

Московский государственный университет имени М. В. Ломоносова  
Факультет вычислительной математики и кибернетики

На правах рукописи



Кривовязь Глеб Робертович

# **Алгоритмы трехмерной реконструкции по изображениям и данным лазерного сканирования**

Специальность 05.13.11 – математическое и программное обеспечение  
вычислительных машин, комплексов и компьютерных сетей

Автореферат  
диссертации на соискание учёной степени  
кандидата физико-математических наук

Москва – 2013

Работа выполнена на кафедре автоматизации систем вычислительных комплексов факультета вычислительной математики и кибернетики Московского государственного университета имени М. В. Ломоносова.

Научный руководитель: Кандидат физико-математических наук, доцент  
Баяковский Юрий Матвеевич

Официальные оппоненты: Доктор физико-математических наук, доцент, старший научный сотрудник Института прикладной математики им. М.В. Келдыша Российской академии наук  
Богуславский Андрей Александрович

Кандидат физико-математических наук, старший научный сотрудник механико-математического факультета МГУ имени М.В. Ломоносова  
Иванов Денис Владимирович

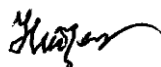
Ведущая организация: Государственный научно–исследовательский институт авиационных систем (ГосНИИАС)

Защита состоится 29 марта 2013 г. в 11 часов на заседании диссертационного совета Д 501.001.44 при Московском государственном университете имени М.В. Ломоносова по адресу: 119991, ГСП-1, Москва, Ленинские горы, МГУ, 2-й учебный корпус, факультет ВМиК, аудитория 685. Желаящие присутствовать на заседании диссертационного совета должны сообщить об этом за 2 дня до указанной даты по тел. (495) 939-30-10 (для оформления заявки на пропуск).

С диссертацией можно ознакомиться в Фундаментальной библиотеке МГУ имени М.В. Ломоносова. С текстом автореферата можно ознакомиться на официальном сайте ВМК МГУ имени М.В. Ломоносова <http://www.cmc.msu.ru> в разделе «Наука» - «Работа диссертационных советов» - «Д 501.001.44».

Автореферат разослан 27 февраля 2013 г.

Учёный секретарь  
диссертационного совета  
профессор



Н.П. Трифонов

## Общая характеристика работы

### Объект исследования и актуальность работы

Автоматическая трехмерная реконструкция сцен окружающего мира является одной из классических задач компьютерного зрения и графики. Трехмерные модели находят свое применение в системах виртуальной и дополненной реальности, робототехнике, медицине, геоинформационных системах и ряде других областей.

Постоянный рост вычислительных возможностей, появление новых областей применения и развитие новых технологий ставят новые условия и выдвигают новые требования к задаче трехмерной реконструкции. Так, бурное развитие сети Интернет и разнообразных веб-ресурсов породило задачу автоматического создания трехмерных моделей популярных мест и объектов по большим неупорядоченным наборам пользовательских фотографий, скачанных из сети<sup>1,2</sup> (Рисунок 1).



**Рисунок 1.** Пример изображений и построенной по ним трехмерной модели

Другим важным аспектом, обуславливающим современное состояние задачи трехмерного моделирования, является стремительное развитие различных типов сенсоров. Все большее распространение получают сенсоры глубины (например, Kinect) и лазерные сканеры, используемые для получения информации о трехмерной структуре сцены. Для автоматического создания качественных трехмерных моделей необходимо совместное

---

<sup>1</sup> Snavely N., Seitz S., Szeliski R. Modeling the World from Internet Photo Collections // International Journal of Computer Vision. 2008. 80. N 2. P 189-210.

<sup>2</sup> Agarwal S., Furukawa Y., Snavely N., Simon I., Curless B., Seitz S., Szeliski R. Building Rome in a Day // Communications of the ACM. 2011. 54. N 10. P 105-112.

использование данных, полученных с разных датчиков: изображений, облаков трехмерных точек, карт глубины.

В данной работе рассматривается общая схема современных алгоритмов трехмерного моделирования, выделяется ряд важных задач и предлагаются способы их решения. Первая глава диссертации посвящена задаче сопоставления изображений: описывается новая двухуровневая схема сопоставления, позволяющая найти правильные параметры модели даже в случае малого перекрытия снимков. Во второй главе рассматривается задача плотной стереорекострукции по изображениям и предлагается новый алгоритм бинокулярного стерео, сочетающий высокую вычислительную эффективность и хорошее качество работы на сценах антропогенной природы. В случае, когда на входе системы моделирования имеются как изображения, так и данные лазерного сканирования для их совместного использования необходимо уточнение параметров внешнего ориентирования камер относительно облака трехмерных точек. Этой задаче посвящена третья глава работы.

## **Цель диссертационной работы**

Целью работы является разработка комплекса алгоритмов построения трехмерной модели сцены по набору изображений и данным лазерного сканирования. Разработанные алгоритмы не должны требовать пользовательского ввода и должны превосходить существующие методы по точности.

## **Научная новизна работы**

В диссертации предложена новая двухуровневая схема сопоставления изображений на основе метода голосования Хафа<sup>3</sup>, позволяющая достичь большей надежности сопоставления в случае малого (менее 20%) перекрытия снимков. Показано превосходство предложенной схемы над существующими подходами как на синтетических, так и на реальных данных.

Разработан новый алгоритм плотной стереорекострукции по паре изображений. Алгоритм ищет решение в пространстве плоскостей и использует контрольные точки для генерации гипотез о плоскостях, присутствующих в сцене. Приведена экспериментальная оценка алгоритма на открытом тестовом наборе, состоящем из изображений наземной

---

<sup>3</sup> Hough P. A method and means for recognizing complex patterns // U.S. Patent #3069654. 1962.

городской съемки. Алгоритм показывает более низкий процент ошибок, чем современный метод, сравнимый с ним по скорости работы.

Также в рамках работы предложен новый алгоритм регистрации изображений относительно облака трехмерных точек. В отличие от существующих подходов, предложенный алгоритм учитывает одновременно как статистическую согласованность между изображением и картой глубины, так и эпиполярные ограничения между парой снимков. Показано, что добавление эпиполярных ограничений повышает процент успешных регистраций.

## **Практическая значимость и реализация**

В рамках работы выполнена программная реализация разработанных алгоритмов, удовлетворяющая всем требованиям и ограничениям, сформулированным при постановке задач.

Предложенный алгоритм сопоставления изображений разрабатывался в рамках проекта с компанией ООО «Ракурс». Реализация алгоритма встроена в программный продукт PHOTOMOD.

Тестирование предложенного алгоритма плотной стереорекострукции по паре изображений производилось на опубликованном в 2012 году наборе данных KITTI (Karlsruhe Institute of Technology and Toyota Technological Institute)<sup>4</sup>. Данный набор, включающий в себя несколько сотен пар изображений наземной городской съемки, является на сегодняшний день одним из наиболее сложных тестовых наборов для рассматриваемой задачи.

Для экспериментальной оценки разработанного алгоритма регистрации изображений относительно облака трехмерных точек был использован тестовый набор Vaihingen, предоставленный Немецким Обществом Фотограмметрии, Дистанционного Зондирования и Геоинформатики (DGPF)<sup>5</sup>.

## **Апробация работы**

Основные результаты работы докладывались и обсуждались на:

---

<sup>4</sup> [http://www.cvlibs.net/datasets/kitti/eval\\_stereo\\_flow.php?benchmark=stereo](http://www.cvlibs.net/datasets/kitti/eval_stereo_flow.php?benchmark=stereo)

<sup>5</sup> Cramer M. The DGPF Test on Digital Aerial Camera Evaluation – Overview and Test Design // Photogrammetrie Fernerkundung Geoinformation. 2010. 2. P 73–82.

- 21-ой международной конференции по компьютерной графике и машинному зрению «GraphiCon-2011», Россия, Москва, 2011;
- 6-й летней школе Microsoft для аспирантов (Microsoft Research PhD Summer School), Англия, Кембридж, 2011;
- Международной конференции по фотограмметрии, компьютерному зрению и анализу изображений PCV (Photogrammetric Computer Vision and Image Analysis), Франция, Париж, 2010;
- 17-й международной научной конференции студентов, аспирантов и молодых ученых «Ломоносов-2010», Россия, Москва, 2010;
- семинаре по компьютерной графике и машинному зрению под руководством Ю. М. Баяковского (факультет ВМК МГУ);
- семинаре аспирантов кафедры АСВК факультета ВМК МГУ под руководством Л. Н. Королева.

## **Публикации**

По теме диссертации автором опубликовано 6 научных работ, из них 3 статьи в рецензируемых журналах, включенных в перечень ВАК [3,5,6], 2 статьи [1,4] и 1 тезисная публикация [2] в сборниках трудов международных конференций.

## **Структура и объем работы**

Диссертация состоит из введения, трех глав, заключения и списка литературы. Содержание работы изложено на 115 страницах. Список литературы включает 88 наименований. В работе содержится 32 рисунка и 3 таблицы.

## **Содержание работы**

**Во введении** описывается общая задача трехмерного моделирования по изображениям и данным лазерного сканирования, выделяются основные шаги ее решения. Задача трехмерного моделирования на сегодняшний день не решена: несмотря на то, что существуют алгоритмы, позволяющие получить результат полностью автоматически, есть данные, на которых они не срабатывают. Поэтому в реальных практических системах почти на каждом шаге используется ручной ввод либо коррекция результатов, что обуславливает необходимость разработки новых алгоритмов.

**В первой главе** рассматривается задача сопоставления изображений. Под сопоставлением изображений  $I_1$ ,  $I_2$  некоторой сцены понимается поиск такого преобразования (параметров модели)  $\Psi$ , которое переводит точки одного изображения в соответствующие точки другого:  $\Psi(I_1) = I_2$ . В качестве искомой модели, в зависимости от задачи, может выступать, например, гомография или фундаментальная матрица.

Стандартным подходом к решению задачи сопоставления пары снимков является последовательное выполнение следующих 4 шагов:

1. Поиск, или детектирование, характерных точек на каждом из изображений. *Характерной точкой*, или *точечной особенностью*, называется точка, окрестность которой отличается от окрестностей близлежащих точек по некоторой мере. Одними из наиболее часто используемых на практике точечных особенностей являются уголки Харриса<sup>6</sup>.
2. Представление окрестности каждой найденной точки в виде вектора признаков, называемого *дескриптором* точки. Дескрипторы особых точек различаются по инвариантности к тем или иным преобразованиям изображения, таким как изменения яркости, сдвиг, поворот. Примером широко используемого дескриптора характерных точек является SIFT<sup>7</sup>.
3. Поиск пар соответствующих точек по расстоянию между их дескрипторами. Каждой характерной точке одного изображения ставится в соответствие ближайшая в пространстве дескрипторов точка второго изображения. Как правило, используют также дополнительные проверки: на взаимную однозначность соответствия, на абсолютное значение расстояния, на отношение расстояний до ближайшей и до второй ближайшей точки.
4. Вычисление искомого преобразования путем применения одного из алгоритмов устойчивой оценки параметров модели на основе случайных выборок.

Остановимся на последнем шаге подробнее. Поскольку после 3-го шага неизбежно возникают ложные пары точек (выбросы), для оценки параметров модели необходим устойчивый к шуму алгоритм. Одним из таких алгоритмов, предложенным еще в 1961

---

<sup>6</sup> Harris C., Stephens M. A combined corner and edge detector // Proceedings of The Fourth Alvey Vision Conference. 1988. P 147–151.

<sup>7</sup> Lowe D. G. Distinctive image features from scale-invariant keypoints // International Journal of Computer Vision. 2004. 60. N 2. P 91-110.

году, является алгоритм RANSAC (RANdom SAmple Consensus)<sup>8</sup>. Идеей алгоритма RANSAC является использование случайных выборок из набора входных данных (в случае задачи сопоставления изображений – пар характерных точек) для генерации гипотез и выбор той гипотезы, которая согласуется с наибольшим числом данных. Скорость сходимости алгоритма и вероятность нахождения правильной модели зависят от сложности модели (количества пар точек, необходимых для генерации гипотезы) и от доли выбросов во входных данных.

В последние годы предложено множество модификаций алгоритма RANSAC, призванных повысить скорость сходимости и надежность алгоритма за счет учета особенностей той или иной задачи. В качестве базового метода в данной работе рассматривается алгоритм PROSAC (PROgressive SAmple Consensus)<sup>9</sup>, ключевая идея которого – в сортировке пар характерных точек по евклидову расстоянию между дескрипторами и использовании для генерации гипотез в первую очередь пар с меньшим расстоянием. Таким образом, более надежные пары проверяются первыми, что повышает скорость нахождения хорошей гипотезы.

Однако, как уже упоминалось выше, существенное влияние на скорость сходимости алгоритмов, основанных на случайных выборках, оказывает доля выбросов в исходном наборе точечных соответствий. Предположим, искомой моделью является фундаментальная матрица, требующая 7 пар точек для вычисления ее параметров. Если доля выбросов составляет 50%, то для нахождения правильного решения с вероятностью 0.95 алгоритму RANSAC достаточно около  $4 \cdot 10^2$  итераций. Но если доля выбросов достигает 90%, то уже необходимо порядка  $3 \cdot 10^7$  итераций, а при 95% -  $4 \cdot 10^9$  итераций. Алгоритм PROSAC может достигать сходимости быстрее, но при значительном уровне выбросов сортировка пар точек по расстоянию между дескрипторами не позволяет надежно отделить верные пары точек от неправильных, поэтому для него также характерен резкий рост необходимого числа итераций. На практике это приводит либо к существенному увеличению времени работы алгоритма, либо к большему числу неудачных сопоставлений.

---

<sup>8</sup> Fischler M., Bolles R. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography // Transactions of ACM - Graphics and Image Processing. 1981. 24. P 381–395.

<sup>9</sup> Chum O., Matas J. Matching with PROSAC - progressive sample consensus // Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2005. P 220-226.



В реальных задачах столь высокий уровень выбросов в исходных парах точек – не редкость. Как правило, подобная ситуация возникает в случае малого перекрытия снимков. При условии более или менее равномерного распределения характерных точек по изображению долю правильных соответствий можно считать примерно равной отношению площади перекрытия снимков к площади одного изображения. В случае больших неупорядоченных наборов изображений случаи малого перекрытия возникают неизбежно. При этом важно успешно сопоставлять и такие изображения, поскольку, чем больше связей между снимками будет получено, тем точнее будет результат последующего глобального уравнивания параметров камер.

В рамках диссертации предлагается новая двухуровневая схема сопоставления изображений, позволяющая достигать более надежных результатов в случае малого (менее 20%) перекрытия снимков. Предлагается разбить процесс сопоставления на 2 этапа:

1. Поиск области перекрытия путем оценки более простой, чем искомая, модели преобразования, а именно модели сдвига-поворота.
2. Нахождение искомой модели путем сопоставления изображений только по области перекрытия.

Таким образом, исходная сложная задача представляется в виде композиции двух более простых: сначала при большом количестве выбросов оценивается простая модель, затем при малой доле выбросов оценивается искомая сложная модель. Упрощение модели на первом этапе позволяет вместо алгоритмов, основанных на случайных выборках, использовать для оценки ее параметров процедуру голосования Хафа, дающую детерминированный результат и более устойчивую к высокой доле выбросов.

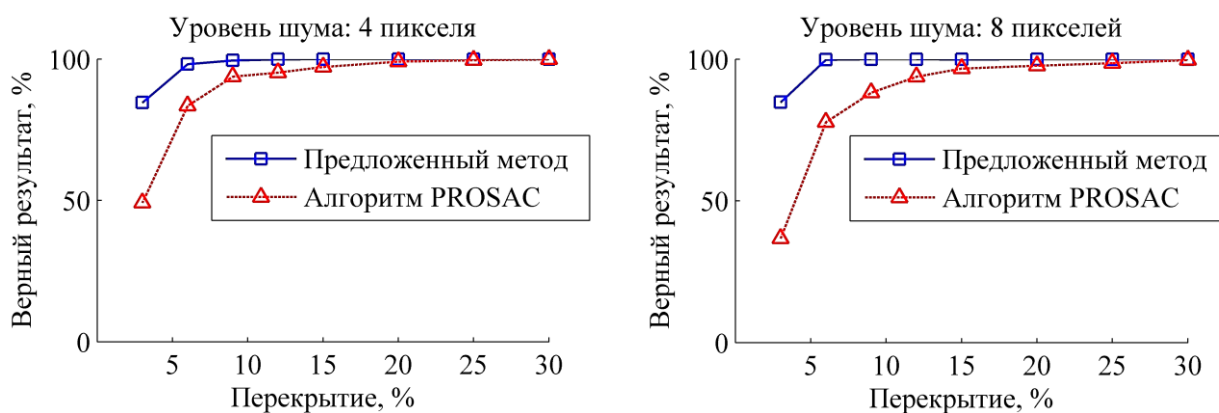
Поиск упрощенной модели осуществляется в два шага: сначала необходимо компенсировать поворот между изображениями, после чего становится возможным определить сдвиг. Пусть  $M$  - набор начальных соответствий и  $m_1 \in M, m_2 \in M$  - две определенные пары точек из этого набора. Положим  $m_1 = \{P_1, P_1'\}$  и  $m_2 = \{P_2, P_2'\}$ , где  $P_1, P_2$  и  $P_1', P_2'$  - точки на изображениях  $I$  и  $I'$  соответственно. Тогда вектор  $\overline{P_1 P_2}$  на изображении  $I$  соответствует вектору  $\overline{P_1' P_2'}$  на изображении  $I'$ , и, исходя из этого, может быть вычислен угол поворота. Пара соответствий  $m_1, m_2$  вносит свой голос за данный угол. Голоса накапливаются в аккумуляторе голосов. Значение угла, для которого достигается максимум голосов, выбирается в качестве относительного угла поворота снимков. Отношение величины первого максимума к величине второго максимума может

быть использовано в качестве меры надежности решения. Аналогичным образом происходит голосование за параметры сдвига, с той лишь разницей, что свой голос вносит каждое отдельное соответствие  $m_i \in M$ , а аккумулятор голосов является двумерным.

В рамках работы была проведена экспериментальная оценка предложенного алгоритма как на синтетических, так и на реальных данных. Пусть  $\tilde{\Psi} = (\tilde{X}, \tilde{Y}, \tilde{\theta})$  - искомые параметры модели сдвига-поворота, а  $\Psi = (X, Y, \theta)$  - найденное решение. Решение считается верным, если выполняется условие:

$$\sqrt{\left(\frac{\tilde{X} - X}{W}\right)^2 + \left(\frac{\tilde{Y} - Y}{H}\right)^2 + \left(\frac{|\theta - \tilde{\theta} - \pi| - \pi}{2\pi}\right)^2} \leq \tau,$$

где  $W$  и  $H$  - ширина и высота изображения соответственно,  $\tau$  - параметр, задающий требуемый уровень точности решения. На Рисунке 2 представлены результаты тестирования алгоритма в сравнении с алгоритмом PROSAC при  $\tau = 0.04$ . Стоит отметить, что для моделирования погрешностей локализации характерных точек в координаты правильных соответствий вносился гауссов шум амплитудой 4 и 8 пикселей.



**Рисунок 2.** Экспериментальная оценка предложенного метода и алгоритма PROSAC на синтетических данных.

Из графиков на Рисунке 2 видно, что предложенный метод позволяет добиться более надежного сопоставления изображений в случае, когда перекрытие снимков не превышает 20%.

Алгоритм был также протестирован на реальных данных, в качестве которых использовались данные аэрофотосъемки. Для них не было известно точных параметров преобразования, поэтому корректность решения проверялась вручную. Предложенный алгоритм показал 71% успешных сопоставлений против 54% у PROSAC.

**Вторая глава** посвящена задаче плотной стереорекострукции по паре изображений, или бинокулярного стерео. Два изображения сцены  $I_1, I_2$  называются *ректифицированными*, если они выровнены по строкам, т.е. для любой пары соответствующих точек  $p_1 \in I_1, p_2 \in I_2$  верно утверждение:  $\exists d : p_1 = (x, y) \Rightarrow p_2 = (x - d, y)$ . Величина смещения  $d$  называется *диспаратетом* в точке  $p_1$ . Матрица  $D_1$ , каждый элемент которой содержит диспаратет соответствующего пикселя изображения  $I_1$ , называется *картой диспаратета* для изображения  $I_1$ . Задача плотной стереорекострукции по паре изображений формулируется следующим образом: имея ректифицированные изображения  $I_1, I_2$ , найти карту диспаратета  $D_1$  для одного из них. Пример ректифицированных изображений и эталонной карты диспаратета приведен на Рисунке 3.



**Рисунок 3.** Пример ректифицированных изображений и эталонной карты диспаратета.

Бинокулярное стерео является давно известной и хорошо изученной задачей компьютерного зрения. Дальнейшее развитие методов во многом зависит от появления новых, отвечающих практическим задачам тестовых баз. На протяжении последних 10 лет основной и фактически единственной тестовой базой был набор Middlebury<sup>10</sup>. Однако его недостатком является малое количество тестовых пар (рейтинг алгоритмов составляется по результатам на всего 4 стереопарах), а также оторванный от практики характер изображений, полученных в лабораторных условиях. Поэтому многие методы, показывающие хорошие результаты на этой базе, оказываются неприменимы на практике. По этой причине в качестве тестовой базы в рамках данной работы используется набор KITTI (Karlsruhe Institute of Technology and Toyota Technological Institute)<sup>11</sup>, состоящий из нескольких сотен пар изображений наземной городской съемки. Для подобных данных,

<sup>10</sup> <http://vision.middlebury.edu/stereo/>

<sup>11</sup> [http://www.cvlibs.net/datasets/kitti/eval\\_stereo\\_flow.php?benchmark=stereo](http://www.cvlibs.net/datasets/kitti/eval_stereo_flow.php?benchmark=stereo)

как и для большинства других сцен антропогенной природы, характерно наличие большого числа плоских областей, не параллельных плоскости камеры и обладающих слабо выраженной текстурой (Рисунок 4).



**Рисунок 4.** Примеры сцен с большим числом наклонных плоскостей.

Существующие алгоритмы, работающие в пространстве диспаратетов и не оперирующие плоскостями в явном виде, обладают характерным недостатком: они разбивают наклонные плоскости на множество фронтально-параллельных плоскостей, что сильно искажает сцену при последующей реконструкции. Для решения данной проблемы уже предлагался ряд алгоритмов, отличающихся друг от друга по двум параметрам: каким образом вычисляются плоскости, и как они затем присваиваются пикселям. Целая группа методов основана на идее цветовой сегментации изображений и вписывания плоскостей в найденные сегменты. Однако цветовая сегментация снижает вычислительную эффективность подобных алгоритмов, а также накладывает жесткое ограничение: каждый сегмент должен аппроксимироваться только одной плоскостью, что не всегда верно. Несколько другой подход используется в алгоритме PatchMatch Stereo<sup>12</sup>: каждый пиксель инициализируется случайной плоскостью, после чего производится итерационное распространение плоскостей по изображению и их уточнение. Но недостатком такого алгоритма является крайне низкая скорость работы: в авторской реализации обработка изображения разрешением 0.1 мегапикселя занимала около минуты.

В данной диссертационной работе предлагается новый алгоритм плотной стереорекострукции по паре изображений, работающий в пространстве плоскостей и состоящий из следующих шагов:

1. Нахождение контрольных точек путем сопоставления изображений.
2. Вычисление набора плоскостей путем сегментации облака трехмерных точек.
3. Выбор оптимальной плоскости в каждом пикселе путем агрегации штрафов по минимальному покрывающему дереву.

---

<sup>12</sup> Bleyer M., Rhemann C., Rother C. PatchMatch Stereo - Stereo Matching with Slanted Support Windows // Proceedings of the British Machine Vision Conference (BMVC). 2011. P 14.1-14.11.

На первом шаге производится сопоставление ректифицированных изображений, результатом которого является набор *контрольных точек* (ground control points, GCP), т.е. точек, значение диспаратета в которых известно с высокой степенью уверенности. Стоит отметить, что в ряде практических сценариев множество контрольных точек может быть дано на входе алгоритма: например, из этапа сопоставления изображений, предшествующего ректификации снимков, либо из результатов лазерного сканирования, если они имеются.

Множество контрольных точек *GCP* фактически представляет собой облако трехмерных точек в пространстве  $\{x, y, d\}$ . Поэтому для вычисления набора плоскостей  $F$  на втором шаге предлагается использовать алгоритм сегментации облака трехмерных точек, состоящий из следующих шагов:

- a. Вычисление алгоритмом RANSAC параметров плоскости, поддерживаемой наибольшим числом точек из *GCP*. Добавление плоскости в множество  $F$ .
- b. Исключение поддерживающих плоскость точек из множества *GCP*.
- c. Повторение шагов b-c пока существуют плоскости, поддерживаемые достаточным числом точек.

На третьем шаге каждому пикселю  $p$  изображения присваивается одна из плоскостей  $f \in F$ , при этом значение диспаратета получается из выражения  $d_p^f = a_f x + b_f y + c_f$ . Выбирается та плоскость, при которой достигает минимума следующий функционал:

$$E_p(f) = E_p^{Agg}(f) + \lambda \cdot E_p^{GCP}(f)$$

Здесь величина  $E_p^{Agg}(f)$  отражает стоимость присвоения плоскости  $f$  точке  $p$ , агрегированную по всем пикселям изображения, а  $E_p^{GCP}(f)$  представляет собой штраф за отклонение от регуляризирующего решения  $D^{GCP}$ , полученного с использованием только контрольных точек.

Для вычисления стоимости  $E_p^{Agg}(f)$  используется подход, предложенный в одной из недавних работ<sup>13</sup>. Изображение представляется в виде графа, вершины которого соответствуют пикселям и соединены ребрами по принципу 4-хсвязности, при этом вес ребра равен абсолютному значению разности яркостей соединяемых им пикселей:

---

<sup>13</sup> Yang Q. A Non-Local Cost Aggregation Method for Stereo Matching // IEEE Conference on Computer Vision and Pattern Recognition. 2012. P 1402-1409.

$w_{rs} = |I_r - I_s|$ . Для этого графа строится *минимальное покрывающее дерево*, т.е. такое дерево, которое включает в себя все вершины исходного графа, и при этом сумма его ребер минимальна. Пусть  $D_{pq}$  – сумма весов ребер вдоль кратчайшего пути в дереве, соединяющего вершины  $p$  и  $q$ , а  $C_q(d)$  – штраф за диспаратет  $d$  в точке  $q$ , зависящий от яркостей и градиентов в соответствующих пикселях левого и правого изображений.

Тогда  $E_p^{Agg}(f) = \sum_{q \in I} \exp(-\frac{D_{pq}}{\sigma}) C_q(d_q^f)$ , где  $\sigma$  – параметр алгоритма, контролирующий степень влияния пикселей друг на друга. Стоит подчеркнуть, что в отличие от алгоритма Янга, в предлагаемом методе стоимость агрегируется не для каждого значения диспаратета, а для каждой плоскости из набора  $F$ .

Выражение штрафа за отклонение от регуляризирующего решения  $D^{GCP}$  имеет вид

$$E_p^{GCP}(f) = -\ln((1-\eta) \exp(-\frac{|d_p^f - d_p^{GCP}|}{\gamma}) + \eta), \text{ где } \eta \text{ и } \gamma - \text{параметры функции, а } d_p^{GCP} -$$

диспаратет в точке  $p$  согласно  $D^{GCP}$ . Для вычисления  $D^{GCP}$  применяется описанная выше процедура агрегации стоимостей плоскостей по минимальному покрывающему

дереву, но с модифицированной функцией  $C_q(d): C_q^{GCP}(d) = \begin{cases} (d - d_q^{GCP})^2, & q \in GCP \\ 0, & q \notin GCP \end{cases}$ , где

$d_q^{GCP}$  – значение диспаратета в контрольной точке  $q$ .

В работе приведены результаты тестирования предложенного алгоритма на 194 парах изображений из набора КИТТИ, для которых были доступны эталонные карты диспаратета. В качестве меры ошибки использовался процент пикселей, для которых разница между найденным и эталонным значением диспаратета превысила порог  $\tau$  при значении  $\tau = 3$ . Алгоритм показал заметное улучшение результатов относительно алгоритма Янга, сопоставимого с предлагаемым методом по скорости работы: средняя ошибка по тестовой выборке составила 18.6% против 33.03% у метода Янга.

**В третьей главе** диссертации исследуется задача регистрации изображений относительно облака трехмерных точек. Пусть имеется набор из  $k$  изображений и облако трехмерных точек  $T$ . Задача в общем виде формулируется как поиск уточненных значений параметров внешнего ориентирования камер  $\tilde{C}^1, \tilde{C}^2, \dots, \tilde{C}^k$  по начальным значениям  $C_0^1, C_0^2, \dots, C_0^k$ , где параметры внешнего ориентирования  $C = (x, y, z, \omega, \phi, \kappa)$

включают в себя координаты центра проекции и углы поворота камеры в мировой системе координат.

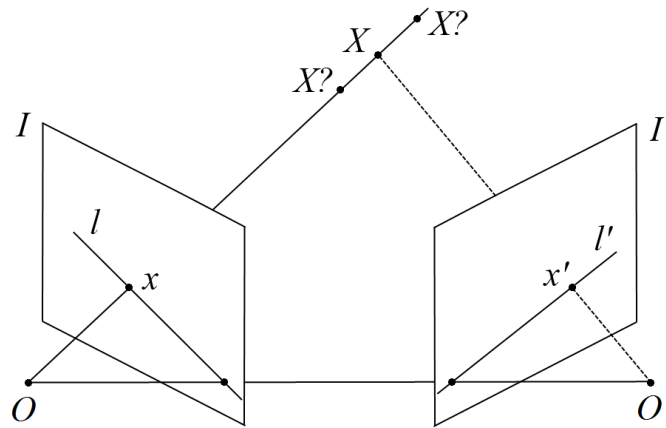
Все существующие алгоритмы регистрации изображений относительно облаков трехмерных точек можно разбить на два класса. К первому относятся методы, восстанавливающие разреженное облако точек по изображениям и находящие такие параметры камер, при которых оно наиболее совпадает с облаком точек, полученным с лазерного сканера. Недостатком подобных методов является вычислительная сложность и дополнительные погрешности, обусловленные ошибками стереорекострукции. Методы второго класса, напротив, получают плоские представления имеющегося облака трехмерных точек в виде карт глубины и ищут параметры для их оптимального сопоставления с изображениями. К недостаткам подобных алгоритмов стоит отнести потерю части информации при переходе от облака трехмерных точек к картам глубины. Тем не менее, проведенный обзор литературы показывает их преимущество над алгоритмами первой группы на сегодняшний день.

Одним из наиболее эффективных алгоритмов второй группы является алгоритм Мастина<sup>14</sup>. Его идея заключается в поиске оптимальных параметров камеры путем минимизации функции  $q_{MI}(C) = -m(I, C)$ , где  $m(I, C)$  – величина взаимной информации между изображением  $I$  и картой глубины, сгенерированной по облаку трехмерных точек при параметрах камеры  $C$ . Для оптимизации применяется метод симплексного спуска. Ключевым недостатком алгоритма Мастина является работа лишь с одним изображением: в контексте общей задачи это означает поочередную оптимизацию параметров для каждого снимка, тогда как одновременный учет нескольких изображений и геометрических связей между ними могли бы повысить надежность результатов.

В диссертационной работе предлагается новый алгоритм регистрации изображений относительно облака трехмерных точек, основанный на ряде улучшений базового подхода Мастина. Пусть имеются два перекрывающихся снимка,  $I$  и  $I'$ , центры проекций которых находятся в точках  $O$  и  $O'$  соответственно, а  $F$  – фундаментальная матрица. Известно, что пара точек  $x$  и  $x'$ , являющихся изображением одной и той же точки  $X$  пространства, лежат на эпиполярных линиях  $l' = F \cdot x$  и  $l = x'^T \cdot F$  соответственно (см. Рисунок 5).

---

<sup>14</sup> [11] Mastin A., Kepner J., Fisher J. Automatic Registration of LIDAR and Optical Images of Urban Scenes // IEEE Conference on Computer Vision and Pattern Recognition. 2009. P 2639-2646.



**Рисунок 5.** Принципы эпиларной геометрии

Однако в реальных задачах, когда неизбежны погрешности в локализации соответствующих точек на изображении и в вычислении фундаментальной матрицы, принято вводить метрику ошибки, основанную на расстоянии от соответствующей точки до эпиларной линии  $d(l, x)$ . Пусть  $P = \{(x_1, x_1'), (x_2, x_2'), \dots, (x_N, x_N')\}$  – набор пар соответствующих точек для данной пары изображений. Определим ошибку на  $P$  как среднее расстояние от каждой точки до эпиларной линии, на которой она должна лежать:

$$e(F, P) = \frac{1}{N} \sum_{i=1}^N (d(x_i^{T \cdot F}, x_i) + d(F \cdot x_i, x_i'))$$

Расширим функционал из алгоритма Мастина, включив в него величину ошибки  $e(F, P)$  и функцию взаимной информации для второго снимка:  $q(C, C') = \psi(I, C) + \psi(I', C') + \lambda \cdot e(f(C, C'), p(I, I'))$ , где  $\lambda$  – параметр, контролирующий влияние эпиларных ограничений. При  $\lambda = 0$  получаем независимую оптимизацию по каждому из изображений аналогично базовому методу.

Также предлагается модификация функции взаимной информации, призванная повысить устойчивость к локальным изменениям яркости на изображениях и перепадам высот на картах глубины. Разобьем изображение и карту глубины на множество непересекающихся областей  $\{S_1, S_2, \dots, S_N\}$  и усредним значение взаимной информации, посчитанное по каждой из областей:

$$\bar{m}(I, C) = -\frac{1}{N} \sum_{k=1}^N m(I_{S_k}; D_{S_k}^C),$$

где  $I_{S_k}$  и  $D_{S_k}^C$  – изображения области  $S_k$  снимка и карты высот соответственно.



В работе представлены результаты тестирования базового алгоритма и предложенных улучшений на тестовом наборе Vaihingen, предоставленном Немецким Обществом Фотограмметрии, Дистанционного Зондирования и Геоинформатики (DGPF)<sup>15</sup>. В Таблице 1 приведена доля успешных регистраций для нескольких пар тестовых изображений.

Алгоритм	Тестовые данные				
	1	2	3	4	среднее
Базовый алгоритм	22%	8%	12%	11%	13.25%
Базовый алгоритм + усредненная взаимная информация	21%	26%	29%	32%	27.00%
Базовый алгоритм + усредненная взаимная информация + эпиполярные ограничения (параметры одной из камер фиксированы)	<b>48%</b>	<b>43%</b>	<b>47%</b>	<b>45%</b>	<b>45.75%</b>
Базовый алгоритм + усредненная взаимная информация + эпиполярные ограничения	25%	25%	26%	25%	25.25%

**Таблица 1.** Результаты работы базового алгоритма и предложенных улучшений

Регистрация считается успешной, если для каждого параметра  $\tau \in \{x, y, z, \omega, \varphi, \kappa\}$  выполняется условие  $|\tilde{\tau} - \tau^*| < \mu \cdot A_\tau$ , где  $\tilde{\tau}$  и  $\tau^*$  – найденное и эталонное значения параметра соответственно, а  $A_\tau$  – максимально возможное начальное отклонение параметра от правильного значения. Приведенные результаты соответствуют значениям  $A_x = 15$  м,  $A_y = 15$  м,  $A_z = 15$  м,  $A_\omega = 0.5^\circ$ ,  $A_\varphi = 0.5^\circ$ ,  $A_\kappa = 0.5^\circ$  и  $\mu = 0.4$ .

## Основные результаты работы

1. Предложена новая двухэтапная схема сопоставления изображений, основанная на процедуре голосования Хафа. Алгоритм превосходит существующие подходы в случае малого (до 20%) перекрытия снимков.

---

<sup>15</sup> Cramer M. The DGPF Test on Digital Aerial Camera Evaluation – Overview and Test Design // Photogrammetrie Fernerkundung Geoinformation. 2010. 2. P 73–82.

2. Разработан новый алгоритм плотной стереорекострукции по паре изображений. По сравнению с существующими методами, сравнимыми с ним по скорости работы, предложенный алгоритм позволяет получать более точные карты диспаратета на сценах антропогенной природы, для которых характерно наличие большого числа слабо текстурированных плоских областей.
3. Предложен новый алгоритм регистрации изображений относительно облака трехмерных точек, учитывающий не только точность наложения изображения и карты глубины, но и эпиполярные ограничения между парой снимков.

В контексте общей задачи трехмерной реконструкции, разработанные алгоритмы позволяют расширить класс данных, на которых возможно полностью автоматическое получение результата, а также позволяют достичь большей точности при совместном использовании изображений и облаков трехмерных точек.

### **Список публикаций**

1. Mizotin M., Krivovyaz G., Velizhev A., Chernyavskiy A., Sechin A. Robust matching of aerial images with low overlap // International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. 2010. 38. N 3A. P 13-18.
2. Кривовязь Г., Мизотин М. Алгоритм автоматического совмещения аэрофотоснимков с малым перекрытием // Сборник тезисов XVII Международной научной конференции студентов, аспирантов и молодых ученых «Ломоносов-2010». 2010. С. 60-61.
3. Кривовязь Г., Мизотин М. Надежное сопоставление аэрофотоснимков с малым перекрытием // Информационные технологии в проектировании и производстве. 2011. Т. 3. С. 92-98.
4. Кривовязь Г., Черников А., Велижев А. Автоматическое сопоставление изображений и облака трехмерных точек // Труды конференции Графикон. 2011. С. 208-211.
5. Кривовязь Г., Птенцов С., Конушин А. Алгоритм плотной стереорекострукции на основе контрольных точек и разметки плоскостями // Программные продукты и системы. 2012. Т. 4. С. 236-241.
6. Кривовязь Г., Велижев А. Алгоритм уточнения параметров внешнего ориентирования камеры по изображениям и данным лазерного сканирования // Известия высших учебных заведений. Геодезия и аэрофотосъемка. 2013. Т. 1. (в печати)