

На правах рукописи

Боков ✓

Боков Александр Александрович

**ИССЛЕДОВАНИЕ И РАЗРАБОТКА АЛГОРИТМОВ
ВОССТАНОВЛЕНИЯ ФОНА В ЗАДАЧЕ ПРЕОБРАЗОВАНИЯ
ВИДЕО В СТЕРЕОСКОПИЧЕСКИЙ ФОРМАТ**

Специальность 05.13.11 —
«Математическое и программное обеспечение вычислительных машин,
комплексов и компьютерных сетей»

Автореферат
диссертации на соискание ученой степени
кандидата физико-математических наук

Челябинск — 2019

Работа выполнена на кафедре интеллектуальных информационных технологий факультета вычислительной математики и кибернетики ФГБОУ ВО «Московский государственный университет имени М.В. Ломоносова».

Научный руководитель: **Ватолин Дмитрий Сергеевич**
кандидат физико-математических наук, старший научный сотрудник ФГБОУ ВО «Московский государственный университет имени М.В. Ломоносова»

Официальные оппоненты: **Турлапов Вадим Евгеньевич**,
доктор технических наук, доцент,
ФГАОУ ВО «Национальный исследовательский Нижегородский государственный университет им. Н.И. Лобачевского», профессор

Игнатенко Алексей Викторович,
кандидат физико-математических наук,
ООО «Ормко» (Москва), старший менеджер по разработке программного обеспечения

Ведущая организация: ФГУП «Государственный научно-исследовательский институт авиационных систем» (Москва)

Защита состоится 11 декабря 2019 г. в 12:00 часов на заседании диссертационного совета Д 212.298.18 при ФГАОУ ВО «Южно-Уральский государственный университет (национальный исследовательский университет)» по адресу: 454080, г. Челябинск, пр. Ленина, 76, ауд. 1001.

С диссертацией можно ознакомиться в библиотеке Южно-Уральского государственного университета и на сайте: <https://www.susu.ru/ru/dissertation/d-21229818/bokov-aleksandr-aleksandrovich>.

Автореферат разослан «___» _____ 2019 г.

Ученый секретарь
диссертационного совета



М. Л. Цымблер

Общая характеристика работы

Актуальность темы. Несмотря на то, что история стереокинематографа насчитывает уже более 90 лет, современный его расцвет начался с выхода кинофильма «Аватар» в 2009 году и обусловлен существенным техническим прогрессом как в области производства стереофильмов, так и в области технологий показа стереоскопического видео. Характерной чертой современного этапа развития данной области является преобладание конвертации в стереоскопический формат как основного способа производства стереофильмов. Например, в 2018 году среди 36 стереофильмов вышедших на экраны кинотеатров 27 были созданы именно путем конвертации в стереоскопический формат. В отличие от стереосъемки, предполагающей наличие специальной установки из камер, позволяющей снять сцену одновременно с двух различных ракурсов, в ходе конвертации новые ракурсы строятся искусственно из обычного моноскопического видео. Конвертация в стереоскопический формат обладает рядом преимуществ по сравнению со стереосъемкой, начиная с повышенного контроля над воспринимаемой зрителем степенью объемности каждого из объектов сцены по отдельности для достижения того или иного художественного эффекта и заканчивая возможностью конвертировать в стереоскопический формат старые популярные фильмы, которые изначально были сняты в моноскопическом формате.

Однако, на сегодняшний день конвертация фильмов в стереоскопический формат остается крайне трудоемким процессом, требующим большой объем ручной работы специально обученных художников. В связи с этим учеными активно исследуются алгоритмы и интерактивные программные системы, которые позволяют упростить процесс конвертации посредством уменьшения объема ручного труда, требуемого для достижения удовлетворительного результата. Задача преобразования видео в стереоскопический формат включает в себя следующие основные подзадачи:

- Построение карты глубины для всех кадров входного видео, которая описывает расстояние от каждой точки сцены до плоскости линзы камеры;
- Заполнение областей открытия, которые содержат фрагменты фона, скрытые за объектом переднего плана в исходном кадре видео, но становящиеся видимыми при изменении ракурса просмотра;
- Обработка полупрозрачных областей и границ объектов сцены.

Данная диссертационная работа посвящена решению подзадачи заполнения областей открытия путем восстановления изображения фона, расположенного за объектами переднего плана. Помимо преобразования видео в стереоскопический формат задача восстановления фона в видео также имеет ряд других приложений, таких как бесследное удаление нежелательных объектов и устранение дефектов пленки в ходе реставрации архивных видеозаписей. Однако стоит отметить, что в данной области до сих пор отсутствует общепринятая методика оценки качества в связи с тем, что традиционные методы оценки качества видео плохо отражают особенности визуального восприятия восстановленного фона. Также, современные алгоритмы восстановления фона в видео формулируют задачу в виде единой задачи оптимизации сразу по всем кадрам, что приводит к чрезмерно высокой вычислительной сложности, затрудняющей применение данных алгоритмов на практике. Таким образом, актуальной является задача выработки специализированной методики оценки качества восстановленного фона, а также задача разработки нового алгоритма восстановления фона с более низкой вычислительной сложностью в сравнении с аналогами.

Степень разработанности темы. Задача восстановления фона активно исследуется учеными с начала 2000-х годов. Наиболее ранние подходы к ее решению на основе моделирования средствами дифференциальных уравнений исследуются в работах Марсело Бертальмио (M. Bertalmio). Методы на основе копирования заплаток, позволяющие восстанавливать более крупные области, были развиты в работах Антонио Криминиси (A. Criminisi). Современные методы базируются на явном моделировании движения фона и формулировке единой задачи оптимизации на все кадры видео, что позволяет достичь высокого качества результата. Данный подход был существенно развит в работе Джиа-Бин Хуанга (J.-B. Huang).

Целью данной работы является исследование и разработка нового решения проблемы заполнения областей открытия на базе восстановления фона за объектами переднего плана, которое бы позволило увеличить степень автоматизации процесса преобразования видео в стереоскопический формат.

Для достижения поставленной цели требуется решить следующие **задачи**:

1. Выработать методику оценки качества восстановленного фона и систематизировать известные алгоритмы его построения.

2. Разработать новый алгоритм восстановления фона в видео, обладающий более низкой вычислительной сложностью по сравнению с аналогами.
3. Создать алгоритм построения новых ракурсов на основе разработанного алгоритма восстановления фона в видео.

Основные положения, выносимые на защиту:

1. Предложен новый подход к решению проблемы заполнения областей открытия на основе проведения многослойного восстановления фона в видео с использованием непараметрической модели движения для каждого слоя.
2. Разработан новый алгоритм восстановления фона в видео с пространственной сложностью, не зависящей от количества кадров во входной видеопоследовательности. По результатам экспериментальной оценки предложенный алгоритм превзошел существующие аналоги по точности найденного решения и не уступил аналогам по результатам экспертной оценки.
3. Предложено два новых функционала качества работы алгоритмов восстановления фона в видео, основанных на критерии локального сходства с эталоном, которые обладают более высокой корреляцией с экспертной оценкой по сравнению с ранее использовавшимися для этой цели функционалами.
4. Все предложенные алгоритмы и функционалы качества были реализованы, а также проведены вычислительные эксперименты, подтверждающие их эффективность. Алгоритм восстановления фона был протестирован на реальных данных, предоставленных коммерческой компанией.

Научная новизна:

1. Впервые предложен алгоритм многослойного восстановления фона в видео, который позволяет моделировать изображение фона в областях перекрытия трех и более объектов.
2. Предложен новый алгоритм восстановления фона в видео, основанный на покадровом уточнении результата восстановления, который обладает более низкой вычислительной сложностью в сравнении с известными аналогами.
3. Предложено два новых функционала качества восстановленного фона, измеряющих локальное сходство с эталоном и постоянство локально-

го сходства с эталоном с течением времени, которые лучше отражают визуальное качество фона, чем ранее известные функционалы.

Теоретическая значимость. В работе задача восстановления фона в видео сформулирована в виде задачи покадровой оптимизации. Показано, что она сводится к решению набора СЛАУ с разреженной симметричной положительно определенной матрицей.

Практическая значимость. Предложенный алгоритм восстановления фона в видео был реализован в виде программного инструмента, позволяющего автоматически восстанавливать фон в произвольных областях видеопоследовательности с более высоким качеством по сравнению с доступными на момент проведения исследования коммерческими аналогами. Данный инструмент может быть применен в киностудиях для удешевления процесса конвертации моноскопического видео в стереоскопический формат за счет автоматизации восстановления фона за объектами переднего плана. Разработанный программный инструмент был апробирован на данных из реальных кинофильмов, которые были предоставлены коммерческой компанией YUVsoft. Предложенные функционалы качества были включены в разработанную методику оценки качества восстановленного фона. Данная методика вместе с подготовленной базой видеопоследовательностей с эталонными изображениями фона легли в основу открытого для новых участников сравнения алгоритмов восстановления фона в видео, доступного в сети Интернет по адресу <http://videocompletion.org>.

Методология и методы исследования. В диссертационной работе применялись методы линейной алгебры, теории алгоритмов, а также методы численного решения разреженных систем линейных алгебраических уравнений.

Достоверность полученных результатов подтверждается публикациями в рецензируемых научных изданиях, а также апробацией в рамках российских и международных конференций.

Апробация работы. Основные результаты работы докладывались на:

- 8-ой международной конференции по 3D визуализации (IC3D 2018), 5 декабря 2018 г., Брюссель, Бельгия;
- 25-ой международной конференции по обработке изображений (ICIP 2018), 7-10 октября 2018 г., Афины, Греция;
- Международной конференции по мультимедийным системам (ICMEW 2017), 10 июля 2017 г., Гонконг, Китай;

- 5-ой международной научно-практической конференции «Инновационные технологии в кинематографе и образовании», 12-13 ноября 2018 г., Москва, Россия;
- Всероссийской научной конференции «Ломоносовские чтения-2018», 16-27 апреля 2018 г., Москва, Россия;
- Всероссийской научной конференции «Ломоносовские чтения-2017», 17-26 апреля 2017 г., Москва, Россия.

Личный вклад автора. В работах [1—6] научному руководителю Д.С. Ватолину принадлежит постановка задачи, А.А. Бокову – все полученные результаты. В статье [3] М.В. Ерофееву и Ю.А. Гитману принадлежит основа разработанной Интернет-платформы <http://videocompletion.org> для открытого сравнения алгоритмов восстановления фона (стр. 607).

Публикации. Основные результаты по теме диссертации изложены в 6 печатных изданиях, 2 из которых изданы в журналах, рекомендованных ВАК [1; 2], 4 входят в международную базу цитирования Scopus [3—6].

Объем и структура работы. Диссертация состоит из введения, трех глав и заключения. Полный объем диссертации составляет 110 страниц, включая 14 рисунков и 7 таблиц. Список литературы содержит 73 наименования.

Содержание работы

Во **введении** обоснована актуальность диссертационной работы, сформулирована цель и аргументирована научная новизна исследований, показана их практическая значимость, представлены выносимые на защиту научные положения.

Первая глава посвящена решению задачи восстановления фона в видео. Задача восстановления фона в видео является подзадачей рассматриваемой в данной работе задачи преобразования видео в стереоскопический формат, так как в ходе такого преобразования возникают так называемые области открытия, где отсутствует информация об изображении фона, и ее требуется тем или иным способом восстановить (подробнее данный процесс описан в разделе 3.1). Формально задачу восстановления фона можно поставить следующим образом. Обозначим область определения видеопоследовательности как $I = [1, W] \times [1, H] \times \{1, \dots, T\}$, где W , H – ширина и высота кадра в пикселах, соответственно, T – количество кадров. Тогда входными данными алгоритма являются значения

цвета вне требующей восстановления области $\Omega \subset I$. То есть, по значениям цвета $C : I \setminus \Omega \mapsto \mathbb{R}^3$ вне области неизвестного фона требуется определить цвет внутри рассматриваемой области $C(\Omega)$. В контексте преобразования видео в стереоскопический формат, как правило, в качестве такой области Ω выступает область, занимаемая определенным объектом переднего плана. Тогда $C(\Omega)$ описывает фон, находящийся за таким объектом. Далее в описании предложенного алгоритма будем предполагать, что областью значений C является лишь \mathbb{R} , то есть значения яркости пикселей в видео. Построенный алгоритм тривиально обобщается на случай использования полного цветового пространства RGB, как описано в разделе 1.3.5. Очевидно, для корректной постановки задачи требуется ввести ряд дополнительных предположений о природе входных данных. Стоит отметить, что данные предположения существенно различаются между различными алгоритмами, ранее предложенными в научных работах.

Ключевой особенностью задачи восстановления фона в видео, которая отличает ее от аналогичной задачи для изображений, является возможность найти фрагменты фона, скрытые в текущем кадре, в других кадрах входного видео. Эта особенность существенно упрощает задачу, особенно если для любой скрытой точки фона $p \in \Omega$ можно найти соответствующую точку $q \notin \Omega$, где фон является видимым, однако часто это оказывается слишком сильным предположением. В то же время, работа с видео порождает дополнительные требования к алгоритмам восстановления фона:

- Результат восстановления должен быть устойчив во времени, ввиду высокой чувствительности зрительной системы человека к неестественным различиям между соседними кадрами. Применение алгоритмов восстановления фона в изображениях независимо к каждому кадру будет приводить к неприемлемым результатам именно по этой причине;
- Алгоритм должен иметь адекватную объемам видеоданных временную и пространственную сложность. Как показано в разделе 1.2, многие существующие алгоритмы трудно применять на практике к видео большой продолжительности и высокого разрешения из-за чрезмерно высокой вычислительной сложности.

Раздел 1.2 содержит обзор существующих алгоритмов решения задачи восстановления фона в видео. Раздел 1.3 содержит описание предлагаемого алгоритма. Предложенный алгоритм опирается на использование непараметрической модели движения фона на основе оптического потока. Оптический поток можно

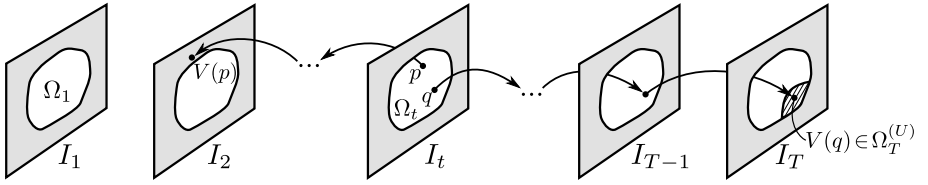


Рис. 1 — Иллюстрация используемого подхода к восстановлению фона. Точка $p \in \Omega^{(K)}$ может быть отображена в область известного фона композицией межкадровых отображений $O^{t \rightarrow t-1}$, что позволяет восстановить фон простым копированием. Точка $q \in \Omega^{(U)}$ не имеет прямого отображения в область известного фона, но цвет в ней может быть восстановлен путем копирования из области $\Omega_T^{(U)} \subset \Omega^{(U)}$ на последнем кадре, к которой был применен алгоритм восстановления фона в изображениях.

определить как отображение, которое описывает межкадровое движение изображения каждой точки в видео. Искомое же отображение из одного кадра видеопоследовательности $I_t \subset I$ в соседний $O^{t \rightarrow t \pm 1} : I_t \mapsto \mathbb{R}^2 \times \{t \pm 1\}$ моделирует движение фона как в области $I \setminus \Omega$, так и внутри области неизвестного фона Ω . Такие отображения $O^{t \rightarrow t+1}$, $O^{t \rightarrow t-1}$ вычисляются для каждой пары соседних кадров с номерами $(t, t+1)$ и $(t, t-1)$, соответственно, с помощью алгоритма, описанного в разделе 1.3.2. Определим множества точек $R^+[p]$, $R^-[p]$, достижимых из точки $p \in \Omega$ посредством применения композиции межкадровых отображений вида $O^{t \rightarrow t+1}$ и $O^{t \rightarrow t-1}$, соответственно, оставаясь в пределах кадра $[1, W] \times [1, H]$. Затем представим множество Ω как объединение двух непересекающихся множеств $\Omega^{(K)}$ и $\Omega^{(U)}$, где $\Omega^{(U)} = \{p \mid \forall q \in R^+[p] \cup R^-[p] : q \in \Omega\}$, то есть множество точек, которые остаются скрытыми на протяжении всей видеопоследовательности, а $\Omega^{(K)} = \Omega \setminus \Omega^{(U)}$. В $\Omega^{(K)}$ цвет фона может быть однозначно восстановлен как:

$$C(p) = C(V(p)), \text{ где } V(p) = \begin{cases} \arg \min_{q \in R^-[p], q \notin \Omega} \|p - q\|_t, & \exists q \in R^-[p] : q \notin \Omega \\ \arg \min_{q \in R^+[p], q \notin \Omega} \|p - q\|_t, & \text{иначе} \end{cases} \quad (1)$$

Здесь $\|p - q\|_t$ обозначает модуль разницы в номере кадра между p и q . То есть, в точку неизвестной области p копируется цвет из ближайшей (по номеру кадра) точки с известным цветом q из множества достижимых точек в предыдущих кадрах $R^-[p]$, если она существует. Иначе цвет копируется из аналогичной точки в будущем кадре. Для области $\Omega^{(U)}$ в данной работе делается предположение, что

$\forall p \in \Omega^{(U)} : \exists(x,y,1) \in R^- [p] \vee \exists(x,y,T) \in R^+ [p]$, то есть достаточно полностью знать фон в первом и последнем кадрах, чтобы восстановить фон везде в $\Omega^{(U)}$. Для вычисления значений цвета в $\Omega^{(U)}$ на первом и последнем кадрах используется известный алгоритм восстановления фона в изображениях [7]. Данное предположение часто выполняется на практике и позволяет построить алгоритм покадрового восстановления фона с пространственной сложностью, не зависящей от количества кадров во входной видеопоследовательности. Иллюстрация предложенного подхода к восстановлению фона приведена на рисунке 1.

Предложенная модель движения фона базируется на алгоритме вычисления оптического потока DeepFlow [8], краткое описание которого приведено в разделе 1.3.1. Ранее известные алгоритмы непараметрического моделирования движения фона включали в себя два отдельных шага: вычисление оптического потока в известной области и интерполяция в область неизвестного фона. Отличительной особенностью предложенной модели, подробное описание которой приведено в разделе 1.3.2, является объединение этих двух шагов в единую задачу оптимизации. Вычислим отображение $O^{t \rightarrow t+1}$ из кадра $I_t \subset I$ в соседний кадр с номером $t + 1$ на регулярной дискретной сетке $I_t^{(s)} \subset I_t$ с шагом s пикселей по обоим пространственным размерностям. Вокруг каждой точки $p = (x,y,t) \in I_t^{(s)}$ определим квадратную окрестность $P_{s \times s}(p)$, представляющую собой множество из s^2 точек пикселей. Тогда значения оптического потока как внутри области неизвестного фона $\Omega_t \subset I_t$, так и вне ее определяются путем решения следующей оптимизационной задачи:

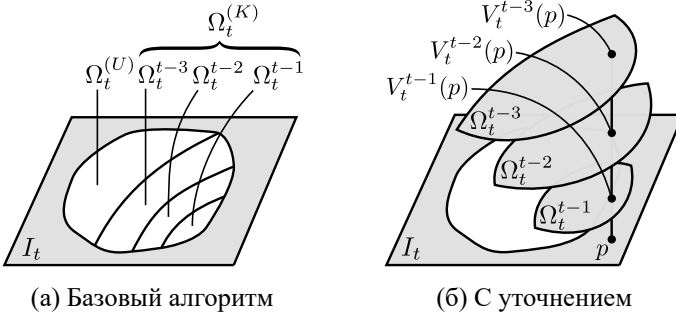
$$O^{t \rightarrow t+1} = \arg \min_O \sum_{\substack{p \in I_t^{(s)} \\ q \in P_{s \times s}(p) \\ q \in I_t \setminus \Omega_t \\ q + F_O(p) \in I_{t+1} \setminus \Omega_{t+1}}} \Phi \left(\sum w_q \left(\nabla_{F_O(p)} C(q) - \frac{\sum_q w_q \nabla_{F_O(p)} C(q)}{\sum_q w_q} \right)^2 \right) + \alpha \sum_{p \in I_t^{(s)}} \Phi \left(\left\| \frac{\partial F_O}{\partial x}(p) \right\|^2 + \left\| \frac{\partial F_O}{\partial y}(p) \right\|^2 \right). \quad (2)$$

Здесь $\Phi(x) = \sqrt{x + \epsilon^2}$, где $x \geq 0, \epsilon = 10^{-3}$, – устойчивая к выбросам функция потерь, $\Omega_{t+1} \subset I_{t+1}$ – область неизвестного фона в соседнем кадре, $w_q = (1 + \|\nabla_2 C(q)\|^2)^{-1}$ – веса, учитывающие значение модуля пространственно-градиента яркости $\nabla_2 C = (\frac{\partial C}{\partial x}, \frac{\partial C}{\partial y})$, аналогичные используемым в базовом алгоритме DeepFlow. $F_O(p) = (O(p) - p)$ обозначает вектор смещения в точке p , $\nabla_{F_O(p)} C(q)$ обозначает разницу между яркостью соответствующих точек: $C(q + F_O(p)) - C(q)$. α – это константа, определяющая вес слагаемого миними-

зируемой функции, отвечающего за гладкость поля векторов смещений F_O ; для упрощения обозначений \sum_q обозначает сумму по тому же множеству значений, что и первая сумма по q . Оптимизационная задача (2) сводится к последовательному решению набора систем линейных алгебраических уравнений с разреженной симметричной положительно определенной матрицей.

Отличительной особенностью предложенного алгоритма является низкая пространственная сложность. Она достигается за счет использования покадрового восстановления фона, которое позволяет снизить пространственную сложность с $O(WHT)$ до $O(WH)$. То есть, объем памяти, используемый предлагаемым алгоритмом, не зависит от количества кадров во входной видеопоследовательности, если не учитывать память, занимаемую входными и выходными данными. В разделе 1.3.3 приведено описание базового алгоритма покадрового восстановления фона. Данный алгоритм строит отображение $V(p)$ из уравнения (1) посредством нескольких проходов по входной последовательности кадров в прямом и обратном порядке. Положим, $V_t : \Omega_t^{(K)} \mapsto I$ – текущее отображение из кадра с номером t в известный фон с областью определения $\Omega_t^{(K)} \subset \Omega_t$. Изначально присвоим $\Omega_t^{(K)} := \emptyset$, где $t = 1, \dots, T$. Тогда алгоритм включает в себя следующие шаги:

1. **Прямой проход.** Для всех $p \in \Omega_t$ присвоим $V_t(p) := O^{t \rightarrow t-1}(p)$, если $O^{t \rightarrow t-1}(p) \in I_{t-1} \setminus \Omega_{t-1}$; $V_t(p) := V_{t-1}(O^{t \rightarrow t-1}(p))$, если $O^{t \rightarrow t-1}(p) \in \Omega_{t-1}^{(K)}$. Иначе p остается в неопределенной области $\Omega_t^{(U)} = \Omega_t \setminus \Omega_t^{(K)}$. Проведем восстановление фона в $\Omega_t^{(K)}$ согласно уравнению (1). Номер кадра t проходит все значения от 2 до T .
2. **Пространственное восстановление.** Применим алгоритм [7] к оставшейся невосстановленной области на последнем кадре $\Omega_T^{(U)}$. После этого шага положим $V_T(p) := p$ в рамках данной области.
3. **Обратный проход.** Для всех $p \in \Omega_t^{(U)}$ присвоим $V_t(p) := O^{t \rightarrow t+1}(p)$, если $O^{t \rightarrow t+1}(p) \in I_{t+1} \setminus \Omega_{t+1}$; $V_t(p) := V_{t+1}(O^{t \rightarrow t+1}(p))$, если $O^{t \rightarrow t+1}(p) \in \Omega_{t+1}^{(K)}$. Аналогично шагу 1 проведем восстановление в текущем кадре. Номер кадра t проходит все значения от $T - 1$ до 1.
4. **Пространственное восстановление.** Аналогично шагу 2 проведем восстановление на первом кадре.
5. **Прямой проход.** Аналогично шагу 1 для всех $p \in \Omega_t^{(U)}$ расширим область определения $V_t(p)$ в порядке $t = 2, \dots, T$ и восстановим фон.



(а) Базовый алгоритм

(б) С уточнением

Рис. 2 — Иллюстрация декомпозиции области Ω_t на непересекающиеся области определения отображений в исходные кадры в базовом алгоритме восстановления фона (а), и перекрывающиеся области определения $\Omega_t^{t_i}$ в предложенном алгоритме восстановления фона с уточнением (б).

Можно отметить, что на каждом шаге алгоритма область Ω_t можно представить в виде объединения непересекающихся областей определения отображений в различные исходные кадры и области $\Omega_t^{(U)}$, где отображение V_t на данном шаге еще не определено:

$$\Omega_t = \Omega_t^{(U)} \cup \Omega_t^{t_1} \cup \dots \cup \Omega_t^{t_n}, \text{ где } \Omega_t^{t_i} = \{p | V_t(p) \in I_{t_i}\}. \quad (3)$$

Отображение V_t тогда также можно разложить на отображения в отдельные исходные кадры $V_t^{t_i} : \Omega_t^{t_i} \mapsto I_{t_i}$.

Проблемой вышеописанного алгоритма покадрового восстановления фона является возможное накопление ошибок модели движения фона в значениях V_t по мере увеличения (уменьшения) номера кадра t . Для компенсации такого накопления ошибки был предложен новый алгоритм, описанный в разделе 1.3.4, который основан на проведении покадрового уточнения V_t . Для этого определим V_t как многозначную функцию, то есть $V_t(p) \subset I$, где мощность множества $|V_t(p)| = N(p)$, причем $\forall q_1, q_2 \in V_t(p) : q_1 \neq q_2 \Rightarrow \exists t : q_1, q_2 \in I_t$, то есть в каждой точке p есть $N(p)$ отображений в различные исходные кадры. Также расширим допустимую область определения V_t до $\bar{\Omega}_t$, которая является результатом морфологического расширения Ω_t с квадратным структурным элементом размера $(2d+1) \times (2d+1)$, где d – константа. Ключевая идея состоит во введении перекрытий ширины d между областями определения отображений в исходные кадры $\Omega_t^{t_i} \subset \bar{\Omega}_t$ из уравнения (3), что позволяет проводить выравнивание фрагментов фона, копируемых из различных исходных кадров, а перекрытие областей $\Omega_t^{t_i}$

с областью известного фона позволяет проводить выравнивание с фоном, окружающим восстанавливаемую область Ω_t . Поясняющая иллюстрация приведена на рисунке 2. Более подробно процедура построения многозначного отображения V_t и проведения покадрового восстановления фона с его помощью описана в разделе 1.3.4. По построению можно показать, что $\forall p : N(p) \leq (2d + 1)^2$, что позволяет сохранить оценку пространственной сложности алгоритма в $O(WH)$. Однако теперь перед восстановлением фона на каждом кадре проведем уточнение V_t путем решения оптимизационной задачи, где изначальное значение V_t , вычисленное по $V_{t\pm 1}$, используется в качестве начального приближения. Положим $V_t(p) = \{V_t^{v_1(p)}(p), \dots, V_t^{v_{N(p)}(p)}(p)\}$, где $v_i(p), i = 1 \dots N(p)$ – упорядоченные по возрастанию модуля разницы с t номера исходных кадров, в которые определены отображения в точке p . Тогда:

$$V_t = \arg \min_{V_t} (E_D(V_t) + \lambda E_S(V_t)), \text{ где} \quad (4)$$

$$E_D(V_t) = \sum_{\substack{p \in I_t^{(1)} \\ p \in \bar{\Omega}_t \setminus \Omega_t}} \sum_{i=1}^{N(p)} \Phi \left(\left(C(V_t^{v_i(p)}(p)) - C(p) \right)^2 \right) + \\ \sum_{\substack{p \in I_t^{(1)} \\ p \in \Omega_t}} \sum_{i=1}^{N(p)-1} \Phi \left(\left(C(V_t^{v_{i+1}(p)}(p)) - C(V_t^{v_i(p)}(p)) \right)^2 \right), \\ E_S(V_t) = \sum_{i=1}^n \sum_{\substack{p \in I_t^{(1)} \\ p \in \Omega_t^{t_i}}} \left\| \frac{\partial F_{V_t^{t_i}}}{\partial x}(p) \right\|^2 + \left\| \frac{\partial F_{V_t^{t_i}}}{\partial y}(p) \right\|^2.$$

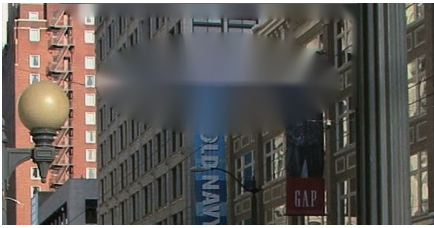
Здесь $I_t^{(1)} \subset I_t$ обозначает множество точек дискретной сетки с шагом в один пиксел; t_1, \dots, t_n – множество номеров кадров, в которые определены отображения в рамках V_t . Аналогично (2), задача (4) сводится к последовательному решению набора систем линейных алгебраических уравнений с разреженной симметричной положительно определенной матрицей. В зависимости от структуры V_t количество неизвестных в данных системах может быть различным, но оно не может превышать $(2d + 1)^2 \cdot 2 \cdot |I_t^{(1)} \cap \bar{\Omega}_t|$, так как $\forall p : N(p) \leq (2d + 1)^2$, как было упомянуто выше.

Изменение условий освещенности во входном видео с течением времени может приводить к появлению видимых швов в результате работы алгоритма,

если восстановление проводить путем копирования цвета согласно уравнению (1). В разделе 1.3.5 представлена модификация алгоритма Poisson Blending [9], позволяющая компенсировать изменения освещенности в процессе такого копирования, что завершает описание предложенного алгоритма. В разделе 1.4 приведен анализ вычислительной сложности построенного алгоритма и проводится сравнение оценок временной и пространственной сложности предложенного алгоритма с его аналогами.

В разделе 1.5 преимущества предложенного алгоритма перед аналогами демонстрируются путем сравнения результатов работы алгоритмов с эталоном по методологии, описанной в главе 2. Экспериментальная оценка была проведена на подготовленном наборе из 7 видеопоследовательностей с эталонным восстановленным фоном, по результатам которой предложенный алгоритм превзошел 6 алгоритмов с доступной реализацией. Также была проведена экспертная оценка на наборе из 26 тестовых видеопоследовательностей, на котором демонстрировались результаты авторами ранее предложенных алгоритмов. Предложенный алгоритм был применен к той же тестовой выборке и результаты работы различных алгоритмов показывались попарно 63 респондентам, принявшим участие в проведенном эксперименте. В каждой паре участнику требовалось выбрать видео с наименее заметными визуальными искажениями. По результатам экспертной оценки предложенный алгоритм не уступил алгоритму из работы 2016 года [10]. Детали данного эксперимента приведены в разделе 1.6. Также, как показано в разделе 1.7, авторская реализация предложенного алгоритма обладает значительно более высокой скоростью обработки видео по сравнению с известными реализациями аналогов. Результаты работы предложенного алгоритма на широком наборе тестовых видеопоследовательностей публично доступны в сети Интернет по адресу http://videocompletion.org/fast_video_completion/.

Вторая глава посвящена методике сравнения алгоритмов восстановления фона в видео. Из обзора существующих подходов к измерению качества восстановленного фона, приведенного в разделе 2.2, следует, что известные методики обладают существенными ограничениями и не предназначены для оценки крупных областей восстановления, которые также имеют большую временную продолжительность. Измерение качества работы алгоритмов восстановления фона в данном случае представляет особую сложность, так как прямые оценки сходства с эталоном имеют сравнительно низкую корреляцию с визуальным качеством



$$(a) \frac{1}{n} \|C(\Omega) - C_{GT}(\Omega)\|_2^2 = 899.16 \quad (b) \frac{1}{n} \|C(\Omega) - C_{GT}(\Omega)\|_2^2 = 955.98$$

Рис. 3 — Результат работы двух алгоритмов восстановления фона. Результат слева (а) обладает значительно более низким визуальным качеством по сравнению с результатом справа (б) согласно экспертной оценке, однако согласно метрике L_2 результат слева ближе к эталону.

результата, определяемым посредством экспертной оценки. На рисунке 3 проиллюстрирован недостаток L_2 нормы разницы между результатом восстановления и эталоном как способа сравнения качества работы алгоритмов.

Предлагаемая методика состоит из двух основных компонент: тестового набора видеопоследовательностей с эталонными изображениями фона и метода сравнения результатов работы алгоритмов с эталоном для определения качества. В процессе составления тестового набора во внимание принимался ряд принципов. Во-первых, каждая тестовая видеопоследовательность должна представлять сложность для существующих алгоритмов, иначе невозможно надежно ранжировать алгоритмы по качеству работы, так как их результаты будут близки к одинаковым. Также тестовая выборка должна покрывать широкий набор типов входных данных, включая последовательности со статичной и свободно перемещающейся камерой, статический и динамический фон, динамические текстуры в видео (например, поверхность воды, огонь). Для обеспечения наличия эталона тестовые примеры создавались посредством наложения различных объектов переднего плана на фоновые видеопоследовательности средствами компьютерной графики. Пример из построенной тестовой последовательности приведен на рисунке 4. Более подробно процесс подготовки тестового набора данных описан в разделе 2.3.

В разделе 2.4 приводятся определения множества различных функционалов качества восстановленного фона, которые были исследованы в данной работе. Предположим, что известны эталонные значения яркости $C_{GT} : I \mapsto R$ (в проведенных экспериментах учет цветовых компонент оказал малое влияние на



(а) Входной кадр

(б) Входная маска

(в) Эталонный фон

Рис. 4 — Пример из составленной тестовой выборки, включающий входной кадр видео (а), соответствующую маску области неизвестного фона (б), и эталонное изображение фона (в).

результат, поэтому снова рассматривается только яркость). Требуется построить функционал качества $\mathcal{F}[C, C_{GT}] = q \in \mathbb{R}$, согласованный с экспертной оценкой. Для определенности положим, что чем меньше значение такого функционала, тем выше должно быть визуальное качество результата. Предлагаемые функционалы качества основаны на широко известном методе SSIM [11], который измеряет качество изображений/видео посредством сравнения с эталоном. Индекс сходства SSIM вычисляется между C и C_{GT} , определенных на двумерных блоках $s \times s$ пикселей $P, Q \subset I$: $-1 \leq \text{SSIM}[C(P), C_{GT}(Q)] \leq 1$ согласно работе [11]. Значение индекса 1 достигается лишь в случае $C(P) \equiv C_{GT}(Q)$, а чем меньше его значение, тем больше заметна разница между C и C_{GT} , определенных на P и Q соответственно, с точки зрения наблюдателя. Ключевой идеей, которая позволила повысить корреляцию с экспертной оценкой по сравнению с известными функционалами, является использование критерия локального сходства восстановленного фона с эталоном. Другими словами, вместо сравнения с эталоном в той же самой точке ($P = Q$), для каждого P предлагается найти соответствующее Q , которое максимизирует сходство $C(P)$ и $C_{GT}(Q)$. На основе этой идеи определяется функционал качества MS- C_{DSSIM} :

$$C_{\text{DSSIM}}[C, C_{GT}] = \frac{1}{N} \sum_{\substack{P \subset I \\ P \cap \Omega \neq \emptyset}} \min_{Q \subset I} (1 - \text{SSIM}[C(P), C_{GT}(Q)]), \quad (5)$$

$$\text{MS-}C_{\text{DSSIM}}[C, C_{GT}] = \sum_{i=0}^{M-1} w_i \cdot C_{\text{DSSIM}}[C^i, C_{GT}^i].$$

Здесь N обозначает количество слагаемых в последующей сумме (все возможные блоки P фиксированного размера, пересекающиеся хотя бы в одном пик-

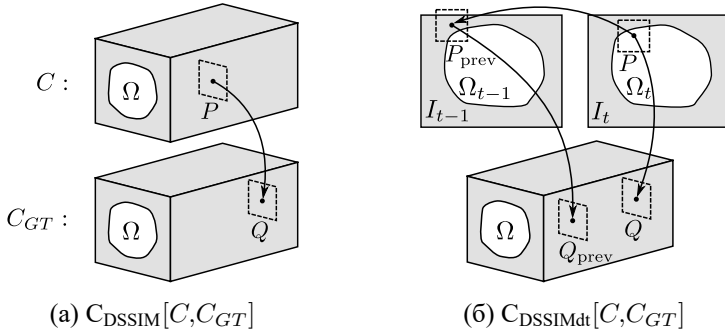


Рис. 5 — Иллюстрация принципа работы двух предложенных функционалов качества восстановленного фона, превзошедших ранее известные по корреляции с экспертной оценкой.

селе с Ω). Оценка проводится независимо на M уровнях Гауссовой пирамиды, результаты которых суммируются с весами w_i , позволяющими учесть различную чувствительность зрительной системы человека к искажениям разного пространственного масштаба. Верхний индекс в C^i и C_{GT}^i , соответственно, обозначает рассматриваемый уровень Гауссовой пирамиды (каждый кадр уменьшается по обоим пространственным размерностям в 2^i раз на i -ом уровне пирамиды).

Зрительная система человека обладает высокой чувствительностью к неестественным различиям между соседними кадрами, или, другими словами, к нестабильности результата обработки видео во времени. На базе критерия локального сходства можно также построить функционал для оценки временной стабильности восстановленного фона вместо качества восстановления в отдельных кадрах, что оценивается функционалом (5). Определим данный функционал как:

$$C_{DSSIMdt}[C, C_{GT}] = \frac{1}{N} \left| \sum_{\substack{P \subset I \\ P \cap \Omega \neq \emptyset}} \min_{Q \subset I} (1 - SSIM[C(P), C_{GT}(Q)]) - \min_{Q \subset I} (1 - SSIM[C(P_{prev}), C_{GT}(Q)]) \right|, \quad (6)$$

$$P_{prev} = \arg \max_{Q \subset \Omega_{prev}^{w \times w}(P)} SSIM[C(P), C(Q)].$$

Здесь $\Omega_{prev}^{w \times w}(P)$ обозначает окрестность размера $w \times w$ пикселей вокруг P , но на предыдущем относительно него кадре. В авторской реализации используется w равное $1/10$ ширины кадра. Таким образом, данный функционал оценивает насколько локальное сходство с эталоном изменяется между соседними кадрами.

Финальный функционал $MS-C_{DSSIMdt}$ определяется по аналогии с функционалом (5). Принцип работы обоих предложенных функционалов качества восстановленного фона проиллюстрирован на рисунке 5.

Для сравнения различных вводимых функционалов качества была проведена экспертная оценка 6 алгоритмов восстановления фона на 7 тестовых видеопоследовательностях путем попарного сравнения результатов их работы 341 респондентом, принявшим участие в данном эксперименте. Его подробное описание приведено в разделе 2.5. Собранные данные позволяют оценить корреляцию функционалов качества с визуальным качеством, измеренным в ходе экспертной оценки, а также определить оптимальные значения весов в (5). В разделе 2.6 приведены результаты данного сравнения и демонстрируется превосходство функционалов, основанных на вышеописанном критерии локального сходства, перед ранее предложенными в научных работах функционалами качества. Четыре функционала были отобраны для использования в предложенной методике, которые позволяют проводить многопараметрическую объективную оценку и сравнение алгоритмов восстановления фона в видео. Для упрощения доступа к результатам проведенного сравнения, а также для обеспечения возможности включать в сравнение новые алгоритмы, был создан сайт <http://videocompletion.org>.

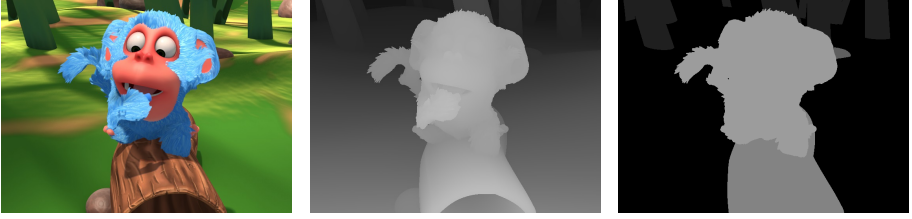
Третья глава посвящена решению задачи преобразования видео в стереоскопический формат на базе описанного в первой главе алгоритма восстановления фона. Для этого требуется построить левый и правый ракурсы выходной стереоскопической видеопоследовательности по одному входному. Трансформация исходного ракурса в целевой может описана с помощью отображения $S(p) = p + (\mathcal{D}(p), 0, 0)$, где $p = (x, y, t) \in I$, а функция горизонтального смещения $\mathcal{D} : I \mapsto \mathbb{R}$ называется диспаратностью соответствующих точек исходного ракурса. Диспаратность может быть однозначным образом вычислена из значения глубины точки p (расстояния от плоскости линзы камеры), фокусного расстояния камеры и значения стереобазы создаваемого стереоскопического видео (расстояния между оптическими осями виртуальных камер, снимающих левый и правый ракурс). Для части точек целевого ракурса q можно найти p такое, что $S(p) = q$; в таком случае цвет целевого ракурса в q может быть определен как цвет точки исходного ракурса с наименьшим значением глубины среди тех, которые отображаются в q отображением S . Однако для точек в области $\{q | \nexists p : S(p) = q\}$ невозможно вычислить цвет целевого ракурса путем копиро-



Рис. 6 — Пример отображения S , преобразующего исходный ракурс в целевой. В области целевого ракурса, обозначенной белым цветом, неизвестны значения цвета, так как в нее не отображается ни одна из точек исходного ракурса. Такая область называется областью открытия в целевом ракурсе. Одной из важных подзадач задачи построения стереоскопических ракурсов является задача вычисления цвета в областях открытия.

вания цвета из соответствующей точки исходного ракурса. Такая область называется областью открытия и включает фрагменты фона, которые были скрыты объектами переднего плана в исходном ракурсе, но становятся видимыми после изменения ракурса просмотра (иллюстрация приведена на рисунке 6). В данной главе приведено описание предлагаемого решения подзадачи вычисления цвета в области открытия посредством восстановления фона.

Раздел 3.2 содержит обзор существующих алгоритмов решения задачи построения стереоскопических ракурсов. Раздел 3.3 содержит описание предлагаемого алгоритма, ключевой отличительной особенностью которого является проведение многослойного восстановления фона, что позволяет точнее вычислять цвет в областях открытия, возникающих при перекрытии трех и более объектов. Предложенный алгоритм опирается на наличие карты слоев $\mathcal{L} : I \mapsto \mathbb{N}$, которая каждой точке $p \in I$ ставит в соответствие номер слоя l , причем слои с более высоким номером могут перекрывать слои с более низким номером, в то время как обратное неверно. Пример такой карты слоев приведен на рисунке 7. Для каждого слоя l можно определить модель движения $O_l^{t \rightarrow t \pm 1}$ согласно уравнению (2) с использованием $\Omega_{\mathcal{L}}^l = \{p | \mathcal{L}(p) > l\}$ в качестве области неизвестного фона. Для всех точек $p \in \Omega_{\mathcal{L}}^{l, (K)} \subset \Omega_{\mathcal{L}}^l$, из которых слой l (то есть, область $\{p | \mathcal{L}(p) = l\}$) достижим путем применения композиции $O_l^{t \rightarrow t+1}$ или $O_l^{t \rightarrow t-1}$, цвет данного слоя $C_l(p)$ может быть вычислен копированием из известной области по аналогии с уравнением (1). То есть, задача вычисления цвета слоев $C_l : \Omega_{\mathcal{L}}^{l, (K)} \mapsto \mathbb{R}^3$ сводится к рассмотренной в первой главе задаче восста-



(а) Исходный кадр (б) Карта диспаратности (в) Карта слоев

Рис. 7 — Пример кадра (а) с дополнительными входными данными в виде карты значений диспаратности \mathcal{D} (б) и карты слоев \mathcal{L} (в).

новления фона в видео. В предположении, что $\mathcal{D}(p) = \mathcal{D}(O^{t \rightarrow t \pm 1}(p))$, то есть постоянства значений диспаратности между кадрами вдоль оптического потока $O^{t \rightarrow t \pm 1}$, точно так же путем копирования могут быть вычислены значения диспаратности слоев $\mathcal{D}_l : \Omega_{\mathcal{L}}^{l, (K)} \mapsto \mathbb{R}$. Зная значения цвета C_l и диспаратности \mathcal{D}_l , целевой ракурс может быть построен путем отображения исходного ракурса и каждого из восстановленных слоев согласно отображению S и композиции в порядке уменьшения глубины (приближения к плоскости линзы камеры). Для вычисления цвета в оставшихся точках, в которые не отображается ни один из восстановленных слоев, предложенный алгоритм использует технику горизонтальной экстраполяции крайнего пиксела с бóльшим значением глубины, описанную авторами работы [12].

Однако если предположение о постоянстве диспаратности между кадрами не выполнено, то требуется также моделировать межкадровое изменение диспаратности $\Delta \mathcal{D}_l^{t \rightarrow t \pm 1} : I \mapsto \mathbb{R}$ для каждого слоя l . В разделе 3.3.1 описано обобщение модели движения фона (2), которое позволяет это учесть. Оно имеет следующий вид:

$$\begin{aligned}
 O_l^{t \rightarrow t+1}, \Delta \mathcal{D}_l^{t \rightarrow t+1} = & \arg \min_{O, \Delta \mathcal{D}} \sum_{p \in I_t^{(s)}} \Phi \left(\sum_{\substack{q \in P_{s \times s}(p) \\ \mathcal{L}(q) \leq l \\ \mathcal{L}(q + F_O(p)) \leq l}} w_q \left(\nabla_{F_O(p)} C(q) - \frac{\sum_q w_q \nabla_{F_O(p)} C(q)}{\sum_q w_q} \right)^2 \right) \\
 & + \Phi \left(\sum_{\substack{q \in P_{s \times s}(p) \\ \mathcal{L}(q) \leq l \\ \mathcal{L}(q + F_O(p)) \leq l}} (\nabla_{F_O(p)} \mathcal{D}(q) - \Delta \mathcal{D}(p))^2 \right) + \sum_{p \in I_t^{(s)}} \alpha_s \Phi \left(\left\| \frac{\partial F_O}{\partial x}(p) \right\|^2 + \left\| \frac{\partial F_O}{\partial y}(p) \right\|^2 \right) \\
 & + \alpha_d \Phi \left(\left\| \frac{\partial \Delta \mathcal{D}}{\partial x}(p) \right\|^2 + \left\| \frac{\partial \Delta \mathcal{D}}{\partial y}(p) \right\|^2 \right). \tag{7}
 \end{aligned}$$

Здесь $\nabla_{F_O(p)} \mathcal{D}(q)$ обозначает разницу между диспаратностью соответствующих точек: $\mathcal{D}(q) - \mathcal{D}(q + F_O(p))$, α_s – константа, определяющая вес слагаемого, отвечающего за гладкость поля векторов смещений F_O , α_d определяет вес слагаемого, отвечающего за гладкость функции межкадрового изменения глубины $\Delta \mathcal{D}$. Остальные обозначения идентичны используемым в (2).

Затем для каждого слоя необходимо построить отображение V в исходные кадры и соответствующее суммарное изменение диспаратности между исходным кадром и текущим $V^{(\Delta \mathcal{D})}$ (для упрощения обозначений номер текущего слоя l был опущен). Схоже с уравнением (1), цвет и диспаратность текущего слоя в точке $p \in \Omega_{\mathcal{L}}^{l,(K)}$ тогда могут быть вычислены следующим образом:

$$\begin{aligned} C_l(p) &= C(V(p)), \text{ где } \mathcal{L}(V(p)) = l \\ D_l(p) &= \mathcal{D}(V(p)) + V^{(\Delta \mathcal{D})}(p) \end{aligned} \quad (8)$$

V и $V^{(\Delta \mathcal{D})}$ строятся посредством нескольких проходов по входной последовательности кадров по аналогии с алгоритмом, описанным в первой главе. В процессе построения отображения V_t из каждого следующего кадра с номером t и соответствующего изменения диспаратности $V_t^{(\Delta \mathcal{D})}$ точно так же можно ввести перекрытия между областями определения отображений в различные исходные кадры и определить оптимизационную задачу уточнения, целью которой является компенсация накопления ошибки в значениях V_t и $V_t^{(\Delta \mathcal{D})}$. Подробное описание алгоритма покадрового восстановления фона с учетом изменения диспаратности приводится в разделе 3.3.2.

В разделе 3.4 преимущества предложенного алгоритма построения стереоскопических ракурсов перед аналогами демонстрируются путем экспертной оценки. Результаты работы различных алгоритмов демонстрировались попарно 152 респондентам; в каждой паре требовалось выбрать результат с более высоким визуальным качеством. Каждый участник данного эксперимента сравнил 20 пар результатов, включая 2 контрольных вопроса. По результатам проведенной экспертной оценки предложенный алгоритм превзошел аналоги по визуальному качеству.

В **заключении** сформулированы основные результаты диссертационного исследования, рассмотрены возможные варианты их применения и обозначены перспективы дальнейших исследований.

Заключение

В ходе диссертационного исследования были получены следующие основные результаты:

1. Предложен новый подход к решению проблемы заполнения областей открытия на основе проведения многослойного восстановления фона в видео с использованием непараметрической модели движения для каждого слоя.
2. Разработан новый алгоритм восстановления фона в видео с пространственной сложностью, не зависящей от количества кадров во входной видеопоследовательности. По результатам экспериментальной оценки предложенный алгоритм превзошел существующие аналоги по точности найденного решения и не уступил аналогам по результатам экспертной оценки.
3. Предложено два новых функционала качества работы алгоритмов восстановления фона в видео, основанных на критерии локального сходства с эталоном, которые обладают более высокой корреляцией с экспертной оценкой по сравнению с ранее использовавшимися для этой цели функционалами.
4. Все предложенные алгоритмы и функционалы качества были реализованы, а также проведены вычислительные эксперименты, подтверждающие их эффективность. Алгоритм восстановления фона был протестирован на реальных данных, предоставленных коммерческой компанией.

Предложенные алгоритмы могут быть использованы для улучшения существующих программных инструментов для полуавтоматического преобразования моноскопического видео в стереоскопический формат, а предложенная методология сравнения алгоритмов восстановления фона в видео может быть использована в ходе разработке новых алгоритмов и подборе их параметров. Дальнейшее развитие темы исследования может включать в себя:

- Исследование альтернативных моделей движения фона, в том числе предсказание векторного поля движения в области неизвестного фона средствами машинного обучения, например с помощью нейронных сетей.
- Исследование нейросетевых алгоритмов восстановления фона в изображениях для повышения качества работы для фрагментов фона, остаю-

щихся скрытыми на протяжении всей входной видеопоследовательности.

Публикации автора по теме диссертации

В изданиях из списка ВАК РФ

1. *Боков А. А., Ватолин Д. С.* Восстановление фона в видео с использованием непараметрической модели движения и покадрового уточнения // Цифровая обработка сигналов. — 2017. — № 3. — С. 13—21.
2. *Боков А. А., Ватолин Д. С.* Методика объективной оценки качества восстановления фона в видео // Цифровая обработка сигналов. — 2016. — № 3. — С. 26—33.

В изданиях, индексируемых в Scopus и Web of Science

3. Toward an Objective Benchmark for Video Completion / A. Bokov, D. Vatolin, M. Erofeev, Y. Gitman // Signal, Image and Video Processing. — 2019. — Vol. 13. — P. 601–608. — DOI: 10.1007/s11760-018-1387-5.
4. *Bokov A., Vatolin D.* Multilayer RGBD-Video Completion For Hole Filling in 3D-View Synthesis // International Conference on 3D Immersion (IC3D). — 2018. — P. 1–8. — DOI: 10.1109/IC3D.2018.8657873.
5. *Bokov A., Vatolin D.* 100+ Times Faster Video Completion by Optical-Flow-Guided Variational Refinement // IEEE International Conference on Image Processing (ICIP). — 2018. — P. 2122–2126. — DOI: 10.1109/ICIP.2018.8451683.
6. *Bokov A., Vatolin D.* Toward efficient background reconstruction for 3D-view synthesis in dynamic scenes // IEEE International Conference on Multimedia Expo Workshops (ICMEW). — 2017. — P. 37–42. — DOI: 10.1109/ICMEW.2017.8026297.

Список литературы

7. *He K., Sun J.* Statistics of patch offsets for image completion // European Conference on Computer Vision (ECCV). — Springer, 2012. — С. 16—29. — DOI: 10.1007/978-3-642-33709-3_2.

8. DeepFlow: Large displacement optical flow with deep matching / P. Weinzaepfel, J. Revaud, Z. Harchaoui, C. Schmid // International Conference on Computer Vision (ICCV). — 2013. — С. 1385—1392. — DOI: 10.1109/ICCV.2013.175.
9. Pérez P., Gangnet M., Blake A. Poisson image editing. — 2003. — DOI: 10.1145/1201775.882269.
10. Temporally coherent completion of dynamic video / J.-B. Huang, S. B. Kang, N. Ahuja, J. Kopf // ACM Transactions on Graphics (TOG). — 2016. — Т. 35, № 6. — 196:1—196:11. — DOI: 10.1145/2980179.2982398.
11. Image quality assessment: from error visibility to structural similarity / Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli // IEEE Transactions on Image Processing (TIP). — 2004. — Т. 13, № 4. — С. 600—612. — DOI: 10.1109/TIP.2003.819861.
12. Vázquez C., Tam W. J., Speranza F. Stereoscopic imaging: filling disoccluded areas in depth image-based rendering // Three-Dimensional TV, Video, and Display V. T. 6392. — 2006. — С. 63920D.

Исследование выполнено при финансовой поддержке РФФИ в рамках гранта 15-01-08632 а.