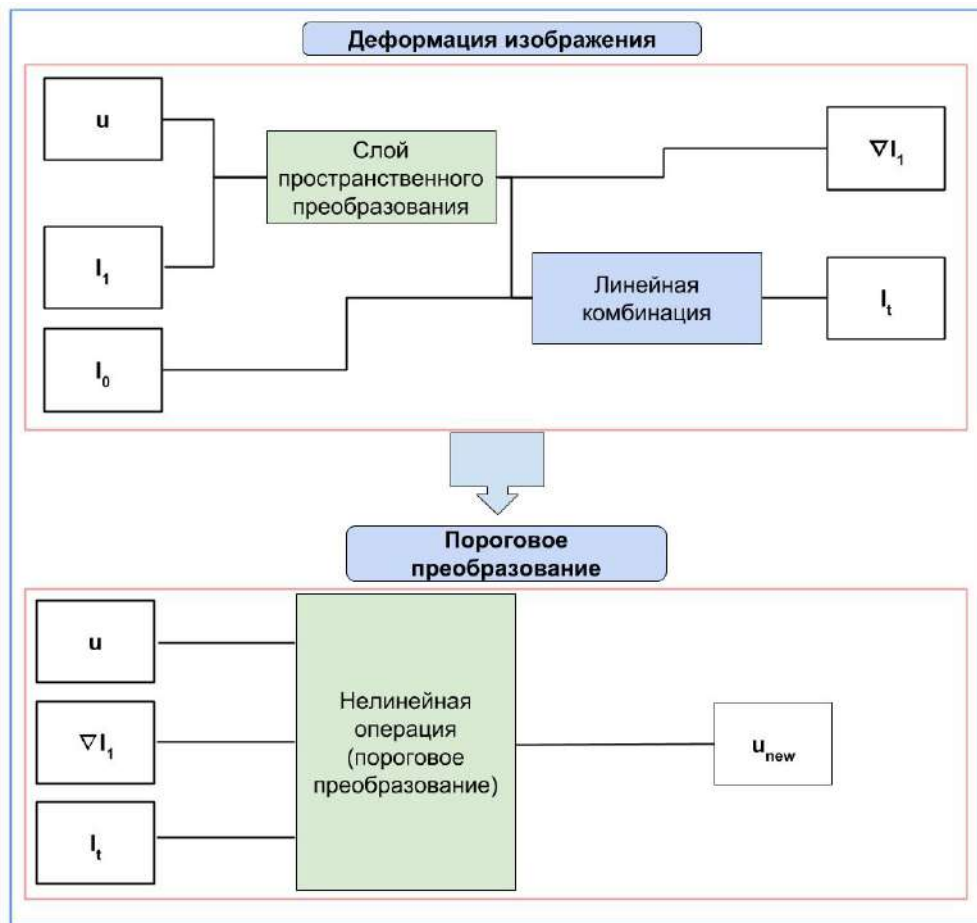


Рис. 4.2: Схематическое представление модуля деформации. Градиенты деформированного изображения по полям смещений вычисляются с использованием слоя пространственной трансформации. Такое преобразование основано на деформации исходного изображения с последующей билинейной интерполяцией и пересэмплированием.

$$\frac{\partial I_1}{\partial y} = \frac{S(I_1, u_{0x}, u_{0y} + \Delta y) - S(I_1, u_{0x}, u_{0y} - \Delta y)}{2\Delta y}. \quad (4.41)$$

Такой слой основывается на параметризации решетки, использованной для сэмплирования деформированного изображения. При этом для сэмплирования используется билинейная интерполяция. Применение данного слоя позволяет вычислять градиент деформированного изображения по компонентам поля смещений. Прочие функции, используемые для построения модуля деформации соответствуют стандартным арифметическим операциям и пороговому преобразованию, заданному соотношением (4.32).

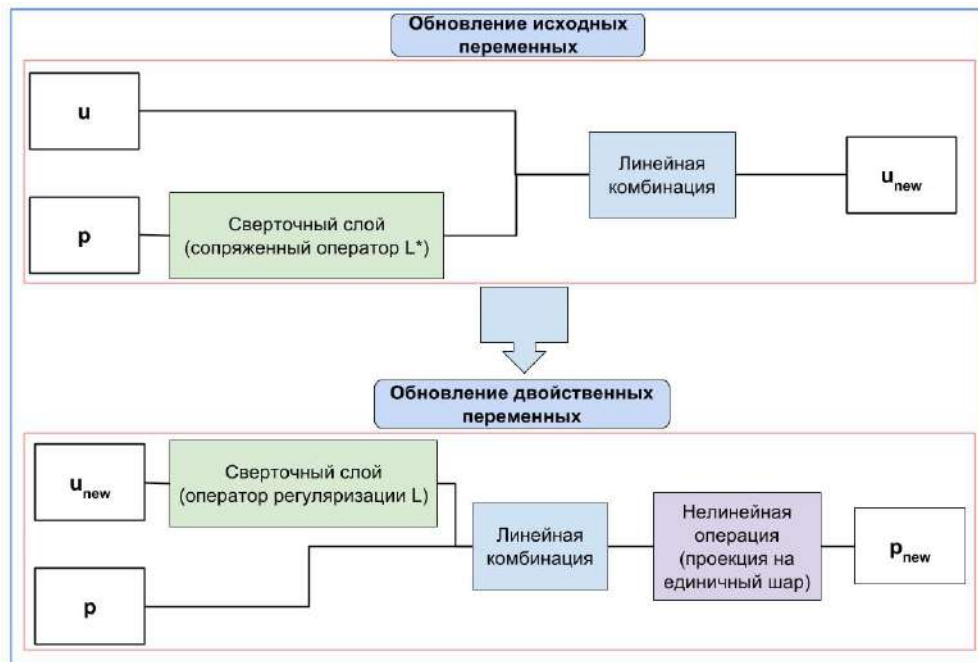


Рис. 4.3: Схема модуля итерации двойственного метода. Модуль принимает на вход начальные значения переменных u, p и вычисляет обновленные значения с использованием обучаемых операторов регуляризации L, L^* с последующим перепроектированием на единичный шар (4.39).

Модуль итерации двойственного метода. В двойственном методе минимизации обновленные значения переменных на каждой итерации вычисляются с использованием операторов регуляризации. Представим эти операторы, $\nabla : R^N \rightarrow R^{N \times 2}$ и $\nabla^* : R^{N \times 2} \rightarrow R^N$ (N - количество пикселей в изображении) в качестве сверток с фильтрами, соответствующими горизонтальной и вертикальной конечной разности:

$$\nabla u_d = \left[\frac{\partial u_d}{\partial x}, \frac{\partial u_d}{\partial y} \right] = [u_d * f_x, u_d * f_y], \quad (4.42)$$

$$\operatorname{div} \mathbf{p} = \frac{\partial p_1}{\partial x} + \frac{\partial p_2}{\partial y} = p_1 * \tilde{f}_x + p_2 * \tilde{f}_y. \quad (4.43)$$

При этом фильтры f_x и f_y соответствуют горизонтальной и вертикальной конечной разности, а фильтры \tilde{f}_x и \tilde{f}_y определяют соответствующий сопряженный оператор ∇^* . При этом в терминах нейросети, двойственная переменная соответствует некоторому двухканальному изображению, которое вычисляется при помощи набора из двух фильтров. Используя аналогичную конструкцию, определим оператор регуляризации $L : R^N \rightarrow R^{N \times K}$ с использованием произвольного количества фильтров K :

$$Lu_d = \begin{bmatrix} u_d * f_1 \\ u_d * f_2 \\ \vdots \\ u_d * f_K \end{bmatrix}. \quad (4.44)$$

При этом сопряженный оператор $L^* : R^{N \times K} \rightarrow R^N$, определенный для двойственной переменной \mathbf{p} , на выходе дает изображение, которое имеет K каналов $\mathbf{p} = [p_1, p_2, \dots, p_K]$. Такой оператор может быть записан в следующем виде:

$$L^* \mathbf{p} = \sum_{i=1}^K p_i * \tilde{f}_i. \quad (4.45)$$

В отличие от оптимизационного метода [2, 62] с фиксированным оператором регуляризации, предложенный метод основан на обучаемой регуляризации, что позволяет уменьшить ошибку метода. Схема модуля итерации двойственного метода приведена на рис. 4.3. Операция перепроектирования на единичный шар (4.39) выступает в качестве нелинейного преобразования, используемого для построения нейросети.

Ввиду применения обучаемой регуляризации, количество итераций алгоритма может быть снижено по сравнению с оптимизационным методом, основанном на фиксированной регуляризации. При этом для каждой итерации, набор фильтров обучается отдельно с целью увеличения способности нейросети адаптироваться ко входным данным.

Предложенный метод имеет низкую вычислительную сложность и может быть использован для расчетов в реальном времени с использованием графического ускорителя (GPU). Количество арифметических операций оценивается как $\mathcal{O}(NDMF^2)$, где N - количество пикселей изображения, D - количество слоев нейросети, M - количество фильтров, а F - размер фильтра.

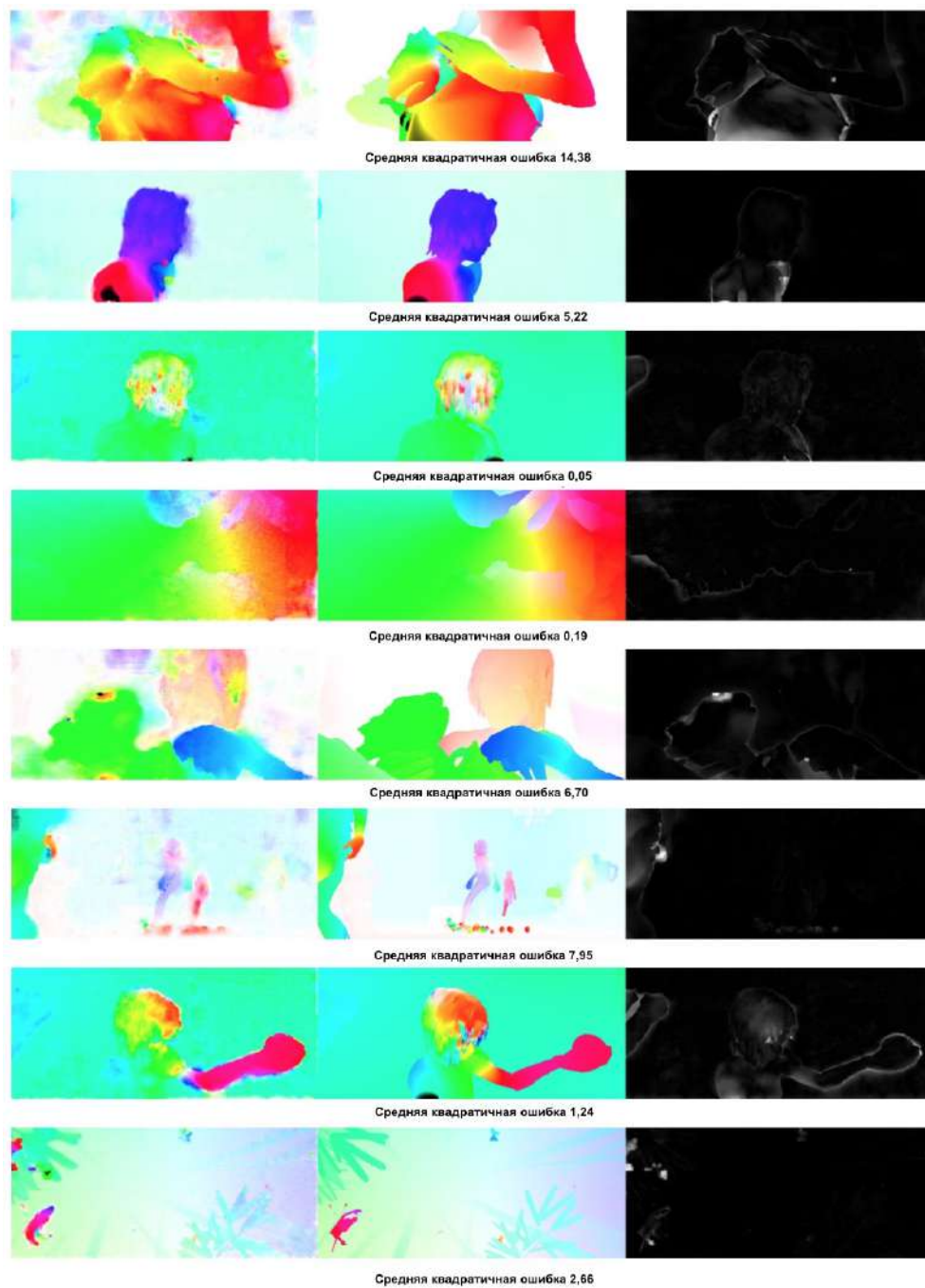


Рис. 4.4: Результаты вычисления оптического потока для публичной коллекции изображений Sintel [72]. Столбцы слева направо: метод, предложенный автором (обучаемая регуляризация), эталонное значение оптического потока, средняя квадратичная ошибка. Средняя квадратичная ошибка метода на тестовой выборке составляет 9,17.

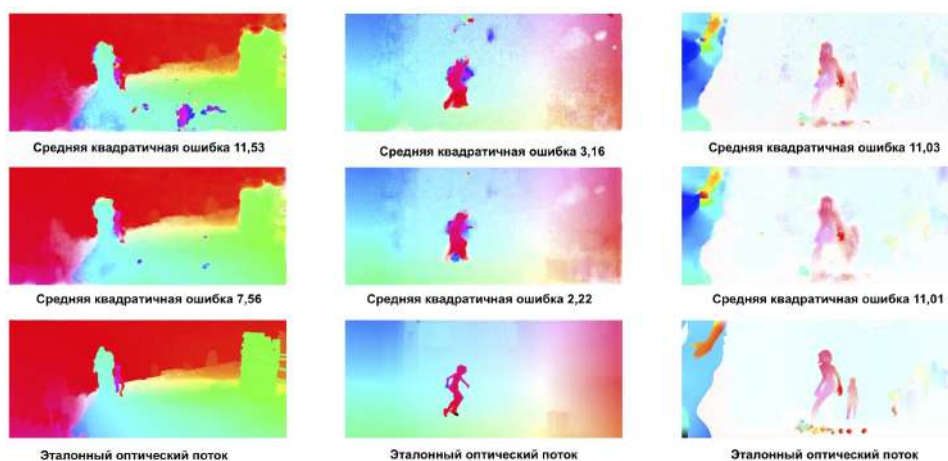


Рис. 4.5: Сравнение фиксированной и обучаемой регуляризации для задачи оптического потока. Первый ряд изображений соответствует фиксированной регуляризации, второй ряд соответствует обучаемой регуляризации, и третий ряд - эталонному значению оптического потока. Обучаемая регуляризация показывает более качественный результат. Полученные поля смещений содержат меньше выбросов и позволяют более четко наблюдать границы движущихся объектов.

4.5.1 Численные эксперименты.

Была проведена серия численных экспериментов по оценке предложенного метода на публичной коллекции изображений Sintel [72]. Для каждой пары изображений определено эталонное поле смещений на разрешении 1024×436 , при этом коллекция содержит сравнительно большие смещения. Вся коллекция была разбита на тренировочное множество (832 пар изображений) и тестовое множество (208 пар изображений).

Была задействована пирамида из 15 масштабов, для перехода между разными масштабами была использована билинейная интерполяция. Использовалась пропорция 0,8 для соседних масштабов и 10 слоев нейросети на каждом масштабе. Каждый слой состоит из модуля деформации и модуля итерации двойственного метода. Банк фильтров для обучаемых операторов регуляризации состоит из 8 фильтров, инициализированных базисными функциями дискретного косинусного преобразования (постоянная компонента удалена из набора). Для операторов обучения операторов L and L^* использовались отдельные значения фильтров. Для тренировки нейросети методом обратного распространения ошибки использовалась Евклидова функция потерь. Применялся адаптивный градиентный метод, описанный в работе [48] с шагом $1,0 \cdot 10^{-5}$. Процесс обучения занимает 10^4 эпох и требует порядка 24 часов на графическом ускорителе.

Финальная евклидова ошибка на тестовом наборе составила 9,17 (рис. 4.4), что показывает улучшение качества по сравнению с оптимизационным методом, использующим фиксированный оператор регуляризации. Ошибка такого метода составляет 10,66. Качественное сравнение результатов для методов с фиксированной и обучаемой регуляризацией представлено на рис. 4.5.

Размер изображения	Время исполнения, с	Частота, кадров/с
1240x375	0,032	31
1024x436	0,031	32
640x480	0,028	36
512x384	0,025	40

Таблица 4.1: Время работы предложенного метода на графическом ускорителе NVIDIA Titan X.

Предложенный метод использует малое количество памяти, что позволяет потенциально использовать его на мобильных платформах. Например, при использовании набора из 8 фильтров, для рассмотренной коллекции изображений нейросеть требует 40 Мб памяти на этапе исполнения при использовании перезаписываемых операций. Среди предложенных в литературе методов, наиболее перспективными для использования в реальном времени являются модели FlowNetS [3] и FlowNet 2.0 [4]. Однако их существенным недостатком является сравнительно большое потребление памяти, а также необходимость претренировки на синтетическом наборе данных, содержащей 22872 пар изображений. Предложенный метод демонстрирует прирост производительности более чем в 2 раза по сравнению с моделью FlowNetS [3] при небольшом увеличении ошибки.

4.6 Многокадровые методы оценки оптического потока

На практике оценка оптического потока как правило производится для видеопоследовательности, содержащей большое количество кадров. Целью многокадровых методов оценки оптического потока является улучшение качества оцениваемых полей смещений посредством использования предположения о пространственно-временной непрерывности движения объектов на сцене.



Рис. 4.6: Иллюстрация метода пространственно-временной фильтрации (изображение из работы [1]). Первый и второй столбец соответствуют двум последовательным кадрам видео. Третий столбец показывает динамику изменения одной из строк изображения по времени (временное измерение соответствует вертикали).

4.6.1 Пространственно-временная фильтрация оптического потока

Классическим подходом для построения многокадрового метода оценки оптического потока является фильтрация в совместных пространственно-временных измерениях [73]. Трехмерный тензор T получается путем конкатенации последовательных кадров видеопоследовательности $I_t(x, y)$ по третьему измерению:

$$T(x, y, t) = I_t(x, y). \quad (4.46)$$

В методе рассматриваются горизонтальные двумерные сечения этого объема (рис. 4.6 (c)), при этом наклон траектории каждого пикселя будет соответствовать горизонтальной компоненте скорости. Так, определение ориентации эквивалентно оценке вектора смещений в единицу времени. В таком подходе, получение устойчивых оценок скорости возможно только при использовании сравнительно больших временных интервалов. Для того чтобы корректно оценивать вектора скоростей на границах объектов, требуется накладывать ограничения на пространственные измерения таких векторов.

4.6.2 Метод, основанный на нахождении траекторий малого ранга

Ввиду того, что визуальной информации не всегда достаточно для сопоставления изображений, большинство методов полагается на регуляризацию. Однако, при анализе видеопоследовательностей возможно также использовать информацию о непрерывности движения. Одним возможным предположением является линейность рассматриваемых траекторий, однако такой подход корректен только для небольших промежутков времени, если на видео представлена нежесткая деформация объектов.

Одним из возможных расширений такого допущения является предположение о том, что траектории, не соответствующие заслоненным объектам, лежат на некотором

многообразии малого ранга. Однако, как показано в работе [74] такое ограничение должно быть добавлено в исходную задачу в некотором мягком виде, что позволяет существенно улучшить качество временной интерполяции. При этом был предложен способ наложения таких ограничений с использованием коэффициентов в некотором базисе траекторий и предположении о разреженности этих коэффициентов. Для последовательности входных изображений, используется следующая функция для представления траекторий $\mathbf{u} = (u_x, u_y)$:

$$\mathbf{u}(x, y, n) = (u_x(x, y, n), u_y(x, y, n)) : \Omega \times \{1, \dots, F\} \rightarrow \mathbf{R}^2, \quad (4.47)$$

где Ω - область определения каждого изображения, n - номер кадра от 0 до F . Каждая из траекторий может быть выражена в виде линейной комбинации R базисных траекторий \mathbf{q}_i :

$$\mathbf{u}(x, y, n) = \sum_{i=1}^R \mathbf{q}_i(n) L_i(x, y) + \boldsymbol{\varepsilon}(x, y, n), \quad (4.48)$$

где $\boldsymbol{\varepsilon}(x, y, n)$ - ошибка представления: наличие такого слагаемого позволяет накладывать ограничение на ранг траекторий в нежестком виде. Базисные траектории $\mathbf{q}_i(n)$ определены как вектор-функции:

$$\mathbf{q}_1(n), \dots, \mathbf{q}_R(n) : \{1, \dots, F\} \rightarrow \mathbf{R}^2, \quad (4.49)$$

при этом такие функции являются общими для всех пикселей изображения. В работе рассматриваются ортонормированные базисы в пространстве траекторий. При этом линейная оболочка набора базисных траекторий называется *подпространством траекторий*. Коэффициенты линейной комбинации $L_i(x, y)$ являются функциями координат.

Для нахождения полей смещений, метод использует постановку оптимизационной задачи, содержащую три члена: член, отвечающий за входные данные, член отвечающий за мягкое ограничение на ранг траекторий и функционал регуляризации. Описанный метод позволяет оценивать поля смещений по многим кадрам видео и эффективно использовать результат для временной интерполяции [74].

4.7 Выводы

Методы оценки оптического потока, демонстрирующие результаты с наименьшей ошибкой как правило можно поделить на две категории: методы глубокого машинного обучения и оптимизационные методы. При этом наиболее качественные результаты достигаются при сочетании сопоставления глубоких дескрипторов с методами регуляризации, основанными на дискретной оптимизации. Такие методы демонстрируют меньшую ошибку при большей вычислительной сложности по сравнению с общими сверточными архитектурами, такими как модель FlowNet [3].

Для построения метода машинного обучения, имеющего низкую вычислительную сложность, автором была предложена сверточная архитектура, полученная путем представления итераций оптимизационного метода в виде слоев нейросети. При этом важным аспектом представляется используемый подход к регуляризации. Автором был предложен метод для обучения оператора регуляризации [12], описанный в данной главе. Такой подход позволяет улучшить качество оценки метода оценки оптического потока в реальном времени с незначительным увеличением вычислительной сложности.

Глава 5

Сопоставление медицинских ультразвуковых изображений

Данная глава посвящена задаче сопоставления ультразвуковых медицинских изображений. Раздел 5.1 содержит постановку задачи эластографии, разделы 5.2 и 5.3 содержат обзоры методов эластографии по двум и многим кадрам. Главу завершает описание метода, предложенного автором (раздел 5.4), приведены результаты численных экспериментов.

Раздел 5.5 содержит постановку задачи сопоставления ультразвуковых сигналов в отражательной томографии, описание метода, предложенного автором, и результаты численных экспериментов на синтетических и реальных данных.

5.1 Задача ультразвуковой эластографии

Ультразвуковая эластография - это метод оценки упругости тканей человека, основанный на сопоставлении ультразвуковых изображений с разной степенью механического сжатия. В данной работе рассматривается задача так называемой квазистатической эластографии, целью которой является реконструкция механической деформации ткани из серии ультразвуковых снимков. Одним из наиболее практических сложных этапов в построении методов эластографии является оценка механической деформации из поля смещений.

Ультразвуковая эластография - пример задачи сопоставления изображений, наиболее сходная с задачей вычисления оптического потока, тем не менее, имеющая одно важное отличие. Целью решения задачи эластографии также является оценка двумерного поля смещений по паре или серии изображений. Однако, для медицинской

диагностики используется не само поле смещений, а его пространственная производная, которая соответствует упругости ткани. Оценка пространственной производной поля смещений требует разработки иных подходов к регуляризации, чем задача вычисления оптического потока.

Исследования по ультразвуковой эластографии соответствуют области анализа медицинских изображений, в связи с этим, нередко имеют терминологию, которая отличается от решения задачи сопоставления изображений, рассматриваемой в рамках компьютерного зрения. Тем не менее, задача является одним из вариантов постановки общей задачи, рассмотренной в данной работе.

5.2 Обзор методов эластографии, основанных на двух кадрах

Большое количество методов эластографии, предложенных в литературе оценивают механическую деформацию как производную от поля смещений, полученного сопоставлением пар ультразвуковых снимков разной степени механического сжатия. Некоторые методы [5, 75, 76] оценивают поля смещений методом кросс-корреляции фрагментов изображения. Применение подобных методов без регуляризации приводит к шумным полям смещений. При этом результат может быть улучшен путем использования дополнительных эвристик, таких как, например, адаптивный размер блока [77].

Серия опубликованных методов основана на применении дискретной оптимизации, при этом целевая функция состоит из члена, отвечающего за соответствие поля смещений входным данным и функционала регуляризации первого порядка. Например, метод предложенный в работе [78] использует метод динамического программирования для минимизации функционала с регуляризацией первого порядка. При этом склонность метода к нулевому решению, характерная для метода ввиду штрафа, накладываемого на пространственное изменение поля смещений, исключается за счет постобработки результата в виде решения дополнительной задачи непрерывной минимизации. Однако, такой метод позволяет вычислять боковые смещения в небольшом ограниченном диапазоне, ввиду использования жадной стратегии поиска. Другой метод, предложенный в работе [79], использует многомасштабный подход для минимизации схожего функционала. При этом метод склонен к сходимости к локальному минимуму, поэтому качество результат сильно зависит от начальной инициализации. Для сильных степеней сжатия, такой метод как правило сходится либо к локальному минимуму, соответствующему излишне гладкому решению, либо расходится.

Для того, чтобы применять эластографию для медицинской диагностики, производится реконструкция локального поля механического сжатия на основе поля смещений. Такая реконструкция основана на вычислении градиента шумных полей смещений. Классическим подход к устойчивому дифференцированию полей смещений является алгоритм [80], основанный на методе наименьших квадратов. Более подробный обзор методов реконструкции локального поля деформации проводится в работе [81], он включает в себя методы численного дифференцирования, основанные на построении сплайнов и методы шумоподавления с использованием вейвлет-преобразования [82, 83]. Численное дифференцирование также может быть записано в виде обратной задачи, где в качестве линейного оператора выступает интегрирование [84].

Одним из наиболее важных аспектов методов эластографии является нахождение достаточно гладкого решения, естественным образом описывающим входные данные. Входные ультразвуковые снимки часто содержат высокий уровень шума, так что локальные алгоритмы сопоставления дают решения с большим количеством выбросов. При этом качества результата зависит как от сложности изучаемой ткани, так и от применяемой степени сжатия. На практике, методы локального сопоставления не дают устойчивого решения при использовании степени сжатия 2-3%. Однако, использование больших степеней сжатия в линейном диапазоне (до 5%) ведет к проявлению дополнительных деталей на изображениях локальной деформации, что увеличивает диагностические возможности метода.

Большинство предложенных методов использует регуляризацию для нахождения полей смещений для высоких степеней сжатия [7, 79]. Часто, при этом, использование сильной регуляризации, необходимой для получения визуально гладкого решения, ведет к пересглаженным изображениям, которые не соответствуют реальным механическим деформациям и возможно не отражают некоторых деталей человеческой ткани.

Одним из перспективных методов для нахождения устойчивой оценки поля смещений для изображений большой степени сжатия, является применение двухэтапного поиска смещений. При этом входное изображение деформируется с применением первоначальной оценки поля смещений и затем применяется для повторного поиска смещений, что позволяет уменьшить эффективную степень сжатия [85].

5.3 Обзор многокадровых методов эластографии

На практике, сопоставление ультразвуковых изображений, встречает ряд трудностей, таких как, например, высокий уровень шума, участки с отсутствующей текстурой. При этом, поле смещений, построенное по двум кадрам, может содержать значительное число выбросов. Одним из перспективных способов по преодолению данных недостатков, является комбинирование информации с серии кадров. Такой подход был рассмотрен в настоящей работе.

5.3.1 Метод ElastMI

Одним из методов по улучшению качества эластографии засчет использования нескольких кадров предложен в работе [7]. При этом используется три ультразвуковых снимка. Метод основан на введение ограничений специального вида, которые основаны на рассмотрении механики материала с жестким включением. Вводится целевая функция, которая объединяет функционал входных данных и регуляризацию. Такая функция минимизируется с использованием EM-алгоритма и метода итеративно взвешиваемых наименьших квадратов с целью уменьшения количества выбросов.

5.3.2 Видеопоследовательность ультразвуковых снимков и видео-эластограмма

Обобщение традиционного метода эластографии на видеопоследовательность представлено в работе [86]. В то время как большинство методов эластографии оперирует парами или тройками снимков, данный метод осуществляет реконструкцию пространственно-временной зависимости полей механических деформации на основе видеопоследовательностей ультразвуковых изображений. Метод использует специальную адаптивную меру близости, основанную на вычислении соотношения сигнала и шума.

При анализе видеопоследовательностей, наиболее практически трудным представляется временное выравнивание рассматриваемых траекторий. При этом используется предположении о непрерывности изучаемого движения. Базовым методом при переходе каждому следующему кадру является накопление поля смещений.

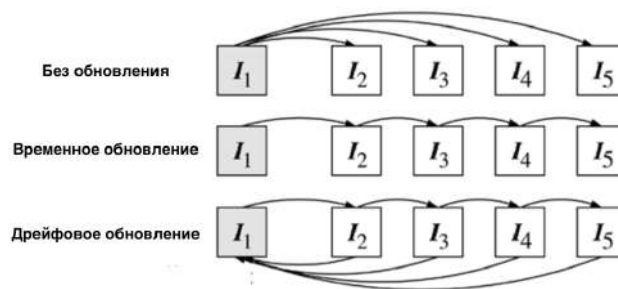


Рис. 5.1: Схематическое изображение различных подходов к пространственно-временному сопоставлению изображений (рисунок из работы [86]). Представлены три различных способа обновления траекторий по времени: метод без обновления, метод обновления по времени и метод дрейфового обновления.

Применение подобного метода корректно только в течение ограниченного промежутка времени. За пределами такого промежутка требуется алгоритм временного обновления, например:

$$I^{t+1}(x, y) = I^t(x + d_x^t(x, y), y + d_y^t(x, y)). \quad (5.1)$$

Несмотря на то, что небольшие смещения между последовательными кадрами видео могут быть оценены достаточно точно, даже небольшая ошибка с течением времени имеет свойство накапливаться. Такой эффект называется временным дрейфом. Для того, чтобы избежать накопления ошибки, в методе применяется специальная схема дрейфового обновления. Метод основан на временном выравнивании, схематически представленном на рис. 5.1.

5.4 Метод регуляризованной многокадровой эластографии

5.4.1 Описание метода

Этот подраздел описывает метод решения задачи ультразвуковой эластографии, предложенный автором. Метод имеет ряд преимуществ по сравнению с другими методами, предложенными в литературе. Основное преимущество состоит в возможности получать более качественные оценки механической деформации за счет использования трех кадров с известным соотношением силы нажатия (большинство предложенных в литературе методов использует два кадра). При этом сила нажатия определяется аппаратно, используя специальный вариант ультразвуковой пробы с

встроенным датчиком силы, описанным в работе [87]. Такая проба была спроектирована с целью увеличения степени контроля за процессом получения изображений, что представляется важным на практике в связи с заметным влиянием неточностей оператора на качество получаемых снимков. Так, использование контролируемой пробы в экспериментах, описанных в данной работе, позволяет применить к датчику предписанное значение силы нажатия, а также гарантировать плавное изменение величины силы при получении нескольких кадров. Однако, разработанный метод также может быть применен с использованием пробы, не имеющей датчика силы, при этом соотношение между величиной силы для разных кадров может быть определено алгоритмически из входных данных. Такой подход увеличивает вычислительную сложность метода и вносит в качество снимков зависимость от точности оператора.

Предложенный автором подход предлагает метод реконструкции механического сжатия, который основан на использовании триплетов ультразвуковых снимков. Реконструкция механической деформации основана на минимизации функционала, где в качестве регуляризации используется адаптивный вариант полной вариации (total variation). Используемый для минимизации двойственный метод градиентного спуска был параллелизован и эффективно реализован автором с использованием графического ускорителя.

Предложенный метод был провалирован в серии экспериментов с использованием данных численного моделирования. В этом случае был смоделирован процесс получения ультразвукового изображения на деформированных объектах известной геометрии. Процесс деформации был рассчитан с использованием метода конечных элементов. Предложенный метод также был применен к фантомам, изготовленным из смеси желатина и агара - такие фантомы имеют известную геометрию и позволяют количественно оценить способность метода к выделению мягких или жестких включений сферической или цилиндрической формы в деформируемых органических материалах. Проведенные эксперименты показали, что метод позволяет оценить поле смещений для вдвое большей силы сжатия по сравнению с методами, основанными на использовании двух кадров, что ведет к увеличению пространственного разрешения получаемых оценок механической деформации. Такой вывод далее подтверждается серией экспериментов на реальных клинических данных, для которых метод позволяет получить более четкие изображения жестких включений ввиду использования большей силы нажатия. Результаты также подтверждают целесообразность использования контролируемой ультразвуковой пробы для эластографии.

5.4.1.1 Оценка поля смещения

Во многих случаях при исследовании тканей человека, в диапазоне от 0 до 5% сжатия, связь между величиной силы нажатия и механической деформацией линейна [7]. Однако, имеется несколько причин, ввиду которых линейное соотношение нарушается, например ввиду наличия источников механического давления внутри тканей, таких как, например, кровеносные сосуды и стенки артерий. Другой ряд причин включает в себя ненулевой вращательный момент, который получает проба от оператора и случайное изменение плоскости, вдоль которой получится изображение. В таком случае целесообразно уточнять первоначальную оценку деформации, основанную на описанном линейном соотношении, для того чтобы учесть нелинейные эффекты - такой подход лежит в основе метода, предложенного автором.

Введем обозначения, используемые при описании метода. Обозначим входные три кадра ультразвукового снимка как I_0, I_1, I_2 . Ультразвуковой снимок состоит из m одномерных сигналов по n временных отсчетов каждый, что соответствует изображению размера $m \times n$. Каждая пара кадров может быть сопоставлена при помощи двумерного поля смещения $\mathbf{d}(x, y) = (d_a(x, y), d_l(x, y))$, определенного для каждого пиксела $\mathbf{p} = (x, y)$ эталонного (первого) кадра, так что пиксел соотносится с пикселом $\mathbf{p} - \mathbf{d}(\mathbf{p})$ на втором кадре. Поля d_a и d_l есть продольные и поперечные компоненты смещений соответственно. Задачей ультразвуковой эластографии является оценка двумерного поля механической деформации $\mathbf{s}(x, y) = (s_a(x, y), s_l(x, y))$, которое может быть получено из поля смещения путем дифференцирования, описанного ниже.

Для каждого кадра, вектор смещения полагается целочисленным и конечным внутри преписанного декартово произведения двух интервалов Λ (окна поиска) и может быть записано следующим образом:

$$\Lambda = \{0, \dots, +D_a\} \times \{-D_l, \dots, +D_l\}. \quad (5.2)$$

Где D_a и D_l - максимальные абсолютные значения продольного и поперечного смещения соответственно.

Первым этапом предложенного метода является вычисление поля смещения для небольшой контактной силы, соответствующей приблизительной степени сжатия 1%. При этом используется стандартный локальный алгоритм поиска соответствующих фрагментов по прямоугольному окну методом полного перебора. Обозначим результат такой процедуры следующим образом:

$$\mathbf{d}_{coarse} = \mathbf{B}(I_0, I_1), \quad (5.3)$$

Где \mathbf{B} - алгоритм поиска смещений описанный выше, который принимает на вход пару кадров и вычисляет двумерное поле смещений. Обозначим за \mathbf{d}_{coarse} двумерный вектор смещений, вычисленный для пары изображений (I_0, I_1) .

Использование описанной выше ультразвуковой пробы позволяет получать значение силы нажатия в режиме реального времени, что позволяет без дополнительной задержки экстраполировать полученное значение поля смещений на большую степень сжатия, соответствующую третьему кадру I_2 (при этом на практике характерная степень сжатия составляет 2-3%). Таким образом получается первоначальная оценка поля смещения на основе линейного соотношения сила-деформация.

Определим операцию деформации изображения. Пусть W - функция деформации, которая принимает на вход изображение I и поле смещений $\mathbf{d} = (d_x, d_y)$:

$$I_d = W(I, \mathbf{d}), \quad (5.4)$$

Тогда интенсивность деформированного выражения выражается следующим образом:

$$I_d(x, y) = I(x - d_x, y - d_y). \quad (5.5)$$

Для того, чтобы уточнить первоначальное приближение $\frac{f_2}{f_1} \mathbf{d}_{coarse}$, полученное для поля смещений, запускается второй проход аналогичной процедуры поиска двумерных смещений, при этом на вход процедуре подается пара изображений I_0 и деформированная с использованием первоначальной оценки смещения версия кадра I_2 , которая может быть обозначена как $W(I_2, \frac{f_2}{f_1} \mathbf{d}_{coarse})$. При этом интервал поиска смещения на данном этапе может быть существенно сужен:

$$\mathbf{d}_{local} = \mathbf{B} \left(I_0, W(I_2, \frac{f_2}{f_1} \mathbf{d}_{coarse}) \right). \quad (5.6)$$

Используя оценки \mathbf{d}_{coarse} и \mathbf{d}_{local} , поле смещения для кадров I_0 и I_2 может быть получено суммированием:

$$\mathbf{d}_{total} = \frac{f_2}{f_1} \mathbf{d}_{coarse} + \mathbf{d}_{local}. \quad (5.7)$$

Полное смещение подается на вход процедуры реконструкции механической деформации, описанной в следующем подразделе.

В качестве метрики для фрагментов изображения используется сумма квадратов разностей интенсивностей (SSD) по прямоугольному окну. Такой метод аналогичен

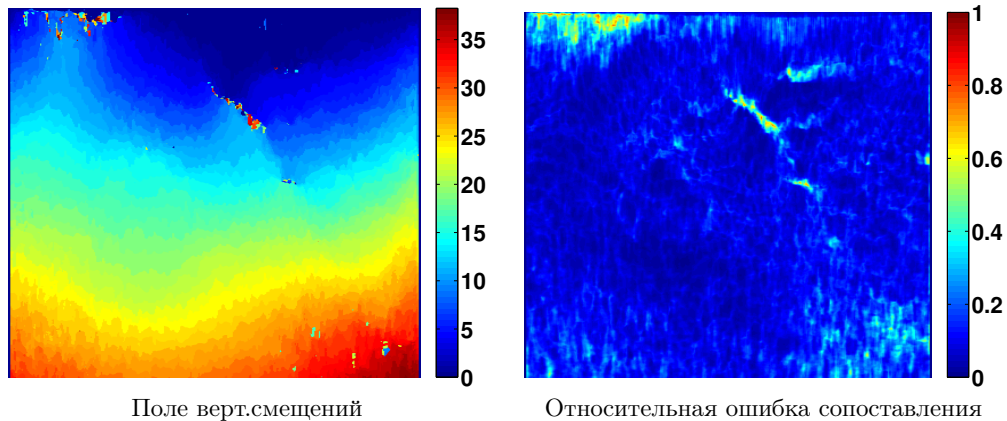


Рис. 5.2: Пример поля смещений для ткани печени человека (данные для эксперимента предоставлены авторами работы [7]).

локальному методу стерео-сопоставления, заданному соотношением 3.2 и описанному в разделе 3.2.1. Помимо поля смещений, процедура выдает на выходе меру качества полученного поля смещения $q(x, y)$, оцененную как относительную ошибку сопоставления в той же метрике.

Описанный метод использует линейное соотношение между контактной силой и деформации только в качестве начального приближения и позволяет вычислять отклонения от линейной модели, которые встречаются на практике.

5.4.1.2 Оценка механической деформации

5.4.1.3 Функционал.

Главной подзадачей ультразвуковой эластографии является вычисление механической деформации на основе полученного поля смещения. При этом оценки поля смещения на практике часто содержат выбросы и большое количество шума, что не позволяет напрямую применять техники численного дифференцирования ввиду их недостаточной устойчивости. При этом могут быть использованы идеи, предложенные в современных методах шумоподавления и методах решения обратных задач, таких как [52, 54, 88]. Исследование возможности подобных методов для задачи ультразвуковой эластографии является одним из предметов данного исследования.

В соответствии со сходной работой [81] оценка механической деформации формулируется в качестве обратной задачи, при этом используется функционал следующего вида:

$$E(s_a) = \|As_a - d_a\|_2 + \lambda\rho(s_a), \quad (5.8)$$

Где s_a - регуляризованная оценка продольной деформации, $\rho(s_a)$ - функционал, соответствующий регуляризации и λ - постоянный коэффициент, контролирующий степень сглаживания. Действие оператора A описывается следующим образом:

$$As_a + d_a(0) = \int_0^L s_a(x)dx + d_a(0) = d(L). \quad (5.9)$$

Для фиксированного поля смещений, задача оценки механической деформации является выпуклой оптимизационной задачей. Качество решения, полученное нахождением глобального минимума, зависит от выбора регуляризационного функционала.

В соответствии с проведенными авторами численными экспериментами, оценка поля смещений часто содержит участки со значительной ошибкой, часто встречаемые на границе различных тканей. В таком случае, большая ошибка в оценке поля смещений даже на небольшом пространственном участке, приводит к существенным выбросам и разрывам в оценке механической деформации. При этом качество сопоставления может быть оценено количественно, используя относительную ошибку интенсивности в L^2 норме (рис. 5.2). Такая оценка позволяет использовать адаптивную регуляризацию: в регионах сопоставления с низкой ошибкой полагаться на входные изображения, а в регионах с высокой ошибкой - полагаться на регуляризацию. На вышеказанном рис. 5.2 приведена относительная ошибка сопоставления и соответствующее поле смещений. Для того, чтобы включить информацию о качестве сопоставления в регуляризационный функционал, вводится матрица весов $D = \text{diag}(w)$, которая зависит от описанной выше меры качества сопоставления:

$$E(s_a) = \|D(As_a - d_a)\|_2 + \lambda\rho(s_a). \quad (5.10)$$

Вычисление произведения As_a в уравнении (5.10) является трудоемким ввиду того, что матрица A содержит около половины ненулевых элементов. Т.к. оценка механической деформации является локальной, вычислительная сложность может быть снижена, используя интегрирование по меньшей окрестности. Численные эксперименты, проведенные автором, показали, что использование окрестности в 100 временных отсчетов приводит к визуально неразличимым результатам по сравнению с использованием полного изображения. Однако, ввиду наличия эффективных параллельных реализаций операции префиксной суммы (`cumsum`), в работе для простоты используется интегрирование по полному интервалу от 0 до L .

В алгоритме используется бинарная маска весов $w(x, y)$ для каждого пиксела, основанная на пороговом отсечении относительной ошибки сопоставления $q(x, y)$ для

того, чтобы разделить пиксели на две категории, соответствующих высокой и низкой ошибке сопоставления:

$$w(x, y) = \begin{cases} 1 & q(x, y) < t \\ 0 & q(x, y) \geq t \end{cases} \quad (5.11)$$

Так, пиксели со значениями весов $w(x, y)$, близкими к нулю, соответствуют неверному сопоставлению - решение в таких регионах полностью полагается на регуляризацию.

5.4.1.4 Регуляризация.

Регуляризация с использованием L^2 нормы. Для того, чтобы форсировать пространственную гладкость оценки поля деформаций, накладывается штраф на пространственные изменения этого поля:

$$E(s_a) = \|D(As_a - d_a)\|_2 + \int (\alpha[D_x s_a]^2 + [D_y s_a]^2), \quad (5.12)$$

где D_x и D_y - операторы дифференцирования методом конечных разностей в горизонтальном и вертикальном направлении соответственно. Т.к. вертикальное разрешение ультразвукового снимка существенно ниже горизонтального, вводится дополнительный коэффициент α , для того чтобы компенсировать эту разницу. Решение находится минимизацией функционала методом наименьших квадратов, при этом наиболее вычислительно трудоемкий этап соответствует решению разреженной линейной системы уравнений.

Регуляризация с использованием полной вариации. В соответствии с проведенными численными экспериментами, использование L^2 нормы в качестве регуляризации ведет к пересглаженным решениям с потерей визуальной информации на границах тканей. Для избежания подобного эффекта, представляется целесообразным использование полной вариации (total variation [51]) в качестве регуляризационного члена. При этом исходный функционал может быть записан в виде:

$$E(s_a) = \|D(As_a - d_a)\|_2 + \lambda TV_\alpha(s_a), \quad (5.13)$$

где второй член выражается как:

$$TV_\alpha(s_a) = \int \sqrt{\alpha[D_x s_a]^2 + [D_y s_a]^2}. \quad (5.14)$$

Использование регуляризации на основе полной вариации ведет к сохранению границ на перепадах между мягкими и жесткими тканями. При этом возможно использование эффективных реализаций двойственного градиентного метода, описанного ниже, для эффективной минимизации. Численные эксперименты, проведенные автором не показали улучшения при использовании регуляризации второго порядка, предложенной в работах [52, 54].

Используя двойственное определение нормы, описанный регуляризационный член может быть записан в виде:

$$TV_\alpha(u) = \sup\left\{\int u \operatorname{div}_\alpha \mathbf{v}, \|\mathbf{v}\|_\infty \leq 1.\right\} \quad (5.15)$$

При этом используются операторы градиента и дивергенции:

$$\operatorname{div}_\alpha \mathbf{v} = \alpha \frac{\partial v_1}{\partial x} + \frac{\partial v_2}{\partial y}, \quad (5.16)$$

$$\nabla_\alpha u = \left(\alpha \frac{\partial u}{\partial x}, \frac{\partial u}{\partial y} \right). \quad (5.17)$$

Для сопряженного оператора выполняется следующее соотношение:

$$(\operatorname{div}_\alpha)^* = -\nabla_\alpha. \quad (5.18)$$

5.4.1.5 Минимизация.

В соответствии с работой [89], автором был использован двойственный градиентный метод минимизации. Диагональная матрица D опущена для простоты, при этом минимизация может быть в виде выпукло-вогнутой задачи:

$$\min_{s_a \in S} \max_{\mathbf{p} \in P, \mathbf{r} \in R} \langle \nabla s_a, \mathbf{p} \rangle + \langle A s_a - d_a, \mathbf{r} \rangle - \frac{\lambda}{2} \|\mathbf{r}\|_2^2. \quad (5.19)$$

Где \mathbf{p} и \mathbf{r} это двойственные переменные по отношению к s_a и значению невязки $(A s_a - d_a)$ соответственно. P и R - единичные шары, соответствующие тем же переменным:

$$P = \{\mathbf{p} \in C^{2mn}, \|\mathbf{p}\|_\infty \leq 1\}. \quad (5.20)$$

Также может быть определено следующее проксимальное отображение:

$$\text{prox}_2^\sigma(\tilde{\mathbf{r}}) = \arg \min_{\mathbf{r} \in R} \frac{\|\mathbf{r} - \tilde{\mathbf{r}}\|_2^2}{2\sigma} + \frac{\lambda}{2} \|\mathbf{r}\|_2^2 = \frac{\tilde{\mathbf{r}}}{1 + \sigma\lambda}. \quad (5.21)$$

Минимизация может быть проведена следующим образом. Положим $s_a = 0$, $\mathbf{p} = 0$, $\mathbf{r} = 0$, выберем $\tau, \sigma > 0$. В соответствии с методом [53] на каждой итерации значения переменных обновляется согласно соотношениям:

$$\mathbf{p} = \text{proj}_P(\mathbf{p} + \sigma \nabla s_a), \quad (5.22)$$

$$\mathbf{r} = \text{prox}_2^\sigma(\mathbf{r} + \sigma(As_a - d_a)), \quad (5.23)$$

$$s = s + \tau(\text{div} \mathbf{p} - A^* \mathbf{r}). \quad (5.24)$$

Процедура выполняется до сходимости s_a . При этом на практике представляется важным выбор шага минимизации, для этого автором была использована стратегия, описанная [90], где значения шагов τ и σ на итерации n описываются следующим образом:

$$\sigma_n = 0,2 + 0,08n, \quad (5.25)$$

$$\tau_n = \left(0,5 - \frac{5,0}{15,0 + n}\right). \quad (5.26)$$

Дискретизация операторов градиента и дивергенции [52] производится с использованием конечных разностей:

$$\text{div} \mathbf{v} = \delta_{x+} v_1 + \delta_{y+} v_2, \quad (5.27)$$

$$\nabla u = (\delta_{x-} u_1, \delta_{y-} u_2), \quad (5.28)$$

где δ_{x+} , δ_{y+} и δ_{x-} , δ_{y-} соответствуют использованию конечных разностей горизонтального и вертикального направлений.

5.4.1.6 Реализация с использованием графических ускорителей.

Двойственный алгоритм градиентного спуска может быть параллелизован на уровне различных пикселей изображения. Такая параллелизация может быть естественным образом реализована на графическом ускорителе (GPU).

При этом важным практическим аспектом является использование локальной памяти. Автором была разработана эффективная реализация метода с использованием локальной памяти: при этом значения прямых и двойственных переменных в

Значение силы (прямое измерение)	Значение силы (оценка из смещений)	Относительная ошибка
1,61	1,51	6%
1,68	1,68	0%
1,75	1,86	6%
1,75	1,84	5%
1,80	1,90	6%
1,86	1,76	5%
2,36	2,21	6%

ТАБЛИЦА 5.1: Ошибка алгоритмической оценки силы нажатия. Таблица содержит результаты оценки силы, произведенной двумя способами: первый столбец содержит значения приложенной силы, полученные путем прямого измерения и второй столбец содержит значения, оцененные алгоритмически с использованием пар соответствующих изображений. Проведенные результаты показывают, что средняя ошибка алгоритмической оценки составляет 5%.

окрестности текущего пиксела загружаются в локальную память, вычисляется одна итерация метода, и результат снова записывается в глобальную память.

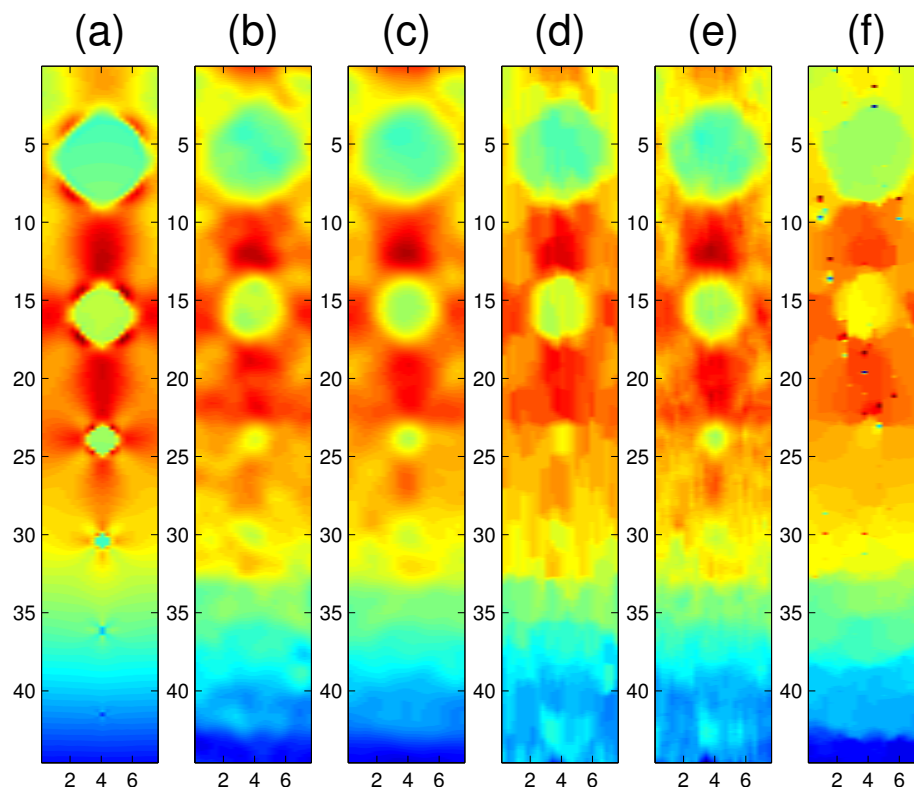
Наиболее трудным на практике представляется реализация префиксной суммы. Автором была использована библиотека `NVIDIA Thrust` [91] для вычисления префиксных сумм, а также для параллелизации базовых векторных операций, используемых в итеративном алгоритме.

Вычисление значения целевого функционала, используемого для проверки на сходимость производится каждые 100 итераций метода с целью экономии вычислительных ресурсов.

5.4.2 Численные эксперименты

Далее, приводятся количественные и качественные результаты применения предложенных методов на следующих экспериментах: моделирование сжатия синтетических фантомов с известной геометрией, снимки специально изготовленных фантомов с приближенно известной геометрией и снимки реальных тканей человека.

Набор сравниваемых методов. Автором были произведены численные эксперименты по сравнению следующих пяти методов, из которых четыре являются разновидностями предложенного подхода. Четыре метода соответствуют использованию двух видов регуляризации (L^2 норма и полная вариация) и разному количеству используемых кадров (пары и тройки кадров). При использовании пары изображений, результат сопоставления используется для оценки механического изображения в соответствии с рядом работ, предложенных в литературе [78, 92, 93].



Оценки поля продольной механической деформации для синтетического фантома.

Рис. 5.3: Оценка вертикальной компоненты поля механических деформаций для фантомов, содержащих 6 включений: (а) эталонные значения, (б) L^2 регуляризация (2 кадра), (в) L^2 регуляризации (3 кадра), (г) полная вариация (2 кадра), (д) полная вариация (3 кадра), (е) метод, основанный на динамическом программировании. Наилучшие результаты получены с использованием 3 кадров. Предложенный метод превосходит метод на основе динамического программирования при использовании любого из двух предложенных подходов для регуляризации

Предложенные методы были сравнены с методом оценки поля смещений на основе динамического программирования, при этом производная поля смещений была вычислена с использованием регуляризации (использовалась полная вариация). Такой метод аналогичен ряду работ [6, 94], что позволяет сравнить их с предложенным подходом. При этом параметры метода были подобраны для получения оптимальных результатов.

Продолжение данного раздела содержит описание используемых данных и результаты численных экспериментов.

Синтетический эксперимент на основе моделирования сжатия фантома известной геометрии.

Для моделирования сжатия синтетических фантомов был использован пакет программ Abaqus [95]. При этом использовалась степень сжатия до 5% в продольном направлении. Для моделирования был использован метод конечных элементов.

Для этого эксперимента были использованы следующие значения диаметра сферических включений: 6, 4, 2, 1, 0,5 и 0,25 мм, расположенных линейно в продольном направлении. Включения имеют модуль упругости 30кПа, в то время как остальной объем имеет модуль упругости 15 кПа. Для моделирования формирования ультразвуковых изображений использовался программный пакет Field-II [96, 97], при этом было задано разрешение 3400×128 . Центральная частота ультразвукового пучка, использованная при моделировании соответствует 5 МГц.

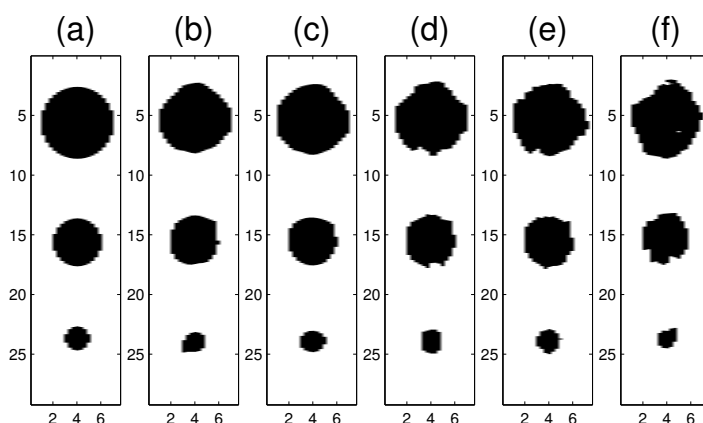
В практических приложениях, наибольший интерес представляет использование эластографии для сегментации опухолей. В связи с этим для предложенных методов была количественно оценена способность выделения жестких включений на основе метода пороговой сегментации. При этом в качестве количественной меры был использован индекс Джаккарта (соотношение пересечения и объединения для оцененной и эталонной сегментации):

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}. \quad (5.29)$$

Т.к. только три наиболее крупных включения были выделены рассмотренными методами, в таблице 5.2 приведены значения индекса Джаккарта, соответствующие трем включениям. Результаты сегментации представлены на рис. 5.4. После проведения экспериментов было экспериментально установлено, что для методов на основе 3 кадров, использование степеней сжатия в 1% и 3% является оптимальным для второго и третьего кадров соответственно.

Приведенные результаты (рис. 5.3) показывают, что метод на основе трех кадров превосходит по качеству метод на основе двух кадров и метод на основе динамического пограммирования. Главной причиной количественного улучшения является возможность использовать большие степени сжатия при использовании трех кадров.

Набор синтетических фантомов случайной геометрии. Ввиду того, что человеческая ткань имеет многослойную структуру, одним из важных аспектов методов эластографии является способность выделять опухоли, т.е. твердые и мягкие включения, расположенные между слоями различной жесткости. Для того, чтобы



Выделение жестких включений с использованием пороговой сегментации.

Рис. 5.4: Результаты сегментации для фантомов, содержащих 6 жестких включений: (а) эталонный результат, (б) L^2 регуляризация (2 кадра), (в) L^2 регуляризация (3 кадра), (г) полная вариация (2 кадра), (д) полная вариация (3 кадра), (е) метод на основе динамического программирования. Результаты показывают превосходство метода на основе трех кадров вне зависимости от выбранного метода регуляризации.

Метод	6 мм	4 мм	2 мм
Динамическое программирование	81,1%	66,2%	57,0%
Полная вариация (2 кадра)	77,7%	77,8%	84,0%
Полная вариация (3 кадра)	80,4%	81,4%	86,0%
L^2 регуляризация (2 кадра)	78,5%	75,5%	69,9%
L^2 регуляризация (3 кадра)	80,9%	78,5%	87,2%

Таблица 5.2: Индекс Джаккарта для трех жестких включений. Метод на основе трех кадров демонстрирует наилучшие результаты.

измерить относительную производительность рассмотренных методов, был смоделирован набор из 7 синтетических фантомов с 4 круглыми включениями каждый (три из них показаны на рис. 5.5).

При этом каждый слой имеет случайную толщину, суммарная высота слоев составляет 44,45 мм. Координаты включений также были выбраны случайным образом внутри объема фантомов. Каждый фантом состоит из включений с модулем упругости 5 кПа, 10 кПа, 50 кПа и 120 кПа. Включения имеют диаметр 2 мм и модуль упругости 30 кПа. Для моделирования также был использован метод конечных элементов.

Сначала была вычислена оценка поля деформаций, используя пару изображений: изображение без сжатия и изображение с 1% сжатия. Затем, были вычислены оценки поля деформаций на основе тройки изображений: изображение без сжатия, изображение с 0,5% сжатия и изображение с 3,0% сжатия.

Метод	Средний инд. Джаккарта	Средняя L2 ошибка
Полная вариация (2 кадра)	31,0%	$7,0 \cdot 10^{-2}$
Полная вариация (3 кадра)	40,5%	$5,9 \cdot 10^{-2}$
Динамическое прогр.	25,6%	$27,6 \cdot 10^{-2}$
L^2 регуляризация (2 кадра)	34,6%	$6,4 \cdot 10^{-2}$
L^2 регуляризация (3 кадра)	41,0%	$5,5 \cdot 10^{-2}$

Таблица 5.3: Количественное сравнение различных методов для 7 синтетических фантомов со случайной геометрией. Фантомы содержат сферические включения со случайными координатами, расположенными между слоями материала различной жесткости.

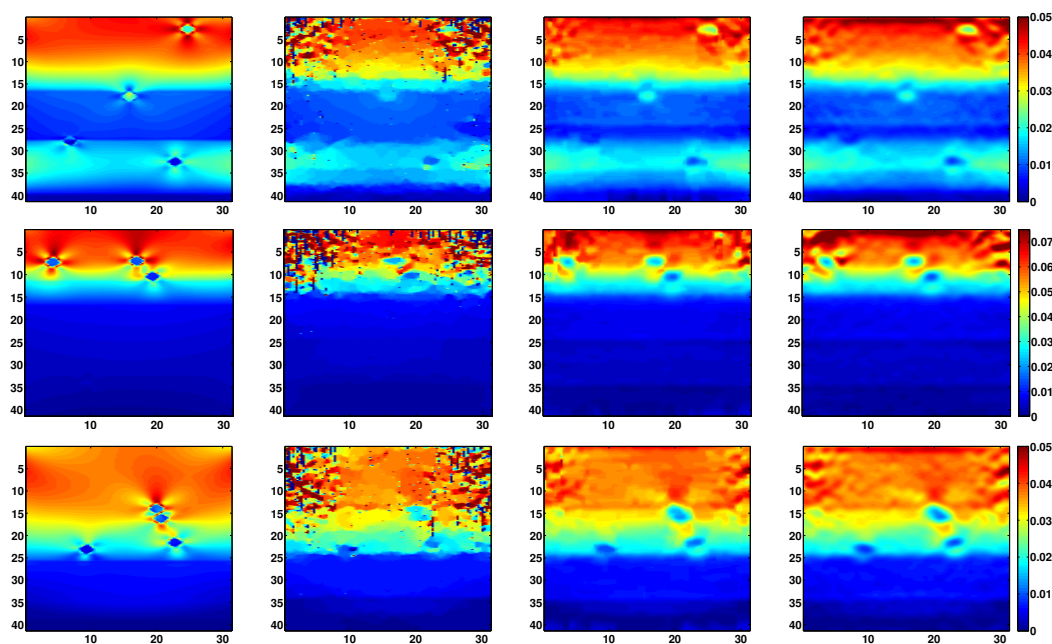


Рис. 5.5: Оценки механической деформации для трех синтетических фантомов. Столбцы слева направо: (а) эталонные значения, (б) метод на основе динамического программирования, (в) полная вариация (3 кадра), (г) L^2 регуляризация (3 кадра)

В таблице 5.3 приведены результаты для набора из 7 случайных фантомов (всего 28 включений): средний индекс Джаккарта и средняя L^2 ошибка для квадратного региона размером 2,5 мм, окружающего каждое из сферических включений (таб. 5.3). Также показана кумулятивная гистограмма ошибки, т.е. доля включений, имеющих ошибку меньше, чем некоторое предписанное значение для обеих мер (рис. 5.7).

Проведенные автором численные эксперименты показали преимущество предложенного метода, основанного на использовании трех кадров по сравнению с методом, использующим 2 кадра. Оба критерия, выбранных для количественной оценки ошибки, показывают превосходство предложенного метода над методом на основе динамического программирования.

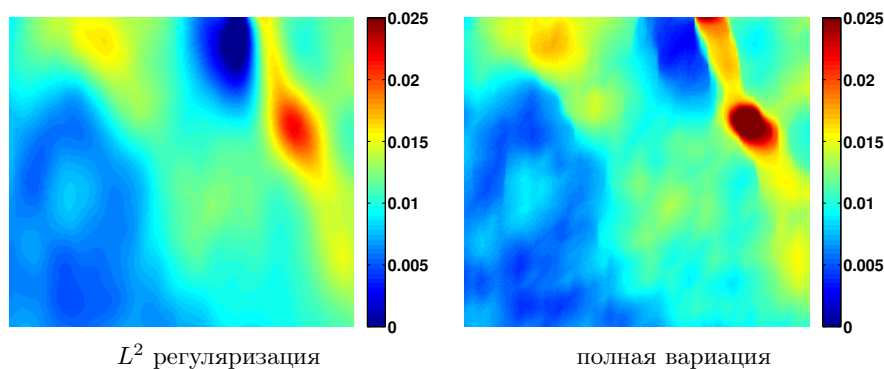


Рис. 5.6: Сравнение L^2 регуляризации и полной вариации (были использованы реальные данные, на рисунке приведена часть изображения 5.12). Использование полной вариации позволяет получить более четкие изображения поля деформации на границах тканей различной жесткости.

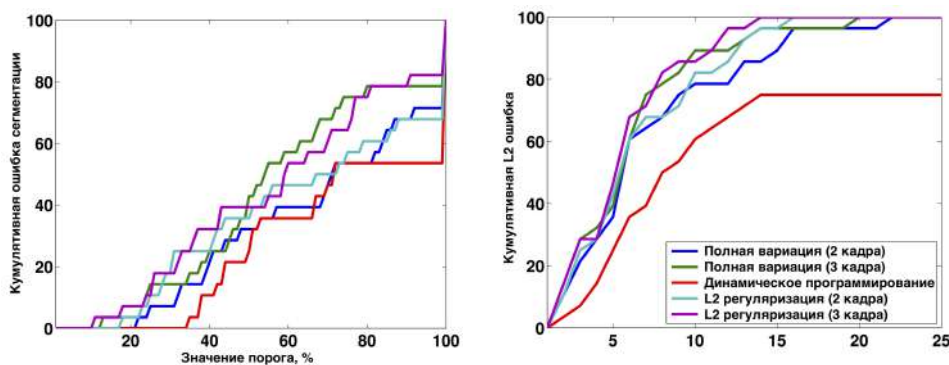


Рис. 5.7: Кумулятивный график ошибки для индекса Джаккарта. Каждый из графиков показывает долю всех включений, которые имеют L^2 ошибку и индекс Джаккарта меньше, чем значение, соответствующее оси абсцисс. Все четыре предложенных метода значительно превосходят метод на основе динамического программирования, тогда как метод на основе двух кадров превосходит метод на основе трех кадров. Использование двух описанных подходов для регуляризации ведет к сравнимым результатам.

Согласно проведенным численным экспериментам не было обнаружено улучшения при использовании регуляризации на основе полной вариации по сравнению с L^2 регуляризацией. Однако качественные наблюдения показывают, что использование полной вариации порождает решение с более четкими границами, что ведет к более легкому выделению жестких включений при визуальном наблюдении.

Фантомы, изготовленные из желатина и агара. Описанная ниже серия экспериментов была произведена с использованием специально изготовленных фантомов из желатина и агара. Для получения ультразвуковых снимков была использована ультразвуковая проба Terason 3000t с частотой 5 МГц. Для измерения контактной силы было использовано устройство, описанное в работе [98].

Проведенные автором эксперименты (рис. 5.8) подтверждают способность предложенного метода и выделять мягкие включения уменьшающегося размера вплоть до 4 мм с высокой степенью достоверности. При этом достаточно использовать метод на основе пары кадров на основе полной вариации. Использование трех кадров позволяет увеличить степень сжатия с 1% до 2% (использование двух кадров в таком случае влечет большое количество выбросов рис. 5.8, нижний ряд). В результате использования более высокой степени сжатия, предложенный метод позволяет выделять включения небольшого размера вплоть до 2 мм, что не представляется возможным для метода на основе двух кадров.

На рис. 5.9 приведена оценка механической деформации для набора из трех фантомов с жесткими цилиндрическими включениями и соответствующие индексы Джаккарта. В качестве эталонной формы были рассмотрены круги соответствующего радиуса, оптимальное расположение которых была выбрано для каждого из изображений механической деформации в пределах окна поиска размера 0,8 мм.

Эксперимент для тканей человека. Для проведения эксперимента, описанного в данном подразделе были использованы ультразвуковые снимки печени человека, содержащей злокачественную опухоль. Снимки получены при помощи ультразвуковой пробы Antares Siemens System. Для использования методов на основе трех кадров были использованы уровни сжатия 1,5% и 3,0%. Эксперимент основан на косвенном измерении степени сжатия и показывает возможность применения предложенного метода для изображений со стандартной ультразвуковой пробы, не содержащей датчика измерения контактной силы.

Результаты, приведенные на рис. 5.12, показывают, что предложенный метод оценки механической деформации на основе трех снимков показывает значительно лучший контраст для жестких и мягких включений, что является следствием возможности использования больших степеней сжатия. Несколько мягких и твердых включений проявляются на изображении благодаря использованию метода, основанного на использовании трех кадров. В дополнение, использование полной вариации позволяет сохранить резкие границы. При этом проявляется кусочно-постоянный характер решений, который не имеет значительного влияния на возможности использования изображений в медицинской диагностике.

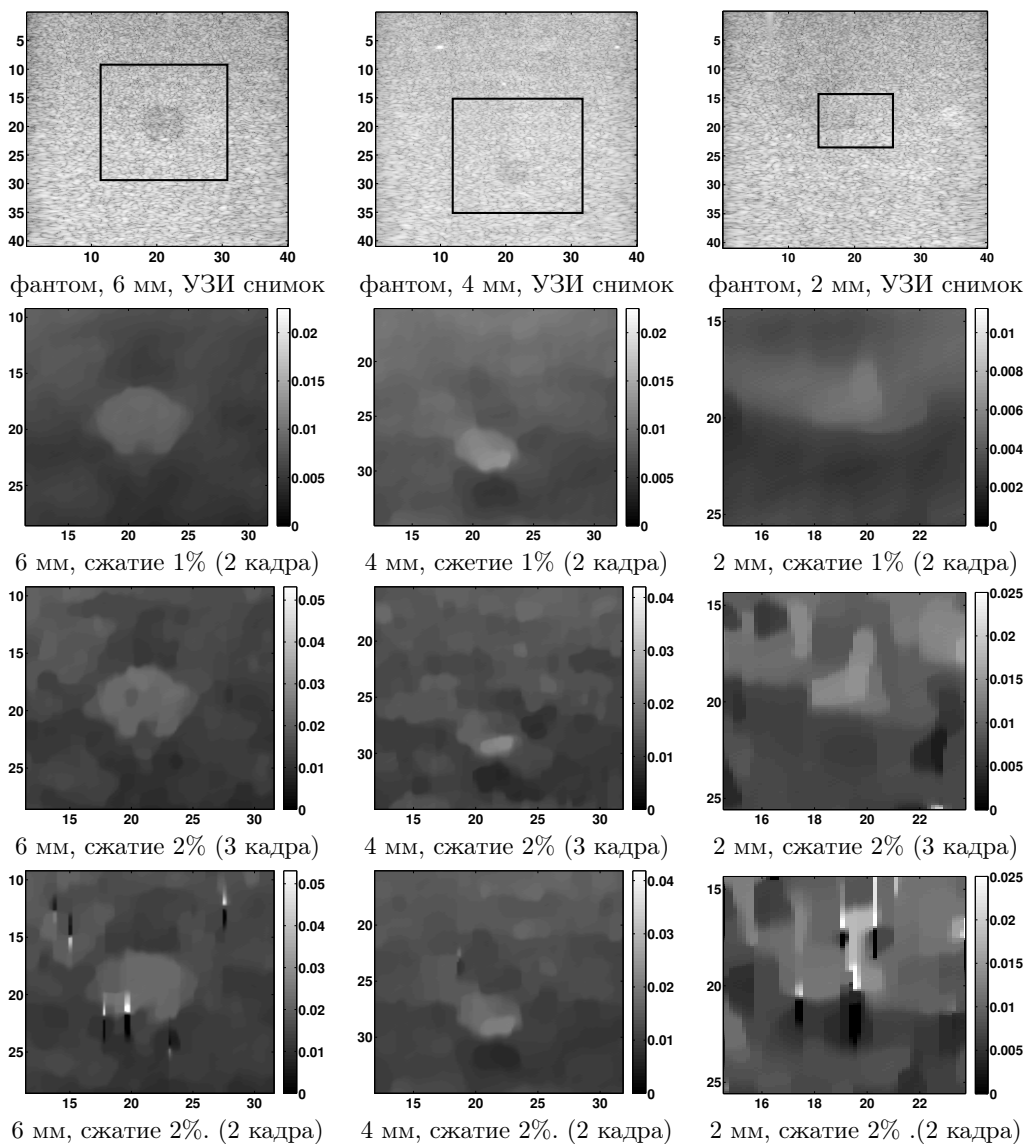


Рис. 5.8: Оценка механической деформации для фантомов из желатина и агара с включениями размера 6,4,2 мм соответственно (слева направо). Первая строка: ультразвуковое изображение, вторая строка: метод с использованием двух кадров с использованием полной вариации, третья строка: метод на основе трех кадров с использованием полной вариации, четвертая строка: оценка механической деформации для высокой степени сжатия с использованием двух кадров.

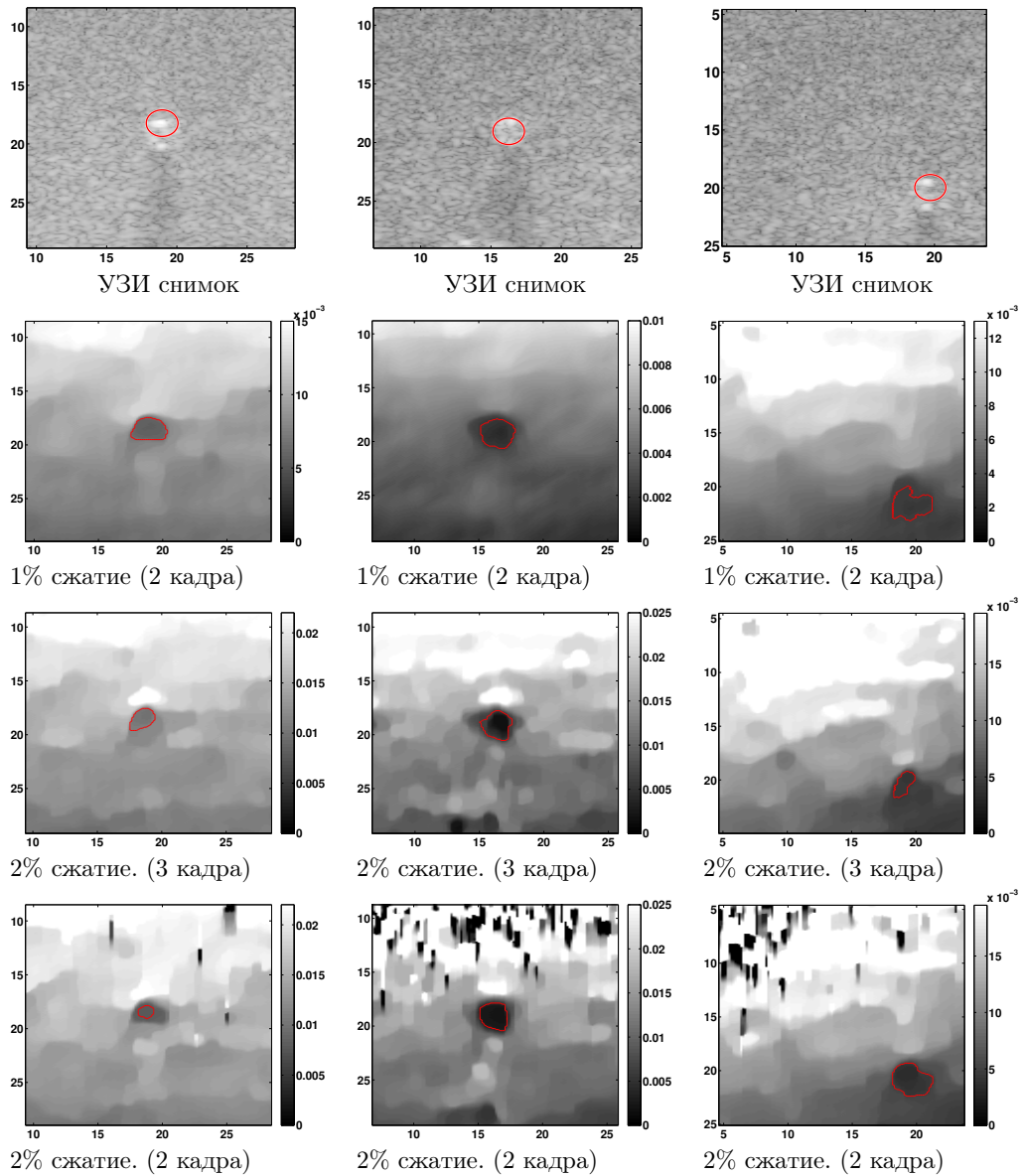


Рис. 5.9: Поля механической деформации для фантомов из желатина и агара с жесткими включениями диаметра (слева-направо). Первая строка: УЗИ снимок, вторая строка: результат метода для двух кадров, полная вариация, третья строка: метод основанный на трех кадрах и использованием полной вариации, четвертая строка: метод на двух кадрах, примененный к первому и третьему кадрам из тройки, соответствующих более сильной степени сжатия. Приводится индекс Джаккарта для каждого случая. Эталонная форма обозначена красными окружностями на УЗИ снимках наряду с оптимальными линиями уровня для сегментации и вычисления индекса Джаккарта.

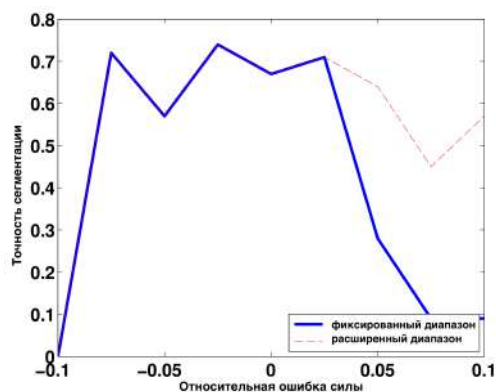


Рис. 5.10: Эксперимент по оценке ошибке косвенного измерения контактной силы, основанной на сопоставлении изображений. График показывает относительную ошибку косвенного измерения по отношению к значениям, полученным с датчика измерения силы. Синяя линия соответствуют фиксированному окну поиска для второго этапа оценки поля смещений (рис. 5.9, третья строка, первый столбец), красная линия соответствует расширенному окну поиска для второго этапа сопоставления.

Вычислительная сложность и время исполнения метода.

Предложенный автором метод состоит из двух этапов: сопоставление блоков (используется два раза) и реконструкция деформации. Количество арифметических операций при сопоставлении блоков оценивается как $\mathcal{O}(ND_aD_l)$, где N - количество пикселей изображения, D_a и D_l - максимальные абсолютные значения продольных и поперечных смещений соответственно. Количество операций при реконструкции деформации составляет $\mathcal{O}(N)$.

На практике для реальных данных были использованы блоки 100x4 пикселей для сопоставления с последующим применением медианного фильтра размера 9x7. Для первого этапа сопоставления изображений используется степень сжатия 1%, в то время как для второго этапа поиск производится внутри меньшего окна. Описание эффективной реализации сопоставления изображения методом сравнения блоков в применении к задаче эластографии описан в работе [99]. Последовательная реализации на языке MATLAB, предложенная автором выполняется в течение 3,5 сек на реальных данных.

Реализация алгоритма реконструкции механической деформации с использованием графических ускорителей, предложенная автором была протестирована на графическом ускорителе GTX 690 с использованием библиотеки CUDA 5.5, при этом время исполнения составляет 0,54 сек для реальных данных размера 450x256 (поле смещений размера 1800x256 после четырехкратного уменьшения разрешения в вертикальном направлении). Соответствующее время исполнения на мобильной версии графического ускорителя NVIDIA GeForce GT 650M с использованием библиотеки

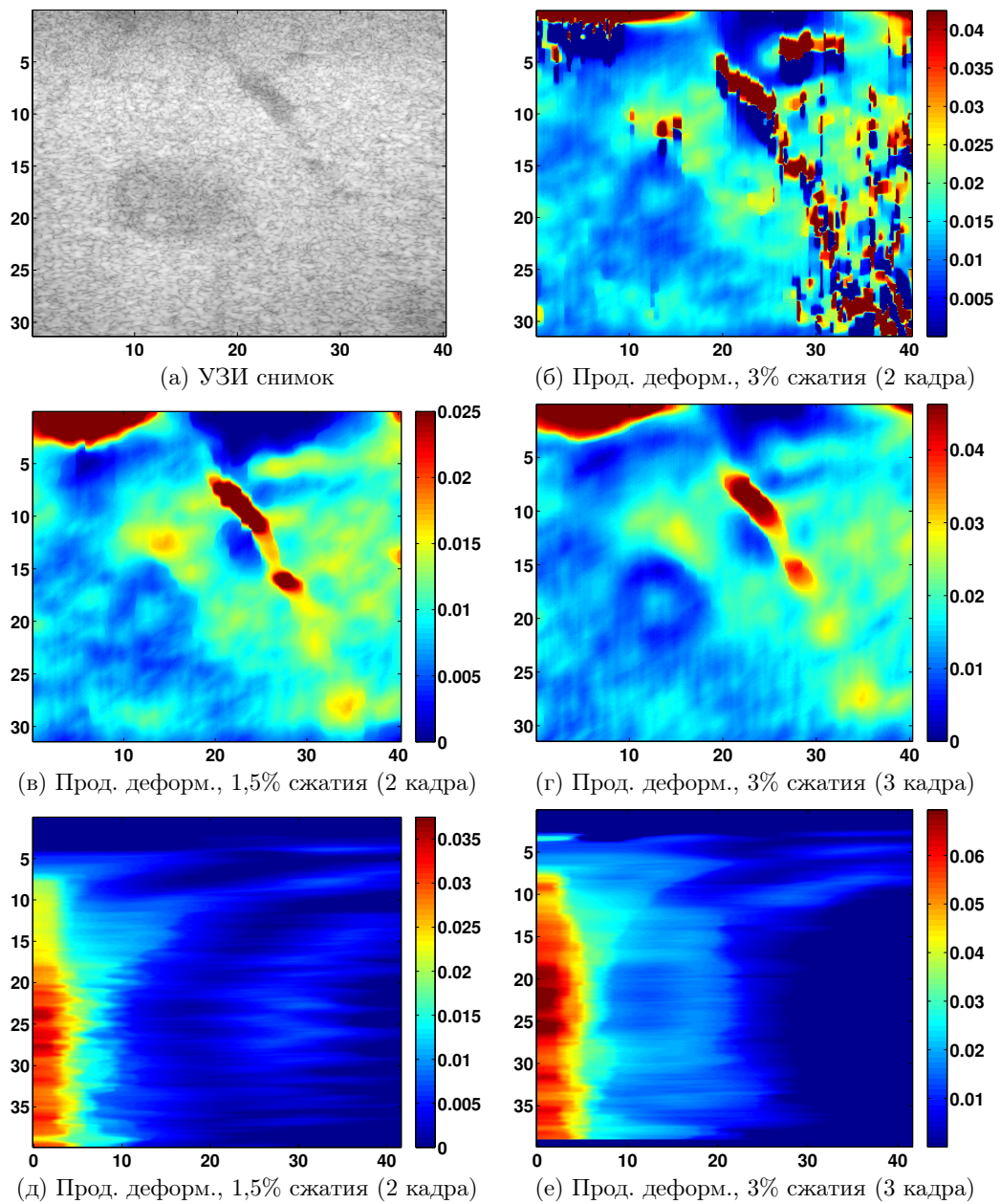


Рис. 5.11: Поле механической деформации для ткани печени человека, содержащей злокачественную опухоль: (а) УЗИ снимок, (б) продольная деформация, 2 кадра, 3% сжатия, (в) продольная деформация, 2 кадра, 1,5% деформация (г) продольная деформация, 3 кадра, 3% деформация, (д) поперечная деформация, 2 кадра, 1,5% сжатия, (е) поперечная деформация 3 кадра, 3% сжатия. Предложенный метод демонстрирует улучшенную способность выделять мягкие и жесткие включения за счет использования более высоких степеней сжатия. Злокачественная опухоль (область высокой деформации слева от центра изображения) становится более легко выделяемой при использовании метода основанном на трех кадрах.

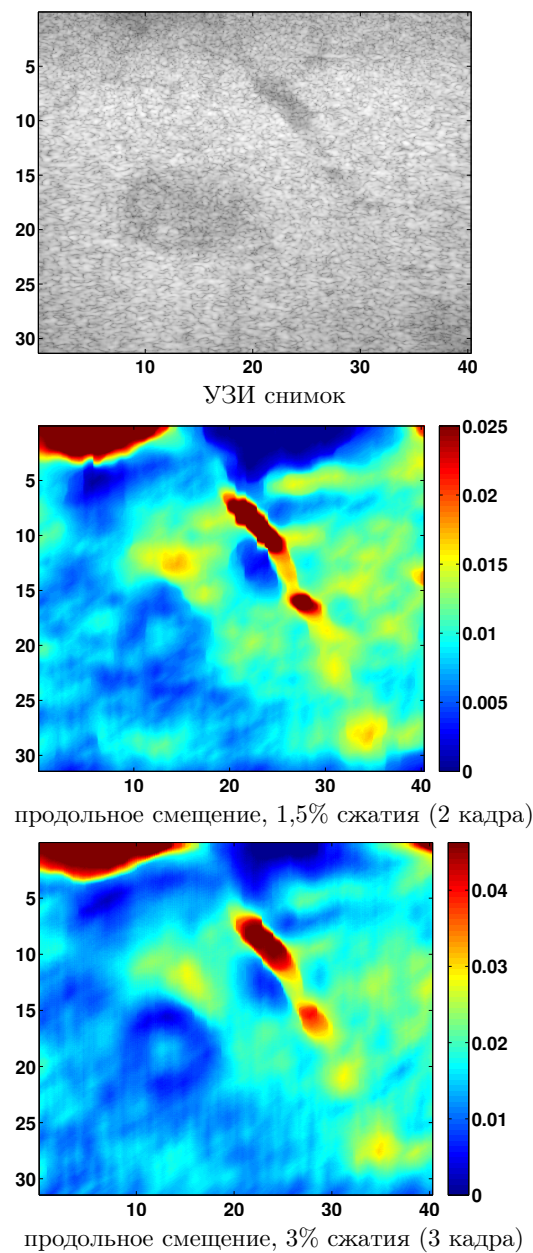


Рис. 5.12: Поля деформации для ткани печени человека, содержащей злокачественную опухоль: (а) УЗИ снимок, (б) продольная деформация, 2 кадра, 3% сжатия, (с) продольное смещение, 2 кадра, 1,5% сжатия, (д) продольная деформация, 3 кадра, 3% сжатия, (е) поперечная деформация, 2 кадра, 1,5% сжатия, (ф) поперечное смещение, 3 кадра, 3% сжатия. Предложенный автором метод демонстрирует улучшенную способность выделять мягкие и жесткие включения за счет использование больших степеней сжатия для метода, основанного на применении трех кадров. Злокачественная опухоль (участок высокой деформации слева от центра изображения) становится более четко видимым при использовании метода на основе трех кадров.

CUDA 6.0 составило 0,98 сек. Приведенные времена исполнения показывают возможность клинического применения метода, т.к. время получения данных при использовании контактной силы составляет порядка нескольких секунд.

Косвенное измерения степени сжатия на основе сопоставления изображений. Метод, предложенный автором, применим для любых ультразвуковых проб до тех пор, пока проба не получает значительный вращательный момент. Была проведена серия экспериментов, в которых значение силы, полученное со специальной ультразвуковой пробы было сопоставлено со значением силы, измеренном алгоритмически. Для этого были использованы специально изготовленные фантомы из желатина и агара. Для определения силы используется соотношение для полей смещения. Степень сжатия, соответствующая кадру с наименьшей силой принимается за единичную, затем вычисляется соотношение смещений для каждого пиксела. Затем относительное значение силы оценивается как медиана соотношения смещений. Таблица 5.1 содержит сравнение значения силы, измеренное напрямую и косвенно для 8 серий снимков.

В соответствии с проведенными экспериментами, относительная ошибка косвенного измерения силы составляет 5%, в то время как ошибка прямого измерения силы ограничена 2%. Для того чтобы далее оценить устойчивость алгоритма, основанного на трех кадрах по отношению к ошибке в силе, был проведен следующий эксперимент. Был построен график зависимости ошибки сегментации по отношению к относительной ошибке в силе. Фиксированное окно поиска было использовано для первого эксперимента (см. рис. 5.10, синяя кривая) и расширенное окно поиска - для второго эксперимента (см. рис. 5.10, красная кривая). Эксперимент показал, что метод может быть применен для данных без значения силы посредством косвенного измерения, при этом, при этом нет необходимости расширять диапазон поиска. Такая необходимость возникает при превышении ошибки измерения силы значения 6%. Однако, косвенное измерения силы увеличивает вычислительную трудоемкость метода за счет использования дополнительного этапа сопоставления изображений.

5.5 Сопоставление сигналов в ультразвуковой томографии

Другой важной задачей, имеющей сходную постановку с задачей сопоставления изображений, является реконструкция в отражательной ультразвуковой томографии. В данном случае, целью является сопоставление массива ультразвуковых сигналов, получаемых с разных углов зрения по отношению к изучаемому объему

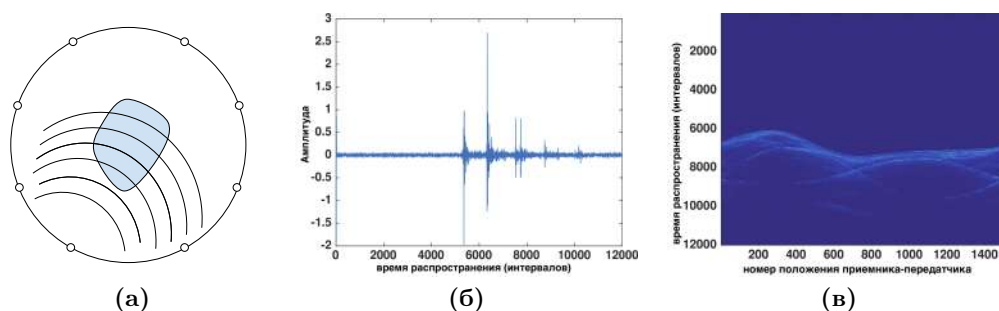


Рис. 5.13: Установка и пример данных, получаемых в отражательной ультразвуковой томографии. (а) Геометрия установки. Пьезоэлементы, используемые для генерации импульсов и приема сигналов расположены по кругу на равном расстоянии. (б) Пример одномерного сигнала, получаемого с одного из элементов. (в) Входные данные, используемые для реконструкции: двумерное изображение получено из массива 1500 одномерных сигналов. Каждый из одномерных сигналов получен с последовательных позиций приемников-передатчиков.

(рис. 5.13 (а)). Реконструкция производится на основе преобразования координат из локальной системы каждого из приемников в глобальную систему координат, привязанную к центру установки. Таким образом, каждый из временных отсчетов отраженного сигнала получает некоторое поле смещений в процессе перепроектирования в глобальную систему координат. Метод реконструкции для данной задачи подробно рассмотрен в последующих подразделах. При этом искажение вдоль направления распространения ультразвукового сигнала обратно пропорционально скорости звука.

Одним из сходств данной задачи с задачей оценки поля смещений в бинокулярном стерео, является сопоставление одномерных сигналов: в случае стерео-реконструкции таковыми являются строки левого и правого изображений, в случае отражательной томографии - амплитуды отраженных сигналов, записанные в разных положениях приемников-передатчиков.

5.5.1 Постановка задачи

Разработанный автором метод реконструкции [100] предназначен для ультразвукового томографа, описанного в работе [101]. В установке использован единственный пьезоэлемент, который перемещается по окружности (рис. 5.13 (а)) и служит приемником-отражателем: в каждом из положений элемент испускает импульс и записывает отраженный сигнал. Каждый из временных отсчетов записанного одномерного сигнала (рис. 5.13 (а)) соответствует временному интервалу, который импульс прошел расстояние от приемника до некоторого отражателя и обратно. После объединения каждого из одномерных сигналов, получается двумерное изображение (рис. 5.13 (в)), которое является входным для метода реконструкции.

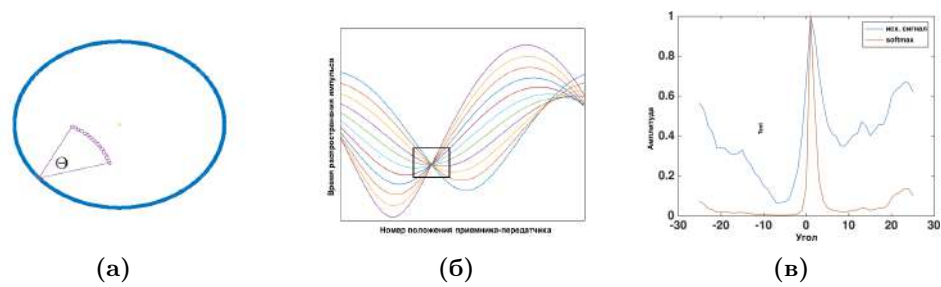


Рис. 5.14: Схема предложенного алгоритма реконструкции. (а) Дуга окружности $A(r, j)$. (б) Набор кривых-отражений, соответствующих каждой из точек дуги. (в) Пример углового распределения до и после применения функции *softmax*.

Процедура калибровки устройства, которая необходима для установления соответствия между временными интервалами при распространения волн и пространственными координатами, описана в работе [101]. После проведения такой процедуры, каждый пиксел входного изображения $I(r, j)$ соответствует расстоянию r от приемника-передатчика до некоторой точки изучаемого объема внутри установки, где j - номер положения приемника-передатчика на окружности.

Такая система координат называется *пространством данных*. Конечной целью является проектирование входных сигналов на *глобальную систему координат*, центр которой привязан к центру томографа. Заметим, что между этими двумя системами координат не существует однозначного соответствия: так, каждый пиксел в пространстве данных, заданный парой (r, j) , соответствует дуге $A(r, j)$ в глобальной системе координат (x, y) . Такая дуга является объединением всех точек на одном расстоянии r от приемника-передатчика (рис. 5.15 (а)). Каждая точка (x, y) в глобальной системе координат, в свою очередь, соответствует некоторой синусоиде в пространстве данных (рис. 5.15 (б)).

5.5.2 Описание метода

5.5.2.1 Анализ углового распределения

Методы реконструкции для задачи отражательной ультразвуковой томографии, предложенные в работах [102, 103], проектируют исходные амплитуды $I(r, j)$ в глобальную систему координат путем их равномерного распределения по соответствующей дуге $A(r, j)$. Данные методы основаны на допущении о том, что амплитуда $I(r, j)$ вызвана набором отражателей, распределенных равномерно вдоль всей дуги. Для того, чтобы избежать использования такого грубого допущения, автором был разработан метод, который оценивает полное угловое распределение для каждой из дуг в явном виде. Было экспериментально показано, что использование такой

оценки вместо равномерного распределения позволяет улучшить результаты реконструкции [100].

Для того, чтобы оценить угловое распределение, для каждой дуги рассмотрим некоторую точку $p(\Theta; r, j) \in A(r, j)$, заданную при помощи угла Θ (рис. 5.15 (а)). При этом каждая точка соответствует некоторой синусоидо-образной кривой $S(\Theta; r, j)$ в пространстве данных, а семейство таких кривых для разных углов Θ пересекается в точке (r, j) (рис. 5.15 (б)).

Такой метод основан на допущении о том, что отражатель, расположенный в точке $p(\Theta; r, j) \in A(r, j)$ вызовет высокую амплитуду сигнала для нескольких последовательных положений приемника $j' \in [j - \delta, j + \delta]$. При этом вычисление углового распределения позволяет установить, какие из точек на дуге $p(\Theta; r, j)$ действительно содержат отражатель.

5.5.2.2 Метод реконструкции

Для вычисления углового распределения будем использовать следующий набор направленных фильтров: каждый фильтр получается из части синусоиды, соответствующей окну вокруг точки (r, j) в пространстве данных (рис. 5.15 (б)). Каждый из фильтров затем сворачивается с фрагментом изображения для того, чтобы оценить угловое распределение в соответствующей точке входных данных (рис. 5.15 (в)):

$$\tilde{p}(\Theta; r, j) = T(r, j) * f_{t, \Theta}, \quad (5.30)$$

где $T(r, j)$ - фрагмент изображения с центром в (r, j) , и $f_{t, \Theta}$ - фильтр, содержащий соответствующую часть синусоиды $S(\Theta; r, j)$.

Полученное угловое распределение $\tilde{p}(\Theta; r, j)$ затем нормируется при помощи функции *softmax*:

$$p(\Theta; r, j) = \frac{\exp(\alpha \tilde{p}(\Theta; r, j))}{\int_{\Phi} \exp(\alpha \tilde{p}(\Phi; r, j)) d\Phi}, \quad (5.31)$$

Полученное нормированное распределение $p(\Theta; r, j)$ используется для проектирования из пространства данных в глобальную систему координат установки:

$$R(x, y) = \sum_{(x, y) = p(\Theta; r, j)} I(r, j) p(r, j, \Theta), \quad (5.32)$$

таким образом интенсивность в точке (x, y) суммируется из точек в пространстве данных (r, j) , таких что $(x, y) \in A(r, j)$, и вклад из точек (r, j) пропорционален интенсивностям $I(r, j)$, взвешенным соответствующими угловыми распределениями.

При этом параметр α в выражении (5.31) позволяет менять степень усиления максимумов при использовании функции *softmax*: более высокие значения параметра соответствуют более резким границам при реконструкции, а более низкие значения повышают устойчивость метода к наличию шумов. В пределе $\alpha \rightarrow +\infty$, нормированное угловое распределение соответствует дельта функции, которая помещает всю интенсивность $I(r, j)$ в единственную точку $p(\Theta; r, j)$ с углом $\Theta = \arg \max_{\Theta} \tilde{p}(\Theta; r, j)$. Далее в подразделе, такой метод имеет название *метода доминантной ориентации*. В пределе при значении α , стремящемся к нулю, получаем метод, соответствующий равномерному угловому распределению.

5.5.3 Численные эксперименты

5.5.4 Эксперименты на синтетических данных

Данный подраздел содержит описание эксперимента на основе численного моделирования. В таком эксперименте результат реконструкции сравнивается с эталонным изображением, который был использован для генерации синтетических входных данных. Приводятся сравнения реконструкции на основе трех методов: метод на основе равномерного распределения, метод на основе доминантной ориентации и метод на основе оценки углового распределения.

Ниже приводятся результаты, основанные на численном моделировании распространения и отражения звуковых волн во временной области. Моделирование производилось с использованием программного пакета K-wave [104]. Был рассмотрен двумерный срез объема, содержащего отражатель эллиптической формы, что соответствует двумерному срезу кости человека. Для внутренней части эллипса была задана скорость света 1800 м/с, для внешней части эллипса - 1509 м/с. Из каждого положения приемника-отражателя был запущен волновой пакет с несущей частотой 5 МГц, затем записано значение амплитуды отраженного сигнала. Итого, входное изображение содержит 1500 позиций приемника-получателя, для каждого записана амплитуда по 1000 временным отсчетам. Форма отражающей поверхности и соответствующие результаты реконструкции показаны на рис 5.15 (первая строка). Для того, чтобы сравнить устойчивость рассмотренных методов, ко входному изображению был добавлен Гауссов шум. Результаты, представленные на рис. 5.15 (вторая строка) показывают превосходство двух методов на основе анализа углового распределения над методом на основе равномерного распределения.

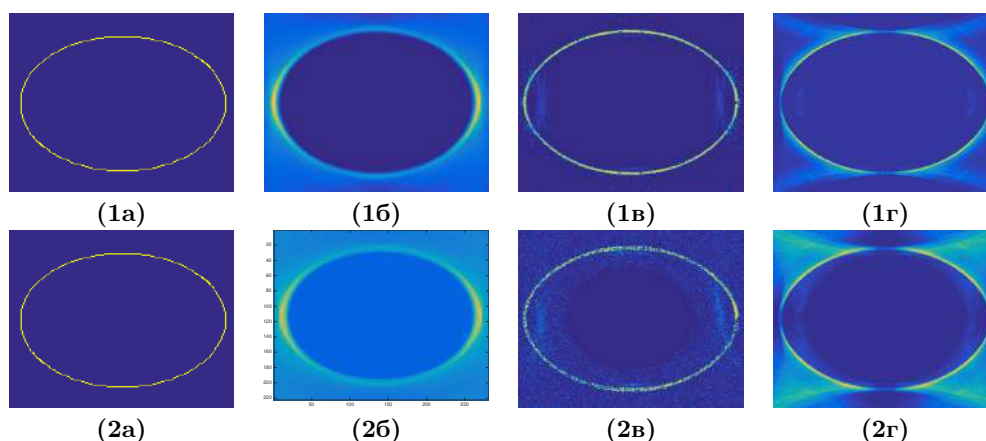


Рис. 5.15: Результаты реконструкции для синтетических входных данных, основанных на моделировании сильного перепада скоростей звука. (первый столбец) Форма границ перепада двух сред. Реконструкция отражательной способности с использованием равномерного распределения (второй столбец), доминантной ориентации, (третий столбец), полного углового распределения (четвертый столбец). Вторая строка содержит результат реконструкции для аналогичного эксперимента с дополнительным Гауссовым шумом, добавленным ко входному изображению.

5.5.5 Эксперименты на реальных данных

Данный подраздел содержит результаты численных экспериментов на реальных данных, полученных с ультразвукового томографа, описанного в работе [101]. Размерность входных данных соответствует 1500 позициям приемника-отражателя, для каждого записана амплитуда для 10000 временных интервалов.

Рис. 5.16 содержит результаты реконструкции для двух срезов руки человека, примерно на 15 см. ниже локтя. Результат реконструкции на основе равномерного распределения, показанный на рис. 5.16 (а,г) имеет большое количество артефактов, сконцентрированных вокруг вероятных положений отражателей. Метод, основанный на доминантной ориентации (рис.5.16 (б,д)) позволяет существенно снизить число артефактов, однако, ведет к появлению шума на выходных изображениях. Наиболее качественный результат получается с использованием метода на основе полного углового распределения (рис. 5.16 (в,е)), который позволяет наблюдать при реконструкции дополнительные детали, которые отсутствуют при использовании других методов.

5.5.6 Оценка вычислительной сложности

Вычислительная сложность предложенного метода реконструкции оценивается как $\mathcal{O}(TSF)$, где T - количество положений приемников-передатчиков, S - количество

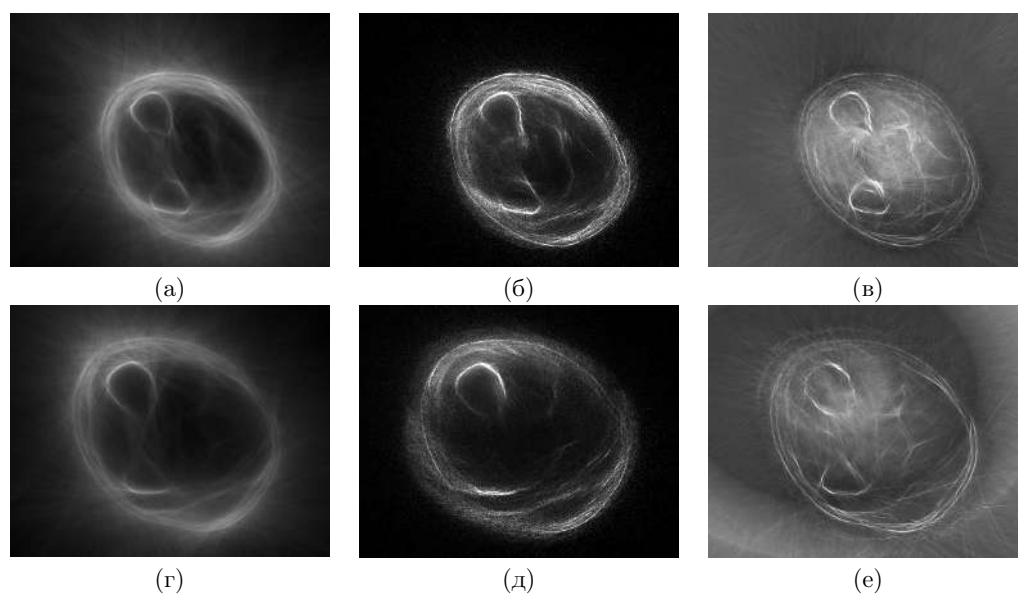


Рис. 5.16: Реконструкция отражательной способности ткани среза руки человека ниже локтя. Первая строка изображений соответствует первому и второму эксперименту, произведенному для разных людей. Первая колонка соответствует использованию равномерного распределения, вторая колонка соответствует методу доминантной ориентации, третья колонка соответствует методу оценки углового распределения.

временных интервалов для каждого положения и F - количество фильтров в наборе, используемом для анализа углового распределения (число дискретных значений угла Θ).

Каждый из рассмотренных методов может быть эффективно реализован с использованием графических ускорителей ввиду независимости вычислений, производимых для каждого из пикселей входного изображения.

5.6 Выводы

Ультразвуковая эластография - задача, сходная задаче вычислению оптического потока, однако в такой задаче возникают практические трудности, специфические для ультразвуковых изображений. Вариационные подходы, основанные на линейаризации изображений не могут быть применены для эластографии ввиду большого количества шума, свойственного ультразвуковым изображениям. Поэтому, большинство методов для эластографии полагается на полный перебор при сопоставлении изображений. Одним из направлений, позволяющих улучшить результат является применение нескольких кадров с целью увеличения используемой степени сжатия.

Другим важным аспектом построения методов для задачи эластографии является разработка подходов к регуляризации, устойчивых к участками неверного сопоставления. Экспериментальное сравнение различных методов регуляризации, проведенное автором, показывает эффективность предложенного адаптивного метода регуляризации в подобных случаях [14].

В данной главе также рассмотрена задача сопоставления ультразвуковых сигналов в отражательной томографии. Такая задача схода к задаче сопоставления строк изображения при стерео-реконструкции. Автором было показана важность анализа угловых распределений при построения метода реконструкции [100].

Глава 6

Заключение

В настоящей работе предложено три метода для задачи сопоставления изображений, основанных на глубоком машинном обучении и решении оптимизационной задачи. По сравнению со многими другими методами, предложенными в литературе, разработанные алгоритмы достаточно эффективны на этапе исполнения и позволяют решать задачу в реальном времени с использованием современных графических ускорителей. Было рассмотрено три варианта задачи сопоставления изображений: плотное бинокулярное стерео-сопоставление, вычисление оптического потока и ультразвуковая эластография. В каждой из задач требуется оценить поле смещений на основе пары или серии кадров: в случае стерео-сопоставления поле смещения является одномерным, в двух других случаях требуется оценить произвольное двумерное поле смещений. В работе было показано, что перспективными методами уменьшения ошибки сопоставления являются использование машинного обучения на реальных данных, а также использование информации с серии, а не пары изображений.

Предложенный метод для задачи стерео-сопоставления основан на агрегировании тензора энергий. В силу того, что поле смещений в этой задаче является одномерным, тензор энергий сопоставления является трехмерным и занимает сравнительно небольшой объем памяти, достаточный для хранения в явном виде и последующей обработки. Метод агрегирования энергий построен на основе рекурсивного фильтра, учитывающего границы объектов на изображении. При этом границы, релевантные для задачи, предсказываются при помощи сверточной нейросети. Так, предложенная сверточно-рекуррентная модель принимает на вход пару изображений и вычисляет поле смещений без необходимости постобработки. В отличие от методов, предложенных в литературе, в такой модели отсутствует необходимость сравнения

большого количества глубоких дескрипторов большой размерности, что существенно снизило вычислительную сложность и позволило получить реализацию, которая работает в реальном времени.

Проведенные численные эксперименты на реальных данных показали важность использования информации о границах объектов на изображении. Модель, предложенная автором основана на применении машинного обучения и позволяет избежать необходимости решения вычислительно трудоемкой задачи дискретной оптимизации.

Предложенный метод для задачи вычисления оптического потока использует обучение оператора регуляризации. Новая архитектура сверточной нейросети основана на представлении графа вычислений метода оптимизации энергий в качестве слоев нейронной сети: был использован быстрый двойственный градиентный метод. В соответствии с классической моделью для числения оптического потока в реальном времени, метод основан на линеаризации изображения с использованием текущей оценки поля смещений. Такая линеаризация позволяет вычислять поля смещений ограниченной величины, однако применение обучаемой регуляризации позволяет заметно уменьшить ошибку сопоставления по сравнению с фиксированной регуляризацией. В частности, обучаемая регуляризация позволяет более качественно оценивать смещения вблизи границ движущихся объектов, а также уменьшить количество выбросов, наблюдаемое при использовании оптимизационного метода.

Проведенные численные эксперименты подтвердили эффективность комбинирования вариационных алгоритмов и современных методов на основе глубоких сверточных нейросетей. Представление графа вычислений вариационных методов в качестве слоев нейросети представляется перспективным способом построения моделей для решения задач обработки изображений.

Предложенный метод сопоставления медицинских ультразвуковых изображений для эластографии основан на применении адаптивной регуляризации. В силу отсутствия обучающих данных, при решении задачи сопоставления ультразвуковых изображений, представляется перспективным подход, использующий информацию о качестве полученного сопоставления для последующего сглаживания полей деформации. Так, в участках совпадения с низкой ошибкой, метод полагается на выходные данные, а в участках выбросов, метод полагается на регуляризацию. Еще одним важным аспектом предложенного метода является использование трех кадров. Использование двухэтапного метода поиска смещений позволяет более точно оценивать смещения для больших значений контактной силы, что позволяет улучшить возможности применения эластографии для медицинской диагностики.

Серия численных экспериментов показала возможность детекции жестких включений в органических тканях размером вплоть до 2 мм с использованием стандартной ультразвуковой пробы, используемой для получения УЗИ снимков. Автором была разработана эффективная реализация метода, которая позволяет решать задачу сопоставления за время, сравнимое с получением данных при использовании графического ускорителя.

Результаты настоящей работы указали нескольких дальнейших направлений для исследований. Исходя из проведенных экспериментов, наиболее перспективным представляется построение моделей, совмещающих построение глубоких дескрипторов и обучаемую регуляризацию, учитывающую границы объектов на изображении. При этом возможно использование серии изображений в случае оценки полей смещения для кадров видеопоследовательности. В случае медицинских ультразвуковых изображений, перспективным представляется построение экспериментальных методов для получения обучающих данных, что позволило бы уменьшить ошибку сопоставления за счет применения машинного обучения.

Литература

- [1] Szeliski Richard. Computer vision: algorithms and applications. Springer Science & Business Media, 2010.
- [2] Zach Christopher, Pock Thomas, Bischof Horst. A duality based approach for realtime TV-L 1 optical flow // Pattern Recognition. 2007. С. 214–223.
- [3] Flownet: Learning optical flow with convolutional networks / Alexey Dosovitskiy, Philipp Fischer, Eddy Ilg [и др.] // Proceedings of the IEEE International Conference on Computer Vision. 2015. С. 2758–2766.
- [4] Flownet 2.0: Evolution of optical flow estimation with deep networks / Eddy Ilg, Nikolaus Mayer, Tonmoy Saikia [и др.] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017.
- [5] Elastography: a quantitative method for imaging the elasticity of biological tissues. / J. Ophir, I. Céspedes, H. Ponnekanti [и др.] // Ultrason Imaging. 1991. Т. 13, № 2. С. 111–134.
- [6] Robust dynamic programming method for ultrasound elastography / Ioana Fleming, Hassan Rivaz, Emad Boctor [и др.] // Proceedings of the SPIE Medical Imaging. Т. 8320. 2012. С. 83201–83206.
- [7] Ultrasound Elastography Using Multiple Images / H. Rivaz, E Boctor, M. Choti [и др.] // Medical Image Analysis. 2014. Т. 18. С. 314–329.
- [8] Zbontar Jure, LeCun Yann. Stereo matching by training a convolutional neural network to compare image patches // Journal of Machine Learning Research. 2016. Т. 17. С. 1–32.
- [9] Luo Wenjie, Schwing Alexander G, Urtasun Raquel. Efficient deep learning for stereo matching // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016. С. 5695–5703.

- [10] Xu Jia, Ranftl René, Koltun Vladlen. Accurate optical flow via direct cost volume processing // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017.
- [11] Kuzmin Andrey, Mikushin Dmitry, Lempitsky Victor. End-to-end Learning of Cost-Volume Aggregation for Real-time Dense Stereo // Machine Learning for Signal Processing, 2017. MLSP 2017. IEEE Conference on / IEEE. 2017.
- [12] Кузьмин А.И. Обучение оператора регуляризации в задаче вычисления оптического потока // Программирование. 2018. Т. 44, № 3. С. 3–15.
- [13] Zbontar Jure, LeCun Yann. Computing the stereo matching cost with a convolutional neural network // IEEE Conference on Computer Vision and Pattern Recognition, CVPR. 2015. С. 1592–1599.
- [14] Multi-frame elastography using a handheld force-controlled ultrasound probe / Andrey Kuzmin, Aaron M Zakrzewski, Brian W Anthony [и др.] // IEEE transactions on ultrasonics, ferroelectrics, and frequency control. 2015. Т. 62, № 8. С. 1486–1500.
- [15] High-resolution stereo datasets with subpixel-accurate ground truth / Daniel Scharstein, Heiko Hirschmüller, York Kitajima [и др.] // German Conference on Pattern Recognition / Springer. 2014. С. 31–42.
- [16] Szeliski Richard. A multi-view approach to motion and stereo // Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on. / IEEE. Т. 1. 1999. С. 157–163.
- [17] Horn Berthold KP, Schunck Brian G. Determining optical flow // Artificial intelligence. 1981. Т. 17, № 1-3. С. 185–203.
- [18] Evaluation of different methods for using colour information in global stereo matching approaches / Michael Bleyer, Sylvie Chambon, Uta Poppe [и др.] // Int. Society for Photogrammetry and Remote Sensing. 2008. С. 63–68.
- [19] Hartley Richard, Zisserman Andrew. Multiple view geometry in computer vision. Cambridge university press, 2003.
- [20] Cross-View People Tracking by Scene-Centered Spatio-Temporal Parsing. / Yuanlu Xu, Xiaobai Liu, Lei Qin [и др.] // AAAI. 2017. С. 4299–4305.
- [21] Zahiri-Azar Reza, Salcudean Septimiu E. Motion estimation in ultrasound images using time domain cross correlation with prior estimates // IEEE Transactions on Biomedical Engineering. 2006. Т. 53, № 10. С. 1990–2000.

- [22] Robust and Precise 3D-Modelling of Traffic Scenes based on Dense Stereo Vision / David Pfeiffer, Alexander Barth, Uwe Franke [и др.] // vldb. informatik. hu-berlin.de. 2009. T. 1.
- [23] Menze Moritz, Geiger Andreas. Object Scene Flow for Autonomous Vehicles // Conference on Computer Vision and Pattern Recognition (CVPR). 2015.
- [24] Clinical application of breast elastography: state of the art / Paolo Ricci, Elena Maggini, Ester Mancuso [и др.] // European journal of radiology. 2014. T. 83, № 3. С. 429–437.
- [25] Hoskins Peter R, Svensson W. Current state of ultrasound elastography // Ultrasound-Journal of the British Medical Ultrasound Society. 2012. T. 20, № 1. с. 3.
- [26] Fast cost-volume filtering for visual correspondence and beyond / Christoph Rhemann, Asmaa Hosni, Michael Bleyer [и др.] // Computer Vision and Pattern Recognition (CVPR), IEEE Conference on / IEEE. 2011. С. 3017–3024.
- [27] Cigla Cevahir. Recursive Edge-Aware Filters for Stereo Matching // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2015. С. 27–34.
- [28] Yang Qingxiong. A non-local cost aggregation method for stereo matching // Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on / IEEE. 2012. С. 1402–1409.
- [29] Hirschmuller Heiko. Accurate and efficient stereo processing by semi-global matching and mutual information // 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) / IEEE. T. 2. 2005. С. 807–814.
- [30] Blake Andrew, Kohli Pushmeet, Rother Carsten. Markov random fields for vision and image processing. Mit Press, 2011.
- [31] Scharstein Daniel, Szeliski Richard. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms // International journal of computer vision. 2002. T. 47, № 1-3. С. 7–42.
- [32] Paris Sylvain, Durand Frédo. A fast approximation of the bilateral filter using a signal processing approach // Computer Vision–ECCV 2006. 2006. С. 568–580.
- [33] Barron Jonathan T, Poole Ben. The fast bilateral solver // European Conference on Computer Vision / Springer. 2016. С. 617–632.

- [34] He Kaiming, Sun Jian, Tang Xiaoou. Guided image filtering // IEEE transactions on pattern analysis and machine intelligence. 2013. T. 35, № 6. C. 1397–1409.
- [35] Gastal Eduardo SL, Oliveira Manuel M. Domain transform for edge-aware image and video processing // ACM Transactions on Graphics (TOG) / ACM. T. 30. 2011. c. 69.
- [36] End-to-End Training of Hybrid CNN-CRF Models for Stereo / Patrick Knöbelreiter, Christian Reinbacher, Alexander Shekhovtsov [и др.] // Computer Vision and Pattern Recognition, 2017. CVPR 2017. IEEE Conference on / IEEE. 2017. C. 1456–1465.
- [37] End-to-End Learning of Geometry and Context for Deep Stereo Regression / Alex Kendall, Hayk Martirosyan, Saumitro Dasgupta [и др.] // Computer Vision and Pattern Recognition, 2017. CVPR 2017. IEEE Conference on / IEEE. 2017. C. 66–75.
- [38] Yoon Kuk-Jin, Kweon In So. Adaptive support-weight approach for correspondence search. // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2006. T. 28, № 4. C. 650–656.
- [39] Pham Cuong Cao, Jeon Jae Wook. Domain transformation-based efficient cost aggregation for local stereo matching // IEEE Transactions on Circuits and Systems for Video Technology. 2013. T. 23, № 7. C. 1119–1130.
- [40] Semantic image segmentation with task-specific edge detection using cnns and a discriminatively trained domain transform / Liang-Chieh Chen, Jonathan T Barron, George Papandreou [и др.] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016. C. 4545–4554.
- [41] Learning phrase representations using RNN encoder-decoder for statistical machine translation / Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre [и др.] // Proceedings of the Conference on Empirical Methods on Natural Language Processing. 2014.
- [42] Conditional random fields as recurrent neural networks / Shuai Zheng, Sadeep Jayasumana, Bernardino Romera-Paredes [и др.] // Proceedings of the IEEE International Conference on Computer Vision. 2015. C. 1529–1537.
- [43] Schwing Alexander G., Urtasun Raquel. Fully Connected Deep Structured Networks // CoRR. 2015. T. abs/1503.02351.
- [44] Zabih Ramin, Woodfill John. Non-parametric local transforms for computing visual correspondence // European conference on computer vision / Springer. 1994. C. 151–158.

- [45] Xie Saining, Tu Zhuowen. Holistically-nested edge detection // Proceedings of the IEEE International Conference on Computer Vision. 2015. С. 1395–1403.
- [46] On the Properties of Neural Machine Translation: Encoder–Decoder Approaches / Kyunghyun Cho, Bart van Merriënboer, Dzmitry Bahdanau [и др.] // Syntax, Semantics and Structure in Statistical Translation. 2014. с. 103.
- [47] Geiger Andreas, Lenz Philip, Urtasun Raquel. Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite // Conference on Computer Vision and Pattern Recognition (CVPR). 2012.
- [48] Kingma Diederik, Ba Jimmy. Adam: A method for stochastic optimization // Proceedings of the International Conference on Learning Representations (ICLR). 2015.
- [49] A Database of Human Segmented Natural Images and its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics / D. Martin, C. Fowlkes, D. Tal [и др.] // Proc. 8th Int'l Conf. Computer Vision. Т. 2. 2001. July. С. 416–423.
- [50] Theano: a CPU and GPU Math Expression Compiler / James Bergstra, Olivier Breuleux, Frédéric Bastien [и др.] // Proceedings of the Python for Scientific Computing Conference (SciPy). 2010. .
- [51] Rudin Leonid I, Osher Stanley, Fatemi Emad. Nonlinear total variation based noise removal algorithms // Physica D: Nonlinear Phenomena. 1992. Т. 60, № 1-4. С. 259–268.
- [52] Bredies Kristian, Kunisch Karl, Pock Thomas. Total generalized variation // SIAM Journal on Imaging Sciences. 2010. Т. 3, № 3. С. 492–526.
- [53] Bredies K., Kunisch K., Pock T. Total generalized variation // SIAM J. Imag. Sc. 2010. Т. 3. С. 492–526.
- [54] Second order total generalized variation (TGV) for MRI / Florian Knoll, Kristian Bredies, Thomas Pock [и др.] // Magnetic resonance in medicine. 2011. Т. 65, № 2. С. 480–491.
- [55] Roth Stefan, Black Michael J. Fields of experts: A framework for learning image priors // Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on / IEEE. Т. 2. 2005. С. 860–867.
- [56] Olshausen Bruno A, Field David J. Sparse coding with an overcomplete basis set: A strategy employed by V1? // Vision research. 1997. Т. 37, № 23. С. 3311–3325.

- [57] Hinton Geoffrey E. Training products of experts by minimizing contrastive divergence // *Neural computation*. 2002. T. 14, № 8. С. 1771–1800.
- [58] Brox Thomas, Bregler Christoph, Malik Jitendra. Large displacement optical flow // *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on / IEEE*. 2009. С. 41–48.
- [59] Epicflow: Edge-preserving interpolation of correspondences for optical flow / Jerome Revaud, Philippe Weinzaepfel, Zaid Harchaoui [и др.] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015. С. 1164–1172.
- [60] DeepFlow: Large displacement optical flow with deep matching / Philippe Weinzaepfel, Jerome Revaud, Zaid Harchaoui [и др.] // *Proceedings of the IEEE International Conference on Computer Vision*. 2013. С. 1385–1392.
- [61] Coarse-to-fine PatchMatch for Dense Correspondence / Yunsong Li, Yinlin Hu, Rui Song [и др.] // *IEEE Transactions on Circuits and Systems for Video Technology*. 2017. с. 43.
- [62] An improved algorithm for TV-L1 optical flow / Andreas Wedel, Thomas Pock, Christopher Zach [и др.] // *Statistical and geometrical approaches to visual motion analysis*. Springer, 2009. С. 23–45.
- [63] Pushing the limits of stereo using variational stereo estimation / Rene Ranftl, Stefan Gehrig, Thomas Pock [и др.] // *Intelligent Vehicles Symposium (IV), 2012 IEEE / IEEE*. 2012. С. 401–407.
- [64] Werlberger Manuel. Convex approaches for high performance video processing. 2012.
- [65] Werlberger Manuel, Pock Thomas, Bischof Horst. Motion estimation with non-local total variation regularization // *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on / IEEE*. 2010. С. 2464–2471.
- [66] Ranftl René, Bredies Kristian, Pock Thomas. Non-local total generalized variation for optical flow estimation // *European Conference on Computer Vision / Springer*. 2014. С. 439–454.
- [67] Steinbrücker Frank, Pock Thomas, Cremers Daniel. Large displacement optical flow computation without warping // *Computer Vision, 2009 IEEE 12th International Conference on / IEEE*. 2009. С. 1609–1614.

- [68] Schmidt Uwe, Roth Stefan. Shrinkage fields for effective image restoration // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014. С. 2774–2781.
- [69] Lefkimmiatis Stamatios. Non-local color image denoising with convolutional neural networks // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017.
- [70] Wang Shenlong, Fidler Sanja, Urtasun Raquel. Proximal deep structured models // Advances in Neural Information Processing Systems. 2016. С. 865–873.
- [71] Spatial transformer networks / Max Jaderberg, Karen Simonyan, Andrew Zisserman [и др.] // Advances in Neural Information Processing Systems. 2015. С. 2017–2025.
- [72] A naturalistic open source movie for optical flow evaluation / Daniel J Butler, Jonas Wulff, Garrett B Stanley [и др.] // European Conference on Computer Vision / Springer. 2012. С. 611–625.
- [73] Heeger David J. Optical flow using spatiotemporal filters // International journal of computer vision. 1988. Т. 1, № 4. С. 279–302.
- [74] Garg Ravi, Roussos Anastasios, Agapito Lourdes. A variational approach to video registration with subspace constraints // International journal of computer vision. 2013. Т. 104, № 3. С. 286–314.
- [75] Bamber J., Bush N. Freehand elasticity imaging using speckle decorrelation rate. // Acoust. Imaging. 1996. Т. 22. С. 285–292.
- [76] Elastography: ultrasonic estimation and imaging of the elastic properties of tissues / Jonathan Ophir, S Kaiser Alam, Brian Garra [и др.] // Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine. 1999. Т. 213, № 3. С. 203–233.
- [77] The AutoQual ultrasound elastography method for quantitative assessment of lateral strain in post-rupture Achilles tendons. / Phillip G. Brown, Joseph Alsousou, Ashley Cooper [и др.] // Journal of Biomechanics. 2013. Т. 46. С. 2695–2700.
- [78] Robust dynamic programming method for ultrasound elastography / I. Fleming, H. Rivaz, E. Voctor [и др.] // Proc. SPIE 8320. 2012.
- [79] Ultrasound elastography based on multiscale estimations of regularized displacement fields. / C. Pellot-Barakat, F. Frouin, M. Insana [и др.] // IEEE Trans. Medical Imaging. 2004. Т. 23, № 2. С. 153–163.

- [80] Kallel F., Ophir J. A Least-squares strain estimator for elastography. // *Ultrason. Imaging*. 1997. T. 19, № 3. C. 195–208.
- [81] Recent Results in Nonlinear Strain and Modulus Imaging. / Timothy J. Hall, Paul E. Barbone, Assad A. Oberai [и др.] // *Curr. Med. Imaging Rev.* 2011. T. 7, № 4. C. 313–327.
- [82] Axial Strain Calculation Using a Low-Pass Digital Differentiator in Ultrasound Elastography. / Jianwen L., Jing Bai, Ping He [и др.] // *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*. 2004. T. 51, № 9. C. 1119–1127.
- [83] Techavipoo U., Varghese T. Wavelet Denoising of Displacement Estimates in Elastography. // *Ultrasound Med. Biol.* 2004. T. 30, № 4. C. 477–491.
- [84] Chartrand R. Numerical differentiation of noise, nonsmooth data. // *ISRN Applied Mathematics*. 2011.
- [85] Chaturvedi Pawan, Insana Michael F., Hall Timothy J. 2-D Companding for Noise Reduction in Strain Imaging // *IEEE TUFFC*. 1998. C. 179–191.
- [86] Revell J., Mirmehdi M., McNally D. Computer vision elastography: Speckle adaptive motion estimation for elastography using ultrasound sequences. // *IEEE Trans. Med. Imag.* 2005. T. 24, № 6. C. 755–766.
- [87] Gilbertson M., Anthony B. Ergonomic control strategies for a handheld force-controlled ultrasound probe // *IEEE IROS*. 2012. C. 1284–1291.
- [88] Zhu Mingqiang, Chan Tony. An efficient primal-dual hybrid gradient algorithm for total variation image restoration // *UCLA CAM Report*. 2008. T. 34.
- [89] Knoll Florian et al. Fast reduction of undersampling artifacts in radial MR angiography with 3D total variation on graphics hardware // *Magn. Reson. Mater. Phy.* 2010. T. 23. C. 103–114.
- [90] Zhu Mingqiang, Chan T. An efficient primal-dual hybrid gradient algorithm for total variation image restoration // *UCLA CAM Report*. 2008. C. 8–34.
- [91] Bell Nathan, Hoberock Jared. Thrust: A productivity-oriented library for CUDA // *GPU computing gems Jade edition*. 2011. T. 2. C. 359–371.
- [92] DiBattista Andrew, Noble J. An efficient block matching and spectral shift estimation algorithm with applications to ultrasound elastography // *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*. 2014. T. 61, № 3. C. 407–419.

- [93] Kallel Faouzi, Ophir Jonathan. A least-squares strain estimator for elastography // Ultrasonic imaging. 1997. T. 19, № 3. C. 195–208.
- [94] Jiang Jingfeng, Hall Timothy J. A generalized speckle tracking algorithm for ultrasonic strain imaging using dynamic programming // Ultrasound in medicine & biology. 2009. T. 35, № 11. C. 1863–1879.
- [95] Smith Michael. ABAQUS/Standard User's Manual, Version 6.9. Simulia, 2009.
- [96] Jensen Jørgen Arendt, Svendsen Niels Bruun. Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers // IEEE Trans. Ultrasonics, Ferroelectrics and Frequency Control. 1992. T. 39, № 2. C. 262–267.
- [97] Jensen Jørgen Arendt. Field: A program for simulating ultrasound systems // 10th Nordicbaltic Conference on Biomedical Imaging. 1996. T. 4. C. 351–353.
- [98] Gilbertson Matthew W, Anthony Brian W. Ergonomic control strategies for a handheld force-controlled ultrasound probe // Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on / IEEE. 2012. C. 1284–1291.
- [99] Real-time processing in dynamic ultrasound elastography: A GPU-based implementation using CUDA / Emmanuel Montagnon, Sami Hissouiny, Philippe Després [и др.] // Information Science, Signal Processing and their Applications (ISSPA), 2012 11th International Conference on / IEEE. 2012. C. 472–477.
- [100] Fast low-cost single element ultrasound reflectivity tomography using angular distribution analysis / Audrey Kuzmin, Xiang Zhang, Jonathan Finche [и др.] // Biomedical Imaging (ISBI), 2016 IEEE 13th International Symposium on / IEEE. 2016. C. 1021–1024.
- [101] A Single Element 3D Ultrasound Tomography System / Xiang Zhang, Jonathan Fincke, Andrey Kuzmin [и др.] // Proc. of 37th annual conference of the IEEE Engineering in Medicine and Biology Society.
- [102] Norton Stephen J, Linzer Melvin. Ultrasonic reflectivity tomography: reconstruction with circular transducer arrays // Ultrasonic Imaging. 1979. T. 1, № 2. C. 154–184.
- [103] Norton Stephen J. Reconstruction of a two-dimensional reflecting medium over a circular domain: Exact solution // The Journal of the Acoustical Society of America. 1980. T. 67, № 4. C. 1266–1273.

-
- [104] Treeby Bradley E, Cox Benjamin T. k-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields // Journal of biomedical optics. 2010. T. 15, № 2. C. 021314–021314.
- [105] Guney Fatma, Geiger Andreas. Displets: Resolving stereo ambiguities using object knowledge // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015. C. 4165–4175.
- [106] Vogel Christoph, Schindler Konrad, Roth Stefan. 3D scene flow estimation with a piecewise rigid scene model // International Journal of Computer Vision. 2015. T. 115, № 1. C. 1–28.
- [107] Einecke Nils, Eggert Julian. A multi-block-matching approach for stereo // 2015 IEEE Intelligent Vehicles Symposium (IV) / IEEE. 2015. C. 585–592.
- [108] Einecke Nils, Eggert Julian. Stereo image warping for improved depth estimation of road surfaces // Intelligent Vehicles Symposium (IV), 2013 IEEE / IEEE. 2013. C. 189–194.
- [109] DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. / LC Chen, G Papandreou, I Kokkinos [и др.] // IEEE transactions on pattern analysis and machine intelligence. 2017.
- [110] Simonyan Karen, Zisserman Andrew. Very deep convolutional networks for large-scale image recognition // Proceedings of the International Conference on Learning Representations (ICLR). 2015.
- [111] A Large Dataset to Train Convolutional Networks for Disparity, Optical Flow, and Scene Flow Estimation / Nikolaus Mayer, Eddy Ilg, Philip Häusser [и др.] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition CVPR. 2016.
- [112] The SYNTHIA Dataset: A Large Collection of Synthetic Images for Semantic Segmentation of Urban Scenes / German Ros, Laura Sellart, Joanna Materzynska [и др.] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition CVPR. 2016.
- [113] Hochreiter Sepp, Schmidhuber Jürgen. Long short-term memory // Neural computation. 1997. T. 9, № 8. C. 1735–1780.
- [114] Barron Jonathan T, Poole Ben. The Fast Bilateral Solver // European conference on computer vision (ECCV). 2016.

- [115] Liu Sifei, Pan Jinshan, Yang Ming-Hsuan. Learning Recursive Filters for Low-Level Vision via a Hybrid Neural Network // European Conference on Computer Vision / Springer. 2016. C. 560–576.
- [116] Pfeiffer David, Franke Uwe. Efficient representation of traffic scenes by means of dynamic stixels // Intelligent Vehicles Symposium (IV), 2010 IEEE / IEEE. 2010. C. 217–224.
- [117] Ranftl Rene, Pock Thomas, Bischof Horst. Minimizing TGV-based Variational Models with Non-Convex Data terms // International Conference on Scale Space and Variational Methods in Computer Vision (SSVM). 2013.
- [118] Steinbrücker Frank, Pock Thomas, Cremers Daniel. Large displacement optical flow computation without warping // ICCV. 2009. C. 1609–1614.
- [119] Elasticity reconstruction from displacement and confidence measures of a multi-compressed ultrasound RF sequence. / J. Li, Y. Cui, M. Kadour [и др.] // IEEE Trans. Ultrason., Ferroelectr., Freq. Control. 2008. T. 55, № 2. C. 319–326.
- [120] Cooper D., Graham J. Estimating motion in noisy, textured image: Optical flow in medical ultrasound. // 7th British Machine Vision Conference. 1996. C. 585–594.
- [121] Cooper D., Madsen B., Graham J. Estimation motion in ultrasound images of the small bowel: Optical flow without image structure // Lecture Notes in Computer Science. 2003. T. 2749, № 7. C. 1017–1026.
- [122] Avarezi L., Weickert J., Sanchez J. Reliable estimation of dense optical flow fields with large displacements. // ICCV. 2000. T. 39, № 1. C. 41–56.
- [123] Vogel C. R., Oman M. E. Iterative methods for total variation denoising. // SIAM J. Sci. Comp. 1996. T. 17, № 1. C. 227–238.
- [124] Real-Time Regularized Ultrasound Elastography. / H. Rivaz, E. Boctor, M. Choti [и др.] // IEEE Trans. Medical Imaging. 2011. T. 30, № 4. C. 928–945.
- [125] Chambolle Antonin. Total variation minimization and a class of binary MRF models // EMMCVPR / Springer. T. 5. 2005. C. 136–152.
- [126] Geiger Andreas, Lenz Philip, Urtasun Raquel. Are we ready for autonomous driving? the kitti vision benchmark suite // Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on / IEEE. 2012. C. 3354–3361.
- [127] Sun Deqing, Roth Stefan, Black Michael J. Secrets of optical flow estimation and their principles // Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on / IEEE. 2010. C. 2432–2439.

- [128] Chen Qifeng, Koltun Vladlen. Full flow: Optical flow estimation by global optimization over regular grids // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016. С. 4706–4714.
- [129] Menze Moritz, Heipke Christian, Geiger Andreas. Discrete optimization for optical flow // German Conference on Pattern Recognition / Springer. 2015. С. 16–28.
- [130] Güney Fatma, Geiger Andreas. Deep discrete flow // Asian Conference on Computer Vision / Springer. 2016. С. 207–224.
- [131] Hu Yinlin, Li Yunsong, Song Rui. Robust Interpolation of Correspondences for Large Displacement Optical Flow // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017.
- [132] Zweig Shay, Wolf Lior. InterpoNet, A brain inspired neural network for optical flow dense interpolation // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017.
- [133] Exploiting semantic information and deep matching for optical flow / Min Bai, Wenjie Luo, Kaustav Kundu [и др.] // European Conference on Computer Vision / Springer. 2016. С. 154–170.
- [134] Optical flow with semantic segmentation and localized layers / Laura Sevilla-Lara, Deqing Sun, Varun Jampani [и др.] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016. С. 3889–3898.
- [135] Hur Junhwa, Roth Stefan. Joint optical flow and temporally consistent semantic segmentation // European Conference on Computer Vision / Springer. 2016. С. 163–177.
- [136] A comparison of TV-L1 optical flow solvers on GPU / Linchao Bao, Hailin Jin, Byungmoon Kim [и др.] // Proc. GPU Technol. Conf.(GTC) Posters. 2014. с. P4254.
- [137] Technical evaluation of Virtual TouchTM tissue quantification and elastography in benign and malignant breast tumors / Quan Jiang, Yuan Zhang, Jian Chen [и др.] // Experimental and therapeutic medicine. 2014. Т. 8, № 4. С. 1059–1064.
- [138] Evaluation of Different Methods for Using Colour Information in Global Stereo Matching / Michael Bleyer, Sylvie Chambon, Uta Poppe [и др.] // The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Т. 37. с. В3а.

- [139] Tissue-mimicking agar/gelatin materials for use in heterogeneous elastography phantoms / Ernest L Madsen, Maritza A Hobson, Hairong Shi [и др.] // *Physics in medicine and biology*. 2005. Т. 50, № 23. с. 5597.
- [140] A single element 3D ultrasound tomography system / Xiang Zhang, Jonathan Fincke, Andrey Kuzmin [и др.] // *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE / IEEE*. 2015. С. 5541–5544.
- [141] Detection of breast cancer with ultrasound tomography: First results with the Computed Ultrasound Risk Evaluation (CURE) prototype / Nebojsa Duric, Peter Littrup, Lou Poulo [и др.] // *Medical physics*. 2007. Т. 34, № 2. С. 773–785.
- [142] Breast Imaging with 3D Ultrasound Computer Tomography: Results of a First In-vivo Study in Comparison to MRI Images / Torsten Hopp, Lukas Šroba, Michael Zapf [и др.] // *Breast Imaging*. Springer, 2014. С. 72–79.
- [143] Waag Robert C, Fedewa Russell J. A ring transducer system for medical ultrasound research // *Ultrasonics, Ferroelectrics, and Frequency Control, IEEE Transactions on*. 2006. Т. 53, № 10. С. 1707–1718.
- [144] Laugier Pascal, Haïat Guillaume. *Bone quantitative ultrasound*. Springer, 2011.
- [145] Van der Heijden Ferdinand. Edge and line feature extraction based on covariance models // *Pattern Analysis and Machine Intelligence, IEEE Transactions on*. 1995. Т. 17, № 1. С. 16–33.