На правах рукописи

Кузьмин Андрей Игоревич

МЕТОДЫ ОБУЧАЕМОЙ РЕГУЛЯРИЗАЦИИ В ЗАДАЧАХ ПЛОТНОГО СОПОСТАВЛЕНИЯ ИЗОБРАЖЕНИЙ

Специальность 05.13.17 —
«Теоретические основы информатики»

Автореферат

диссертации на соискание учёной степени
кандидата технических наук

Москва — 2018

Работа выполнена в АНОО ВО «Сколковский институт науки и технологий».

Научный руководитель: **Лемпицкий Виктор Сергеевич**
кандидат физико-математических наук, доцент центра по научным и инженерным вычислениям для задач с большими массивами данных, АНОО ВО «Сколковский институт науки и технологий»

Официальные оппоненты: **Крылов Андрей Серджевич**
доктор физико-математических наук, профессор, заведующий лабораторией математических методов обработки изображений, Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский государственный университет имени М. В. Ломоносова»

Николаев Дмитрий Петрович
кандидат физико-математических наук, заместитель директора по научной работе, Федеральное государственное бюджетное учреждение науки Институт проблем передачи информации им. Харкевича Российской академии наук

Ведущая организация: Национальный исследовательский университет «Высшая школа экономики»

Защита состоится «25» октября 2018 года в 16:00 на заседании диссертационного совета Д002.073.05 на базе Федерального государственного учреждения «Федеральный исследовательский центр «Информатика и управление» Российской академии наук» по адресу: 119333, г. Москва, ул. Вавилова, д. 40, концеренц-зал.

С диссертацией можно ознакомиться в библиотеке и на сайте ФИЦ ИУ РАН http://www.frccsc.ru/.

Автореферат разослан «___» _____ 2018 года.

Ученый секретарь
диссертационного совета
Д002.073.05, д.ф.-м.н., профессор Рязанов В.В.

Общая характеристика работы

Актуальность темы. Задача сопоставления изображений является одной из наиболее важных задач в компьютерном зрении, которая возникает во многих практических приложениях, таких как бинокулярная стереореконструкция (Szelinski: 2010), детекция движения на видеопоследовательностях (Zach: 2007, Dosovitkiy: 2015, Ilg: 2017) и анализ медицинских ультразвуковых снимков (Ophir: 1991, Fleming: 2012, Rivaz: 2014).

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, афинные преобразования, которые позволяют сопоставить изображения с учетом перспективных искажений (Szelinski: 2010). 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 (Кузьмин: 2018).

Другим важным аспектом постановки задачи является способ сопоставления изображений. 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 (Zbontar: 2016, Luo: 2016, Dosovitskiy: 2015). При этом существенным недостатком большинства таких методов являтся высокая вычислительная сложность, что не позволяет применять их в задачах, требующих сопоставления в реальном времени (Luo: 2016, Xu: 2017), 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 (Kuzmin: 2017, Кузьмин: 2018).

Большинство современных методов сопоставления изображений можно разделить на две категории. К первой категории относятся методы, основанные на глубоком машинном обучении (Zbontar: 2016, Luo: 2016, Dosovitskiy: 2015, Kuzmin: 2017). 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 (Rivaz: 2014, Kuzmin: 2015). 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 стереореконструкция, которая основана на оценке смещений для левого и правого изображений со стерео-камеры, возникающих засчет бинокулярного эффекта (Scharstein: 2002, Szelinski: 2010). Вторым является сопоставление изображений движущихся объектов на видео-последовательности, известная как задача вычисления оптического потока (Horn: 1981, Szelinski: 2010, Dosovitskiy: 2015). Третьим является задача ультразвуковой эластографии (Ophir: 1991, Fleming: 2012, Rivaz: 2014). Она соответствует сопоставлению медицинских ультразвуковых сников 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 машиного 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 бинокулярная стерео-реконструкция, оптическии поток и ультразвуковая эластография. Проведен количественный анализ качества сопоставления.
3. Предложная эффективная параллелизация разработанных методов. Построен комплекс эффективных параллельных программ с

использованием графических ускорителей, демонстрирующих применимость предложенных моделей в режиме реального времени.

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 засчет 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 достигатся засчет обобщения функционала полной вариации. Предложенный функ-

ционал 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 открытих 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”, 2016.
5. Международная конференция “IEEE 37th Annual International Conference on Medicine and Biology Society”, 2015.

Личный вклад. Все результаты получены автором лично.

Публикации. По тематике исследования опубликовано 5 научных работ, в том числе 5 статей в изданиях, рекомендованных ВАК.

Объем и структура работы. Диссертация состоит из введения, пяти глав и заключения и приложения. Полный объем диссертации **133** страницы текста с **47** рисунками и 7 таблицами. Список литературы содержит **145** наименований.

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.

Первая глава посвящена постановке задачи плотного сопоставления изображений и описанию приложений, рассмотренных автором. Глава начинается с разбора различных постановок задачи сопоставления изображений, а также предположений, необходимых для решения задачи на практике. В главе также рассмотрена задача плотного сопоставления изображений в трех приложениях: бинокулярное стерео-сопоставление, оптический поток и ультразвуковая эластография. Главу завершает набор примеров практического использования предложенных в работе методов.

Вторая глава посвящена задаче плотного сопоставления изображений в приложении к бинокулярной стерео-реконструкции (Kuzmin: 2017). Приводится описание метода, предложенного автором, обзор предложенных в литературе методов, и численные эксперименты по сравнению различных методов.

В настоящее время, из всех предложенных в литературе методов, наиболее низкой ошибкой стерео-сопоставления обладают методы глубокого машинного обучения на основе сверточных нейросетей (Zbontar: 2015, Luo: 2016). При этом большинство таких методов основано на обучении дескрипторов большой размерности, которые затем используются для сопоставле-

ния фрагментов изображений. Несмотря на высокое качество реконструкции, сравнение большого количества дескрипторов большой размерности имеет высокую вычислительную сложность (Zbontar: 2015, Luo: 2016). Целью данной работы является разработка альтернативных моделей, позволяющих понизить вычислительную сложность методов машинного обучения на этапе исполнения.

Предложенный метод в явном виде хранит в памяти трехмерный тензор энергий стерео-сопоставления в виде трехмерного массива размера (h,w,d_{max}), где h and w - размеры изображения, а d_{max} - максимальный разрешенный диспаритет. Вычисление тензора энергий производится в соответствии с локальными методами стерео-сопоставления. При этом, энергия равна сумме двух членов:

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

где коэффициент $\alpha \in [0,1]$. Первый член есть абсолютное значение разностей интенсивностей соответствующих пикселов:

$$E_{SAD}(x,y,d) = |I^L(x,y) - I^R(x-d,y)|. \tag{2}$$

Второй член $E_{census}(x,y,d)$ основан на сопоставлении локальных дескрипторов, соответствующих следующему цензус-преобразованию. Для черно-белого изображения I, определим функцию ξ, которая будет будет принимать значение 0 или 1 в зависимости от результата сравнения интенсивностей в пикселах $\boldsymbol{p}$ и $\boldsymbol{q}$:

$$\xi(\boldsymbol{p},\boldsymbol{q}) = \begin{cases} 1, \text{ если } I(\boldsymbol{q}) < I(\boldsymbol{p}), \\ \text{иначе } 0. \end{cases} \tag{3}$$

Используя такую функцию, определим цензус-преобразование (Zabih: 1994), которое ставит в соответствии каждому пикселу изображения следующий многомерный вектор из нулей и единиц:

$$R_\tau(\boldsymbol{p}) = \bigotimes_{[i,j]\in D_w} \xi(\boldsymbol{p},\boldsymbol{p}+[i,j]), \tag{4}$$

где $\otimes$ - операция конкатенации, а D_w - набор возможных двумерных смещений внутри квадратного окна размера $n \times n$ с центром в пикселе $\boldsymbol{p}$. Полученные дескрипторы, заданные битовыми последовательностями, сравниваются, используя расстояние Хэмминга.

В качестве архитектуры для сверточной нейросети, была использована многомасштабная модель для детекции границ (Xie: 2015). Предложенный автором метод агрегирования энергий сопоставления основан на рекурсивном фильтре, учитывающем границы, и его обучаемой версии. 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 $i = 2,...,N$ имеем:

$$y_i = (1 - \omega_i)x_i + \omega_i y_{i-1} \tag{5}$$

Варьирование весов ω_i используется для контроля степени сглаживания, что дает возможность сохранять пространственные границы изображения: в самом деле фрагменты изображения с величиной ω_i, близкой к единице, усредняются между соседними пикселами y_{i-1} и y_i, тогда как величина w_i, близкая к нулю (например, в пикселах, соответствующих границам изображения), ведет к отсутствию пространственного сглаживания, т.е. $y_i = x_i$.

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. При этом 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:

$$I_{filt} = \mathrm{F}(I,W_h,W_v). \tag{6}$$

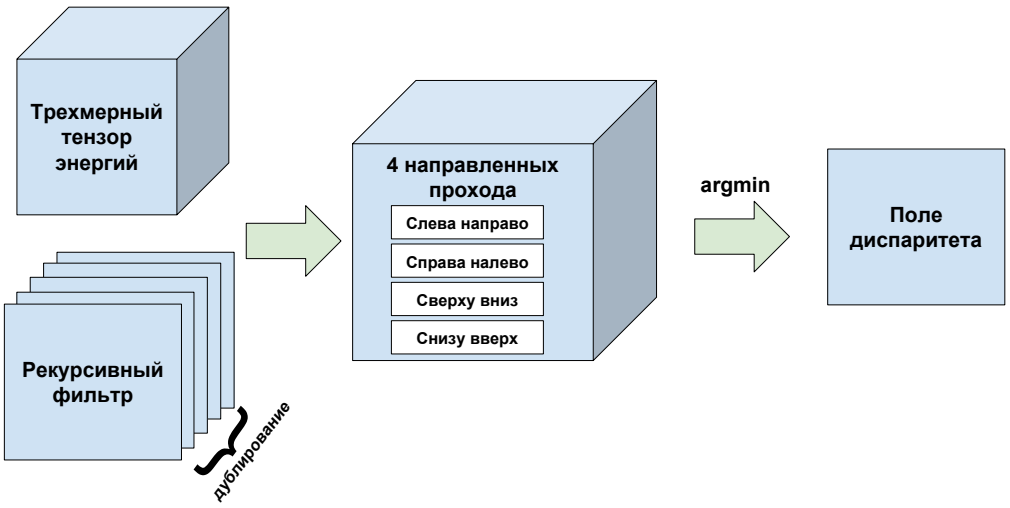


Рис. 1 — Схематическое изображение метода фильтрации тензора энергий.

Алгоритм вычисления описанных выше четырех рекуррентных проходов задается соотношениями:

$$I^L(x,y,d) = (1 - W_h(x,y))\, I(x,y) + W_h(x,y)\, I(x-1,y), \tag{7}$$
$$I^R(x,y,d) = (1 - W_h(x,y))\, I^L(x,y) + W_h(x,y)\, I^L(x+1,y), \tag{8}$$
$$I^T(x,y,d) = (1 - W_v(x,y))\, I^R(x,y) + W_v(x,y)\, I^R(x,y-1), \tag{9}$$
$$I^B(x,y,d) = (1 - W_v(x,y))\, I^T(x,y) + W_v(x,y)\, I^T(x,y+1), \tag{10}$$

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 направленых проходов вычисления осуществляются независимо для строк или столбцов изображения.

Для того, чтобы построить обучаемую модель фильтрации, веса предсказываются на основе входного изображения с помощью сверточной нейросети. Схема алгоритма агрегирования энергий представлена на рис. 1. 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 $E_d = E(x,y,d)$, $x = 0,..,w$, $y = 0,..,h$, применяется дублирование весов по третьему измерению:

$$E_d^{filt} = \mathrm{F}(E_d,W_h,W_v),$$
$$d = \{0,1,...,d_{max}\}. \tag{11}$$

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 $\sigma : \mathbf{R}^K \to [0,1]^K$. Такое отображение для некоторого одномерного вектора энергий E выражается следующим образом:

$$\sigma_j(E) = \frac{e^{E_j}}{\sum_{i=1}^{K} e^{E_i}}. \tag{12}$$

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

$$C_{sm}(x_0,y_0,d) = \sigma\big(C(x_0,y_0,d)\big). \tag{13}$$

Для того, чтобы смоделировать функцию, которая имеет максимум в эталонном значении диспаритета, используется дельта функция:

$$C_{gt}(x,y,d) = \begin{cases} 1, \text{ если } d = D_{gt}(x,y), \\ \text{иначе } 0. \end{cases} \tag{14}$$

В качестве функции потерь используется значение кросс-энтропии:

$$L(C,C_{gt}) = -\sum_{x,y}\sum_{d} C_{sm}(x,y,d)\ \log\ C_{gt}(x,y,d). \tag{15}$$

В процессе обучения веса модели изменяются таким образом, чтобы обнаружить максимальное значение энергии в компоненте, соответствующей эталонному диспаритету. Так, в результате получается обучаемая модель на основе сверточно-рекуррентной нейросети, не требующая постобработки (см. рис. 2).

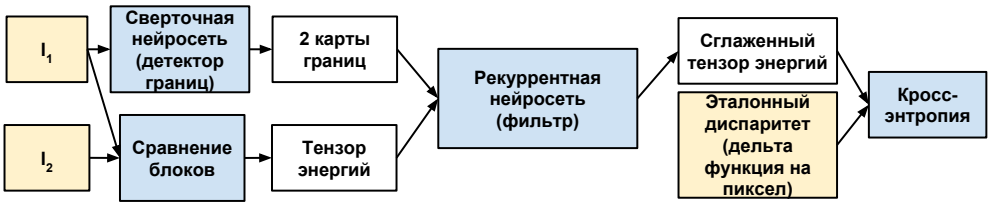


Рис. 2 — Схема предложенной сверточно-рекуррентной нейросети.

Наиболее трудоемким этапом алгоритма является агрегирование тензора энергий, количество операций оценивается как $\mathcal{O}(nd_{max})$, где n - количество пикселов в изображении, а d_{max} - максимальный диспаритет. При этом трудоемкость наиболее быстрого метода, основанного на машинном обучении, предложенного в литературе оценивается как $\mathcal{O}(nd_{max}k)$, где k - размерность глубоких дескрипторов.

Продолжение третьей главы содержит результаты численных экспериментов на реальных данных и сравнение результатов с предложенными в литературе методами. В главе также приводится анализ времени исполнения на графическом ускорителе по отдельным этапам работы алгоритма.

Третья глава посвящена задаче плотного сопоставления изображений в приложении к вычислению оптического потока для кадров видеопоследовательности (Кузьмин: 2018). Приводится описание метода, предложенного автором, обзор предложенных в литературе методов и результаты численных экспериментов.

В задаче вычисления оптического потока требуется вычислить двумерное поле смещений для пары изображений $I_{0,1}$, определенных на области Ω (Horn: 1985). Оптический поток $\boldsymbol{u} = (u_x,u_y)$ 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 пикселов, $I_0(x,y)$ и $I_1(x + u_x(x,y),y + u_y(x,y))$ соответствуют одним и тем же точкам сцены. В отличие от задачи стереосопоставления, которая сводится к нахождению одномерного поля смещений, движение объектов на сцене может быть произвольным, поэтому поле смещением в задаче оптического потока является двумерным.

Серия современных подходов для задачи вычисления оптического потока основана на применении методов машинного обучения с использованием сверточных нейросетей (Dosovitskiy: 2015, Xu: 2017, Ilg: 2017). При этом нейросети могут быть использованы как для попиксельного предсказания оптического потока для входных изображений, так и для обучения дескрипторов для сопоставления изображений. В то время как такие методы позволяют получить оценки достаточно высокого качества, они обладают высокой вычислительной сложностью. В связи с этим, представляется актуальной задача разработки методов машинного обучения, имеющих низкую вычислительную сложность на этапе исполнения.

Развитие вариационных методов для вычисления оптического потока связано с разработкой эффективных методов оптимизации. При этом набор функций энергии, используемых в задаче, ограничен возможностью представить задачу в виде минимизации одного или последовательности выпуклых функционалов (Wedel: 2009, Werlberger: 2010, Ranftl: 2014). Метод, основанный на применении сверточной нейросети, предложенный в данной работе (Кузьмин: 2018), позволяет обучать оператор регуляризации с использованием тренировочной выборки пар изображений, для которых предоставлены эталонные значения оптического потока. Идея построения обучаемой модели состоит в представлении итераций оптимизационного алгоритма в виде слоев сверточной нейросети.

В соотвествии с алгоритмом оптимизации, используемым в методе оценки оптического потока в реальном времени, определим архитектуру нейросети, которая будет состоять из модуля деформации и модуля итерации двойственного метода. 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 средняя квадратичная ошибка для двумерных векторов смещений. Результат, полученный нейросетью, не требует дополнительной постобработки.

Оптический поток вычисляется с использованием многомасштабного подхода (рис. 3). Начиная с наиболее крупного масштаба, для которого ней-

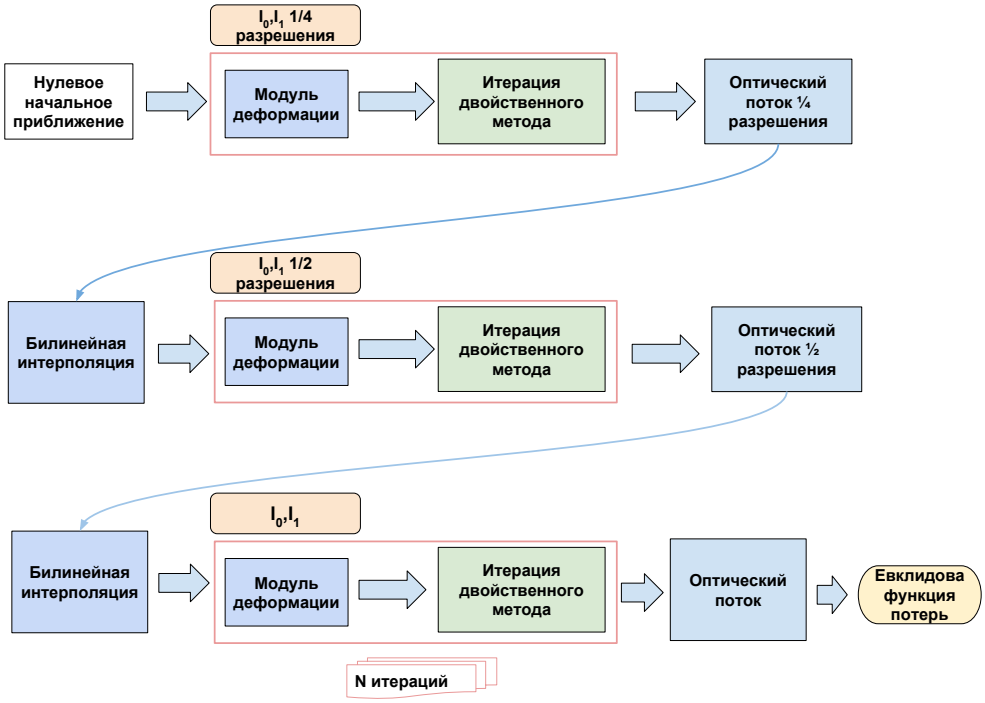


Рис. 3 — Архитектура предложенной нейросети для задачи оптического потока.

росеть 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.

Модуль деформации. Модуль деформации использует операцию порогового преобразования как слой нейросети. Модуль принимает на вход пару изображений $I_{0,1}$ на соответствующем масштабе и начальную оценку поля смещений $\boldsymbol{u}_0$. Для вычисления градиента используется процедура деформации изображения и центральная разностная схема. Для того, чтобы обучать нейросеть методом обратного распространения ошибки, для вычисления деформированного изображения используется слой пространственной деформации (spatial transformer layer), предложенный в литературе (Jaderberg: 2015):

$$\frac{\partial I_1}{\partial x} = \frac{\mathtt{S}(I_1, u_{0x} + \Delta x, u_{0y}) - \mathtt{S}(I_1, u_{0x} - \Delta x, u_{0y})}{2\Delta x}, \tag{16}$$

$$\frac{\partial I_1}{\partial y} = \frac{\mathtt{S}(I_1, u_{0x}, u_{0y} + \Delta y) - \mathtt{S}(I_1, u_{0x}, u_{0y} - \Delta y)}{2\Delta y}. \tag{17}$$

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 стадартным арифметическим операциям и пороговому преобразованию.

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

$$\nabla u_d = \left[\frac{\partial u_d}{\partial x}, \frac{\partial u_d}{\partial y}\right] = [u_d * f_x, u_d * f_y], \tag{18}$$

$$\text{div } \boldsymbol{p} = \frac{\partial p_1}{\partial x} + \frac{\partial p_2}{\partial y} = p_1 * \tilde{f}_x + p_2 * \tilde{f}_y. \tag{19}$$

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

$$Lu_d = = \begin{bmatrix} u_d * f_1 \\ u_d * f_2 \\ \vdots \\ u_d * f_K \end{bmatrix}. \tag{20}$$

При этом сопряженный оператор $L^* : R^{N\times K} \to R^N$, определенный для двойственной переменной $\boldsymbol{p}$, на выходе дает изображение, которое имеет K

каналов $\boldsymbol{p} = [p_1, p_2, \ldots, p_K]$. Такой оператор может быть записан в следующем виде:

$$L^* \boldsymbol{p} = \sum_{i=1}^{K} p_i * \tilde{f}_i. \tag{21}$$

В отличие от оптимизационного метода с фиксированным оператором регуляризации, предложенный метод основан на обучаемой регуляризации, что позволяет уменьшить ошибку сопоставления.

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

Продолжение третьей гловы содержит серию численных экспериментов по оценке предложенного метода на публичной коллекции изображений Sintel (Butler: 2012).

Разработанный метод имеет низкую вычислительную сложность и может быть использован для расчетов в реальном времени с использованием графического ускорителя. Количество арифметических операцией оценивается как $\mathcal{O}(NDMF^2)$, где N - количество пикселей изображения, D - количество слоев нейросети, M - количество фильтров, а F - размер фильтра. Малое потребление памяти (например, 40 МБ для изображений рассмотренной коллекции при использовании перезаписываемых операций) позволяет потенциально использовать предложеный метод на мобильных платформах.

В **четвертой главе** приведено описание разработанного метода сопоставления ультразвуковых изображений на основе многих кадров (Kuzmin: 2015).

Глава описывает метод, предложенный автором, а также содержит результаты сравнения с прочими методами на основе реальных данных. Основное преимущество предложенного подхода состоит в возможности получать более качественные оценки механической деформации засчет использования трех кадров с известным соотношением силы нажатия (большинство предложенных в литературе методов использует два кадра). При этом сила нажа-

тия определяется аппаратно, используя специальную ультразвуковую пробу со встроенным датчиком силы.

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

Для каждого кадра, вектор смещения полагается целочисленным и конечным внутри преписанного декартового произведения двух интервалов. Окно поиска Λ может быть описано следующим образом:

$$\Lambda = \{0, \ldots, +D_a\} \times \{-D_l, \ldots, +D_l\}\,.$$

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 соотвествующих фрагментов по прямоугольному окну методом полного перебора. Обозначим результат такой процедуры следующим образом:

$$\boldsymbol{d}_{coarse} = \mathtt{B}(I_0,I_1)\,,$$

Где $\mathtt{B}$ - алгоритм поиска смещений описанный выше, который принимает на выход пару кадров и вычисляет двумерное поле смещений. Обозначим за $\boldsymbol{d}_{coarse}$ двумерный вектор смещений, вычисленный для пары изображений I_0 и 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 $\boldsymbol{d} = (d_x,d_y)$:

$$I_d = W(I,\boldsymbol{d}), \tag{22}$$

Тогда интенсивность деформированного выражения выражается следующим образом:

$$I_d(x,y) = I(x - d_x,y - d_y). \tag{23}$$

Для того, чтобы уточнить первоначальное приближение $\frac{f_2}{f_1}\,\boldsymbol{d}_{coarse}$, 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 $\mathtt{W}(I_2, \frac{f_2}{f_1}\,\boldsymbol{d}_{coarse})$. При этом интервал поиска смещения на данном этапе может быть существенно сужен:

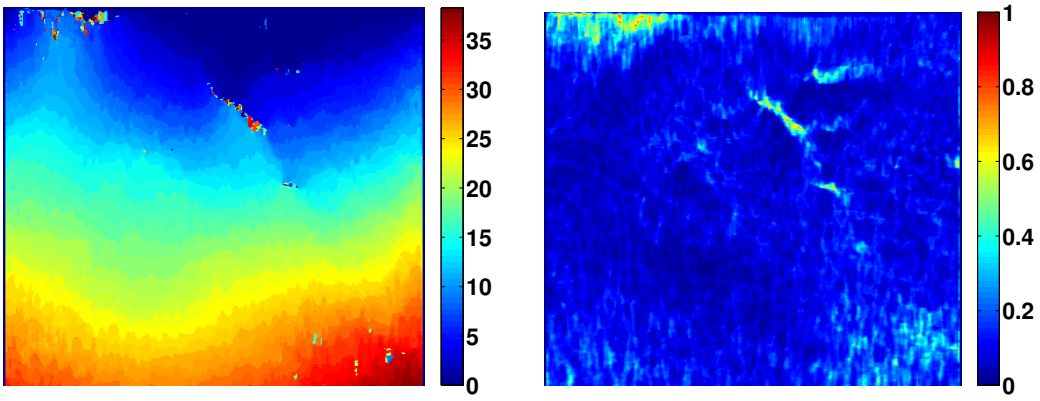
$$\boldsymbol{d}_{local} = \mathtt{B}\left(I_0, \mathtt{W}(I_2, \frac{f_2}{f_1}\,\boldsymbol{d}_{coarse})\right). \tag{24}$$

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

$$\boldsymbol{d}_{total} = \frac{f_2}{f_1}\,\boldsymbol{d}_{coarse} + \boldsymbol{d}_{local}. \tag{25}$$

Полное смещение подается на выход процедуры реконструкции механической деформации.

В качестве метрики для сопоставления фрагментов изображения используется сумма квадратов разностей интенсивностей по прямоугольному окну. Помимо поля смещений, процедура выдает на выходе меру качества



Поле верт.смещений Относительная ошибка сопоставления

Рис. 4 — Пример поля смещений для ткани печени человека.

полученного сопоставления, соответствующую относительной ошибке в той же метрике.

При этом оценка механической деформации формулируется в качестве обратной задачи с использованием функционала следующего вида:

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

Где s_a - регуляризованная оценка продольной деформации, $\rho(s_a)$ - функционал регуляризации и λ - параметр, отвечающий за сглаживание:

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

В соответствии с проведенными автором численными экспериментами, оценка поля смещений часто содержит участки со значительной ошибкой, часто встречаемые на границе различных тканей (рис. 4 слева). 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 сопоставления в L^2 норме (рис. 4 справа). Такая оценка позволяет использовать адаптивную регуляризацию: в регионах сопоставления с низкой ошибкой полагаться на входные изображения, а в регионах с высокой ошибкой - полагаться на пространственное сглаживание. Для того, чтобы включить информацию о качестве сопоставления в регуляризационный функционал, вводится диаго-

нальная матрица весов $D = diag(w)$, которая зависит от описанной выше меры качества сопоставления:

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

В алгоритме используется бинарная маска весов w для каждого пиксела, основанная на пороговом отсечении относительной ошибки сопоставления q_{ij}:

$$w_{ij} = \begin{cases} 1 & q(x,y) < t \\ 0 & q(x,y) \geq t. \end{cases}$$

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

В соответствии с проведенными численными экспериментами, использование L^2 нормы в качестве регуляризации ведет к пересглаженным решениям с потерей визуальной информации на границах различных типов тканей. Для избежания подобного эффекта, представляется целесообразным использование полной вариации в качестве регуляризационного функционала. При этом исходный функционал может быть записан в виде:

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

где полная вариация выражается с использованием пространственных производных продольного поля деформации:

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

Использование регуляризации на основе полной вариации ведет к сохранению границ на перепадах между мягкими и жесткими тканями. При этом была использована эффективная реализация двойственного градиентного метода для эффективной минимизации. Алгоритм параллелизован на уровне различных пикселов изображения, что позволило естественным образом получить реализацию на графическом ускорителе.

Далее в главе приводятся количественные и качественные результаты применения предолженного метода на следующих экспериментах: моде-

лирование сжатия синтетических фантомов с известной геометрией, снимки специально изготовленных фантомов с приближенно известной геометрией и снимки реальных тканей человека.

Автором были произведены численные эксперименты по сравнению пяти методов, из которых четыре являются разновидностями предложенного подхода. Эти методы соотвествуют использованию двух видов регуляризации (L^2 норма или полная вариация) и разному количеству используемых кадров (пара или тройка кадров). При использовании пары изображений, результат сопоставления пары кадров напрямую используется для оценки деформации.

В **заключении** приведены основные результаты диссертационной работы, которые заключаются в следующем:

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 стереореконструкция, вычисление оптический поток и ультразвуковая эластография. Проведен количественный анализ качества сопоставления.
3. Предложная эффективная параллелизация разработанных методов. Построен комплекс эффективных параллельных программ с использованием графических ускорителей, демонстрирующих применимость предложенных моделей в режиме реального времени.

Публикации автора по теме диссертации

Публикации в рецензируемых изданиях из перечня 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.