

© 2025 г. А.В. ГУДЫМ (anton.v.gudym@yandex.ru),
А.П. СОКОЛОВ, д-р техн. наук (alsokolo@bmstu.ru)
(Московский государственный технический университет им. Н.Э. Баумана)

ОБЗОР МЕТОДОВ ПРОИЗВОДСТВЕННОЙ ФОТОГРАММЕТРИИ ПОВЫШЕННОЙ ТОЧНОСТИ¹

Настоящий обзор посвящен новым и классическим методам, составляющим современный фотограмметрический конвейер, применяемый для эффективного высокоточного восстановления 3D-координат облаков точек и позиций объектов по фото- или видеосигналу. Уделяется особое внимание факторам, оказывающим влияние на измерительные погрешности выходной 3D-реконструкции. В зависимости от приложения реконструируемым 3D-точкам могут соответствовать различные признаки, такие как контрастные особенности текстуры объекта, рельефа или специальные метки, нанесенные на поверхность объекта. После выделения и сопоставления признаков следует решение задачи оптимизации пучка проекционных лучей (от англ. «bundle adjustment») для восстановления 3D-координат точек в пространстве. В обзоре освещены актуальные работы по данному направлению, приводятся удобные и практически значимые формулировки различных моделей камер, учитывающие дисторсию и используемые в рамках задачи оптимизации пучка. В экспериментальной части демонстрируется уровень точности, который может быть достигнут на практике с помощью рассмотренных методов для близко-ракурсных измерений. Показано, что повторяемость получаемых 3D-координат точек может превосходить уровень профессиональной фотограмметрии.

Ключевые слова: точная фотограмметрия, визуальная одометрия, компьютерное зрение, сопоставление признаков, оптимизация пучка проекционных лучей.

DOI: 10.7868/S2413977725120013

1. Введение

Фотограмметрия – наука об измерениях по фотографиям имеет давнюю историю, активно развивалась как в России, так и в мире с конца XIX в. Термин «*фотограмметрия*» в 1867 г. ввел немецкий архитектор Альбрехт Мейденбауэр (1834–1921), а ранее в 1858 г. опубликовал свой фотограмметрический метод измерения зданий. Математические основы фотограмметрии заложил немецкий математик С. Финстервальдер (Sebastian Finsterwalder, 1862–1951) [1]. Большой вклад в развитие математических моделей проекции, применяемых по сей день, внес американский исследователь Дуэйн Браун [2–4].

¹ Работа выполнена в рамках госзадания Минобрнауки России (тема № FSN-2024-0083).



Рис. 1. Слева полигональная сетка (3D-модель) поверхности ландшафта, полученная в результате фотограмметрии. Справа участок исходного изображения, где окружности – позиции особых точек или SIFT-признаков для сопоставления и триангуляции в оптимизационной задаче фотограмметрии.

Основоположниками фотограмметрии в России считаются следующие выдающиеся ученые и инженеры, специалисты кафедры фотограмметрии Московского государственного университета геодезии и картографии, внесшие значительный вклад в развитие технологии: профессора Н.М. Алексапольский (1890–1942), Ф.В. Дробышев [5], А.Н. Лобанов, Л.Н. Васильев, В.Б. Дубиновский [6] и многие другие.

Одной из первых отечественных стереофотограмметрических систем стала фотограмметрическая система «Талка», которая была разработана советским инженером Д.В. Тюкавкиным в 1960-х гг. [7].

Исторически, фотограмметрия применялась в области геодезии и картографии (рис. 1), тем не менее в течение последних двух десятилетий, учитывая стремительный рост вычислительных мощностей компьютерных систем и возможностей методов обработки изображений в том числе с помощью искусственного интеллекта (ИИ), фотограмметрия, как измерительная технология, получила широкое распространение в промышленности (*производственная* близко-ракурсная или ближняя *фотограмметрия*, англ. close-range photogrammetry). Использование методов фотограмметрии, в первую очередь, обеспечивает восстановление или реконструкцию модели 3D-поверхности интересующего объекта по набору фотографических изображений. Кроме того, технология имеет широкий спектр применения в современных производственных процессах [8–11]:

- 1) контактные технологии измерения формы поверхности через позиционирование (трекинг) зонда/щупа в режиме координатно-измерительной машины (КИМ);
- 2) бесконтактное измерение формы поверхности при трекинге лазерного 3D-сканера;
- 3) контроль положения крупногабаритных деталей в аэрокосмической отрасли;
- 4) контроль прецизионной обработки деталей в машиностроении;

- 5) трекинг звеньев робота-манипулятора и установленного инструмента [12–15].

Производственная фотограмметрия предполагает применение многочисленных алгоритмов из области компьютерного зрения и анализа изображений (распознавание паттернов, 3D-моделирование). Известным исследователем в области задач распознавания и сопоставления признаков на фотографических изображениях является Ю.В. Визильтер [16].

Предметом настоящей работы является описание методов, составляющих современный фотограмметрический конвейер, применяемый в производственной фотограмметрии. Рассматриваемые в работе методы обеспечивают эффективное высокоточное восстановление 3D-координат точек (далее 3D-точек), сопоставляемых тем или иным характерным признакам, идентифицируемым на поверхности наблюдаемых объектов по фото- или видеосигналу. Построенное облако 3D-точек используется для реконструкции модели 3D-поверхности объекта.

Известные альтернативные технологии построения 3D-поверхности включают:

- 1) проекционные 3D-сканеры со структурной подсветкой или лазерными линиями;
- 2) LIDAR-сканеры, измеряющие задержку времени хода лазерного луча;
- 3) гибридные решения – применение структурной подсветки в сочетании с фотограмметрией.

Преимущество альтернативных решений в том, что они не требуют наличия контрастных текстурных признаков на поверхности объекта. В то же время по соотношению величины измерительной погрешности к габариту объекта, фотограмметрия может уступать лишь LIDAR-системам [9] (при очень высокой цене последних), оставаясь эффективным инструментом для проведения измерений в широком диапазоне размеров. Например, одним и тем же набором инструментов можно измерять объекты с габаритными размерами от 0,1 до 10 м. Погрешность может составлять 1 : 100 000 [17] или даже 1 : 200 000 [8] к габаритной величине объекта, т.е. десять или пять микрометров на 1 м соответственно.

За короткий срок фотограмметрия стала стандартным инструментом эффективного (точного и быстрого) контроля качества, например, в аэрокосмической отрасли производства. Нередко фотограмметрию применяют совместно с проекционными 3D-сканерами для позиционирования устройства и последующего слияния точек в выходную полигональную сетку высокого разрешения.

Фотограмметрия является ключевым инструментом аккуратной навигации по видеосигналу и источником геодезических измерений для беспилотных летательных аппаратов (БПЛА) [14, 18–20]. Пример построения 3D-поверхности ландшафта с использованием фотограмметрического конвейера [21] с данными аэрофотосъемки от БПЛА [22] представлен на рис. 1.

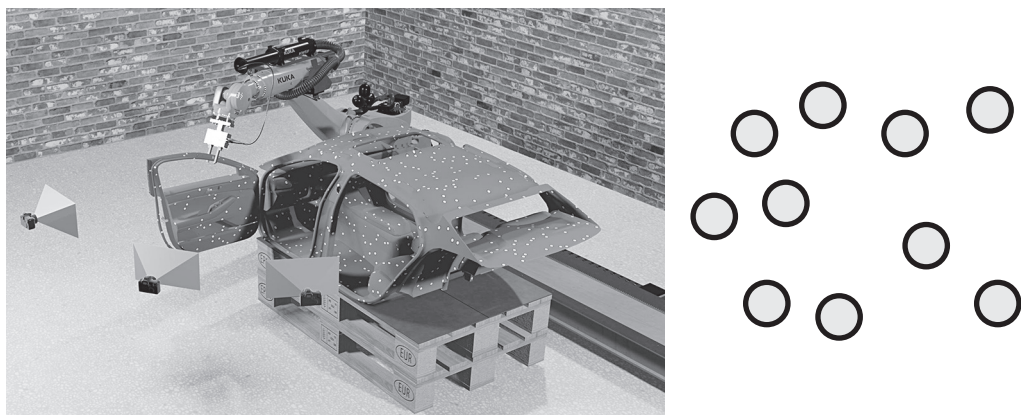


Рис. 2. Слева пример сценария для позиционирования выходного звена робота по сигналу нескольких стационарных камер. В качестве признака для 3D-реконструкции используются круглые черно-белые метки (справа).

Справа на изображении представлены позиции текстурных SIFT-признаков, представляющих собой совокупность 2D-координат и вектора вещественных чисел (дескриптора), который характеризует свойства текстуры объекта на изображении в небольшой окрестности признака.

Известным преимуществом фотограмметрии относительно альтернативных способов 3D-реконструкции является ее универсальность и масштабируемость – те же алгоритмы и модели камер (раздел 6.1) применимы как для геодезических измерений, так и в совершенно иных условиях съемки, когда расстояние между объектом и камерой не превышает нескольких метров. На рис. 2 найденные 3D-координаты меток на поверхности объектов используются для оценки взаимного положения тел, например, позиции выходного звена робота-манипулятора относительно объекта привязки, положение которого также реконструируется на основе идентифицируемых положений размещенных на этом объекте меток.

В работе особое внимание уделяется качественному процессу выделения текстурных признаков на примере искусственных круглых меток, так как это оказывает решающее значение на точность получаемых 3D-координат. Факторы, определяющие этот процесс, также актуальны и для естественных текстурных особенностей в таких известных алгоритмах, как: SIFT [23], SURF [24], SuperPoint [25] и др.

2. Структура обзора

Любая фотограмметрическая система представляет собой программно-аппаратный комплекс. Пример аппаратной части фотограмметрической системы представлен в разделе 4. Программная часть является реализацией алгоритмов, предназначенных для решения задач следующих двух типов, соответствующих этапам фотограмметрического конвейера:

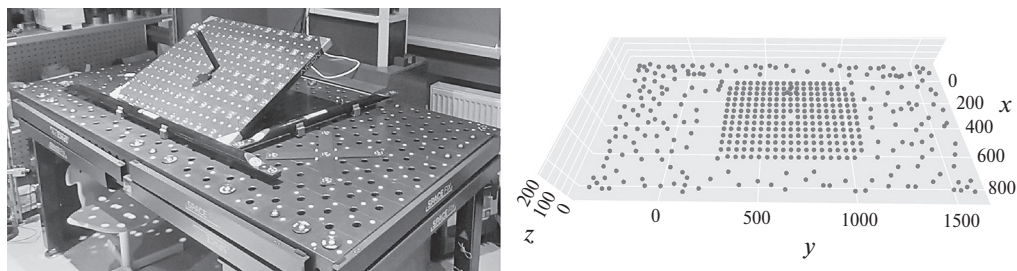


Рис. 3. Тестовая сцена (слева) включает два измеряемых объекта с нанесенными метками: стальной сварочный стол ~ 2000 мм, углепластиковая калибровочная плита ~ 800 мм, а также два углепластиковых масштабных жезла. Результат фотограмметрии (справа) – высокоточная реконструкция меток в виде облака 3D-точек.

- 1) обработка входных изображений (раздел 5);
 - выделение или детектирование максимального числа особых точек на поверхности наблюдаемого объекта (шаг 1);
 - построение дескриптора, обладающего инвариантностью к аффинным преобразованиям метки и оптическим искажениям сигнала (шаг 2);
 - сопоставление дескрипторов (шаг 3);
 - фильтрация найденных соответствий (шаг 4);
- 2) решение задачи оптимизации пучка для проекционных лучей по найденным соответствиям или просто решение задачи оптимизации пучка (раздел 6);
 - определение начальных приближений параметров (координаты наблюдаемых признаков, внутренние параметры камер и их позиции) или просто автокалибровка (шаг 5);
 - решение задачи оптимизации (шаг 6).

Этап обработки данных (раздел 5) сильно зависит от сценария применения. Например, для измерения на производстве на объект наносятся простые круглые или кодовые метки (рис. 3). При этом в аэро- или спутниковой фотосъемке детектируют естественные контрастные особенности на местности. Этап оптимизации, в отличие от обработки входных данных, достаточно универсален и применим практически к любому сценарию работы.

Раздел 7 посвящен результатам экспериментов с различными факторами задачи оптимизации пучка. Также демонстрируется высокий уровень точности, не уступающий профессиональным фотограмметрическим продуктам.

3. Математическая нотация

Векторные величины обозначаются жирным шрифтом, представлены вектор-столбцами скалярных величин, например $\mathbf{p} = [p_1, \dots, p_n]^T \in \mathbb{R}^n$. Для обозначения однородных координат, применяемых в проективной геометрии и являющихся расширением понятия декартовых координат, используется

верхний индекс h , например: $\mathbf{p}^h = [p_1^h, \dots, p_n^h, 1/\lambda]^T \in \mathbb{R}^{n+1}$, $\lambda \neq 0$. Для одно-родных координат верно: $\lambda p_i^h = p_i$, где $i \in [1..n]$, что удобно для лаконичной формулировки различных матричных преобразований.

Линейные операторы преобразования систем координат (СК), как правило, содержат обозначение откуда (индекс **справа внизу**) и куда (индекс **слева вверх**) происходит переход. Например, для описания положения СК объекта (индекс o , object) относительно СК камеры (индекс c , camera) используется матрица ${}^cT_o \in \mathbf{SE3} \subset \mathbb{R}^{4 \times 4}$, $\det({}^cT_o) \neq 0$ (шесть степеней свободы):

$$(1) \quad {}^cT_o \cdot \mathbf{p}_o^h = \begin{bmatrix} {}^cR_o & {}^c\mathbf{t}_o \\ \mathbf{0} & 1 \end{bmatrix} \begin{bmatrix} x_o \\ y_o \\ z_o \\ 1 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_o \\ y_o \\ z_o \\ 1 \end{bmatrix} = \mathbf{p}_c^h,$$

${}^c\mathbf{t}_o$ – сдвиг центра СК объекта относительно камеры (три степени свободы),

$\mathbf{p}_o^h$ – положение точки относительно объекта,

$\mathbf{p}_c^h$ – положение точки относительно камеры,

${}^cR_o \in \mathbf{SO3} \subset \mathbb{R}^{3 \times 3}$ – матрица вращения (три параметра вращения отображаются в матрицу, например, формулой Родригеса [26] или с помощью углов Эйлера), $\mathbf{SO3}$ – специальная ортогональная группа вращений (${}^cR_o^T = {}^cR_o^{-1}$, $\det({}^cR_o) = 1$),

$\mathbf{SE3}$ – евклидова группа движения [27] или преобразование подобия с единичным масштабом, описывает возможные перемещения тела в пространстве.

Значение математической группы для инженерного приложения: если требуется посчитать изменение или приращение позиции камеры в интервале времени от t_a до t_b , то следует использовать умножение на обратный элемент²:

$${}^bT_o \cdot {}^aT_o^{-1} \cdot \mathbf{p}_a^h = {}^bT_o \cdot {}^oT_a \cdot \mathbf{p}_a^h = {}^bT_a \cdot \mathbf{p}_a^h = \mathbf{p}_b^h,$$

aT_o , bT_o – положение объекта относительно СК камеры в момент времени t_a и t_b соответственно,

bT_a – положение камеры в момент времени t_a относительно СК камеры в момент t_b ,

$\mathbf{p}_a^h$, $\mathbf{p}_b^h$ – положение 3D-точки относительно камеры в момент времени t_a и t_b соответственно.

Для лучшего понимания формулировок следует допустить одновременное существование всех СК, связанных с состояниями движущегося тела, и рассматривать моменты времени как идентификаторы той или иной СК. Из этого также следует универсальность представленных ниже формулировок – индексы a , b могут соответствовать не двум моментам времени, а двум различным камерам в один момент времени. $I \in \mathbf{SE3}$ – единичная диагональная матрица для идентичных СК (неподвижное тело).

² Матрицы-элементы $\mathbf{SE3}$ не следует умножать на скаляр или вычитать – это недопустимая операция над элементом группы.

Определение связей СК в индексах матриц необходимо для ясности постановки оптимизационной задачи фотограмметрии, в частности, для описания модели камеры (подраздел 6.1).

4. Оборудование для промышленной фотограмметрии

На рис. 3 представлена сцена эксперимента для построения облака 3D-точек по меткам (маркерам), наклеенным на измеряемые объекты. Также в сцене присутствуют два масштабных жезла (эталона длины) – объекты с заранее известным расстоянием между маркерами³. Масштабные жезлы позволяют определять абсолютные значения расстояний между точками, идентифицируемыми в сцене. В случае отсутствия масштабных жезлов, по набору изображений, возможно восстановить геометрию сцены или объекта с точностью до масштаба [26]. Жезлы часто выполнены из углепластика, что позволяет достигать малого коэффициента линейного температурного расширения вдоль его оси, а также обеспечивает малый вес и достаточную прочность.

Для наилучших результатов, минимизации шумов измерений и максимального диапазона работы, в промышленных системах часто применяют специальные круглые метки или маркеры из светоотражающего материала. Такое свойство чрезвычайно полезно, чтобы увеличить контрастность или резкость контура маркера. С использованием вспышки для камеры (для подсветки объекта), светоотражающий материал возвращает гораздо больше энергии строго в направлении источника, в сравнении с квадратичной функцией рассеяния энергии света для обычных материалов.

Для получения качественных изображений ключевых точек на поверхности измеряемого объекта может использоваться профессиональная зеркальная фотокамера. Для наилучших результатов требуется черно-белый светочувствительный сенсор и оптика высокого разрешения. К примеру, профессиональный фотограмметрический продукт «Hexagon DPA Pro» от немецкого концерна Hexagon включает камеру Canon EOS 5DS с черно-белым сенсором $w \times h = 8700 \times 5800$ (50 Мп); объектив $f = 28$ мм⁴.

Для достижения хороших результатов измерений специализированная оптика (например, с малой дисторсией) не требуется [3], основные факторы – это, естественно, факторы, влияющие на качество сигнала (контрастность изображения), а также геометрическая стабильность (жесткость) оптической системы камера-сенсор⁵ в процессе сбора данных [17, 28]. На геометрическую стабильность влияет жесткость конструкции и методы крепления объектива на камере, масса-габаритные показатели, способ установки вспышки, а также

³ Геометрия маркеров на жезле измеряется или калибруется заранее, в лабораторных условиях.

⁴ Углы зрения составляют $\Delta\varphi_{hor} = 64^\circ$, $\Delta\varphi_{ver} = 45^\circ$, $\Delta\varphi_{diag} = 73^\circ$; предельное угловое разрешение $\approx 2 \frac{\Delta\varphi_{diag}}{\sqrt{w^2 + h^2}} = 0,014^\circ$. Угловое разрешение удобно для сопоставления сенсоров разного разрешения или технологий, например SfM и LIDAR.

⁵ Методы динамической стабилизации сенсора в данном контексте могут оказаться вредны.

температурный режим работы устройства. Таким образом, для лучших результатов (высокоточных измерений) может потребоваться тщательный подбор оборудования:

- 1) калиброванный углепластиковый масштабный жезл;
- 2) зеркальная фотокамера высокого разрешения, обеспечивающая геометрическую стабильность объектива;
- 3) светоотражающие метки на клеевой основе;
- 4) вычислитель, реализующий этапы фотограмметрического конвейера, описанные в следующих разделах.

5. Поиск и сопоставление признаков

Поиск признаков и их сопоставление – чрезвычайно широкая область исследований [21, 29–33], которая, среди прочих, имеет массу приложений в задачах фотограмметрии, SfM (англ. Structure from Motion [26]), SLAM⁶, дополненной реальности, поиска изображений, анализа контекстной информации. В данной работе делается акцент на возможности применения техник для точных измерений.

Цель данного этапа – выявить максимальное количество связей или соответствий между 3D-точками на объекте (будущими измерениями) и их наблюдениями – 2D-точками на изображениях. Ряд работ [25, 32, 34] позволяет найти соответствия между 2D-точками конкретной пары изображений, следовательно, потребуется дополнительная группировка точек по принадлежности к общей точке поверхности. В фотограмметрии именно соответствие 2D-точки на изображении (или проекции) $(u, v) \in \mathbb{R}^2$ и 3D-точки на объекте $(x, y, z) \in \mathbb{R}^3$ определяет будущую систему уравнений в задаче оптимизации пучка (раздел 6.5). На рис. 4 оттенком представлены соответствия, найденные путем сопоставления двоичных кодовых меток, нанесенных на объект.

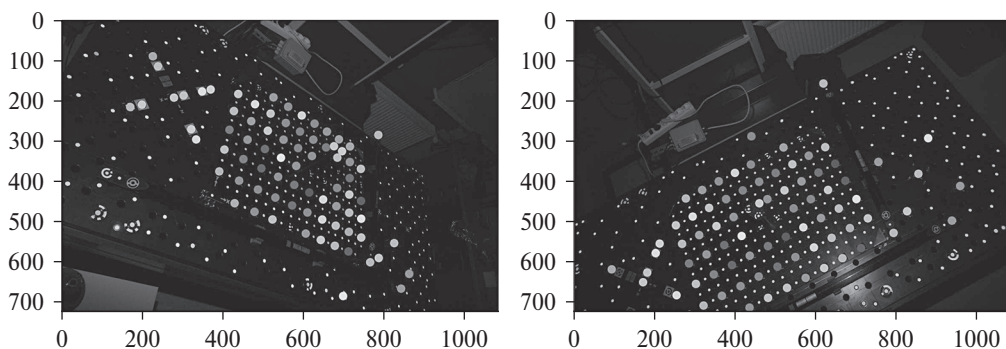


Рис. 4. Исходные изображения с результатом сопоставления 2D-точек по двоичному дескриптору. Оттенок обозначает группу или определенную 3D-точку на объекте.

⁶ SLAM (англ. simultaneous localization and mapping) – одновременная локализация и построение карты, метод навигации в мобильных автономных системах.

5.1. Детектор признаков

Первым шагом фотограмметрического конвейера (стр. 6) является выделение или детектирование максимального числа N_{pts}^a особых точек $\mathcal{P}_a = \{\mathbf{p}_a^k\}_{k=1}^{N_{pts}^a}$, $\mathbf{p}_a^k = [u_a^k, v_a^k]^T \in \mathbb{R}^2$ на поверхности измеряемого объекта, наблюдаемого на изображении $a \in [1 \dots N_{im}]$ от камеры⁷. Данный шаг выполняет алгоритм детекции или просто детектор.

В фотограмметрии общего назначения [21], без искусственных меток, на текстуре объекта выделяют разнообразные естественные особенности – точки, окружности, углы, кресты или подобные структуры с помощью детекторов Харриса [35], GFTT [36], FAST [37]. Некоторые детекторы выделяют и точку, и обобщенную информацию о текстуре окрестности в виде многомерного вектора-дескриптора. Координаты точки вместе с дескриптором часто называют признаком (от англ. «feature»). Такую задачу решают классические подходы SIFT [23], SURF [24] и опирающиеся на машинное обучение SuperPoint [25], DISK [38]. В этом случае выявляются специальные области, наиболее подходящие для последующего определения устойчивого дескриптора. Как правило, естественные признаки не обладают достаточной контрастностью или размером для точной 2D-локализации и последующей 3D-реконструкции. Увидеть это можно на примере, так как в фотограмметрии минимизируется разность наблюдаемых 2D-координат признаков и проекций на изображении соответствующих (искомых) 3D-точек объекта. Данную разность принято называть ошибкой репроекции (от англ. «reprojection error») [26, 39]. Для высокоточной измерительной задачи одно стандартное отклонение ошибки репроекции не превышает 0,1 пикселя (рис. 11), следовательно, нужно выделять признак еще точнее. Шум определения координат признака (детекции) для естественных текстурных особенностей нередко превышает один пиксель [25, 32, 40]. При аккуратной калибровке в [41], но используя уголки, ошибка репроекции составляет $\sim 0,33$ пикселя. Использование глубокого машинного обучения для детектирования текстурных признаков в [42] дает $\sim 0,5$ пикселя. Поэтому для фотограмметрических измерений в промышленности на поверхность объекта наносят черно-белые метки круглой и, реже, квадратной формы.

Круглая метка предпочтительна относительно уголка и подобных структур. Чтобы разобраться в причине, рассмотрим основные факторы, влияющие на устойчивость локализации, для наглядности выделим две категории:

- 1) **геометрические** факторы определяют форму 2D-признака (эффекты перспективной проекции, дисторсия объектива или кривизна поверхности), рис. 5;
- 2) **оптические** факторы влияют на контрастность сигнала, соотношение сигнал-шум⁸ (уровень освещенности и отражательные свойства поверхности, резкость объектива, разрешение сенсора, утечка фотонов), рис. 6.

⁷ Несколько N_{im} стационарных камер или одна мобильная камера в N_{im} позициях, в зависимости от применения.

⁸ Сигналом принято называть интенсивность или цвет пикселя.

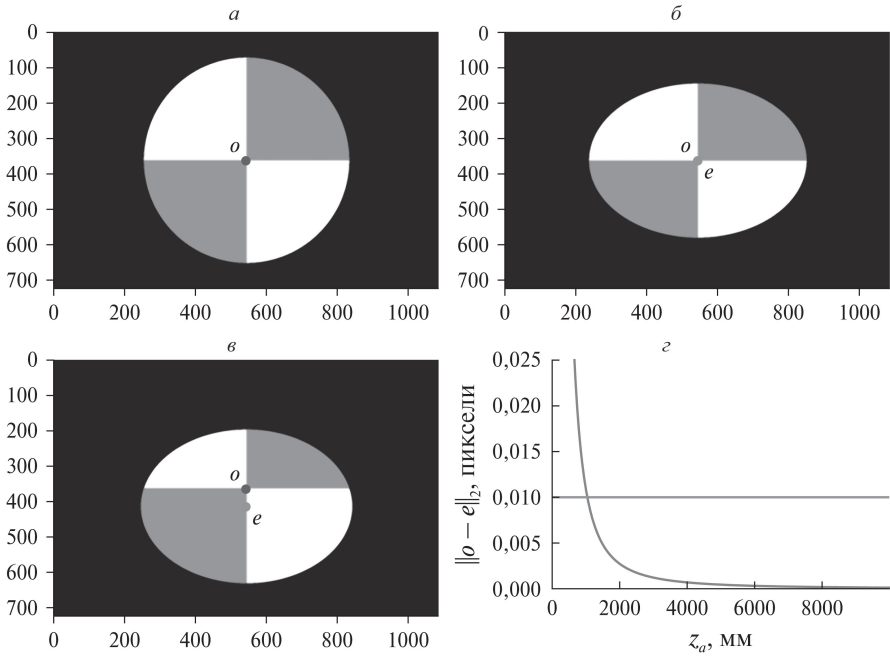


Рис. 5. Круглая метка на объекте – *a*, аффинная проекция метки на экран – *б*, перспективная проекция метки – *в*, зависимость перспективного эффекта $\|o - e\|_2$ от расстояния до метки z_a – *г*.

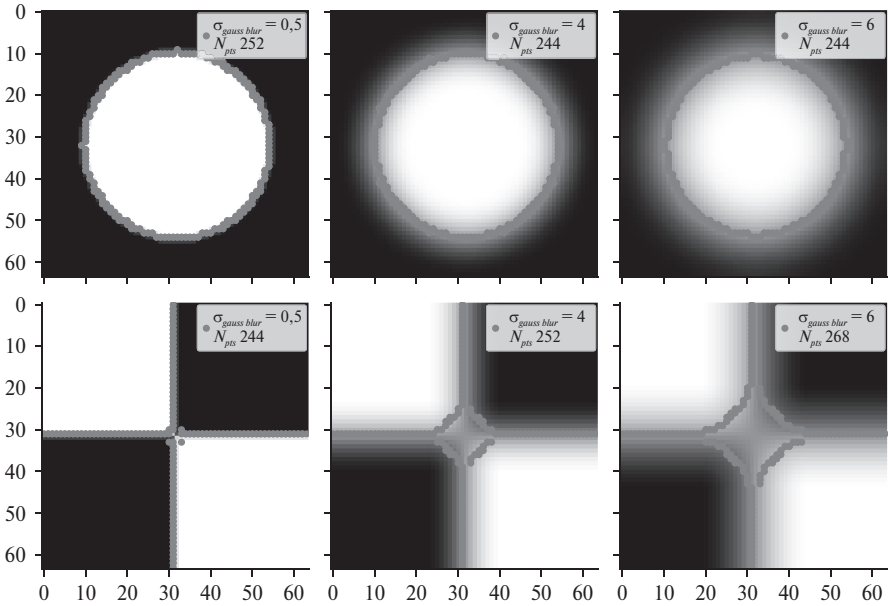


Рис. 6. Влияние снижения резкости (слева-направо) для круга и квадрата: в отличие от круглой метки (верхний ряд), структура квадрата (нижний ряд) существенно искажается. Точками обозначен контур признака – пиксели с наибольшей амплитудой градиента сигнала.

Геометрические факторы детектирования. Все упомянутые детекторы и большинство им подобных опираются на непрерывность поверхности в окрестности интересующей точки. То есть, чем более гладкая, близкая к плоскости поверхность, тем более устойчивым, воспроизводимым на других изображениях будет найденный признак.

К примеру, SIFT-детектор «подавляет» точки на линиях [23], чтобы исключить контуры объекта и возможное попадание информации с фона. Предположение, что окрестность особой точки – это небольшой плоский участок поверхности в 3D, строго говоря, означает, что отображение «3D-поверхность – 2D-изображение»⁹ можно выполнить через аффинное преобразование наблюдаемой области:

$${}^oA_a = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & 0 & 1 \end{bmatrix}, \quad \det({}^oA_a) \neq 0.$$

Оно включает шесть степеней свободы [26]: 2D сдвиг, 2D масштаб, 1D диагональное сжатие и 1D поворот. В более общем случае (без допущения о малом размере участка) имеет место проективное или гомографическое (англ. «homography») преобразование:

$${}^oH_a = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}, \quad \det({}^oH_a) \neq 0.$$

В компьютерной графике аффинная трансформация oA_a соответствует ортогональной проекции, а oH_a – перспективной. Преобразование oA_a переносит некоторый признак с плоского объекта прямо на изображение a , сохраняя параллельность линий и пропорции расстояний на линии, в отличие от более общего преобразования oH_a . Модель этих преобразований в однородных координатах имеет вид:

$$(2) \quad \lambda \begin{bmatrix} u_a \\ v_a \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x_o \\ y_o \\ 1 \end{bmatrix},$$

$a, b \in [1 \dots N_{im}]$ – индексы изображений, u_a, v_a – координаты точки на изображении (сам признак или точка его 3D-окрестности, индекс точки k опущен для краткости), x_o, y_o – координаты точки на объекте в системе координат 3D-плоскости признака ($z_o = 0$), λ – масштабный коэффициент, по сути, необходимый для удобства формулировки, легко исключается из системы линейных уравнений (2) путем подстановки, если требуется найти oH_a .

⁹ Сходно трассировке лучей в компьютерной графике – поиск пересечения луча «из пикселя» изображения и 3D-плоскости.

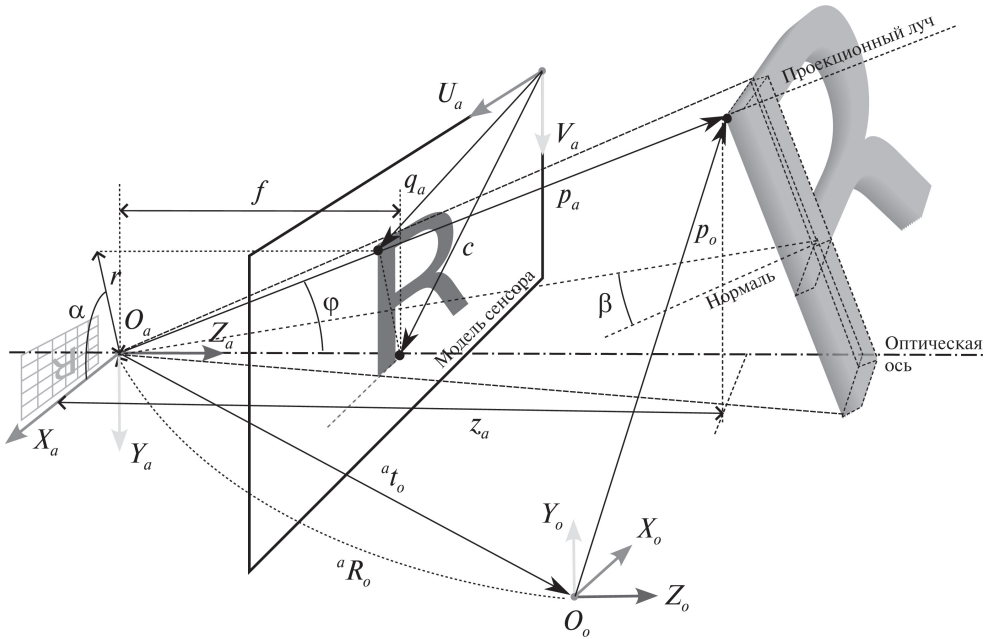


Рис. 7. Перспективная или прямолинейная модель камеры (6): 3D-объект относительно различных систем координат и его проекция на экран или сенсор (подробное описание в тексте).

Рассмотрим пример. Пусть на плоском объекте изображена круглая метка (рис. 5,а) с радиусом $\mathcal{R} = 5$ мм. При аффинной проекции метки на экран (с углом наблюдения $\beta = 45^\circ$, пояснение на рис. 7), центр круга $\mathbf{o} \in \mathbb{R}^2$ будет идеально соответствовать центру $\mathbf{e} \in \mathbb{R}^2$ наблюдаемого эллипса (рис. 5,б). В действительности, при очень близком расстоянии до поверхности $z_a = 15$ мм будет наблюдаться мощное проективное искажение (рис. 5,в), величина которого в зависимости от z_a представлена на рис. 5,г. Условия эксперимента представлены в разделе 7, использовалась камера Canon EOS 5DS.

Зависимость, представленная на рис. 5,г, верна для метки с радиусом $\mathcal{R} = 5$ мм и $\beta = 45^\circ$. Для построения аналогичной зависимости для метки другого радиуса, например $\mathcal{R}' = 10$ мм, при том же угле наблюдения β необходимо умножить (масштабировать) абсциссу на представленном графике (рис. 5,в) на коэффициент $\frac{\mathcal{R}'}{\mathcal{R}}$.

Разница между операторами oA_a и oH_a наглядно представлена на рис. 5. Когда расстояние до поверхности мало, а область для выделения центра признака велика наблюдается все более существенное отклонение между обнаруженной точкой $\mathbf{e}$ и проекцией $\mathbf{o}$ искомой 3D-точки на объекте. Когда отклонение превышает 1 : 10 от стандартной ошибки репроекции в задаче оптимизации пучка (рис. 5,г, красная линия на графике), то нужно оптимизировать уже не только 3D-точку объекта, но и параметры ориентации участка поверхности в пространстве (раздел 6). В некоторых приложениях проективные искажения позволяют эффективно восстанавливать геометрию объекта по двум и более копланарным окружностям [43].

Помимо проективных искажений, описанных выше, отклонение $\|\mathbf{o} - \mathbf{e}\|_2$ может быть вызвано и куда более сложными в моделировании нелинейными эффектами, например, проекционной моделью камеры (дисторсия объектива и пр.) [44] или кривизной поверхности. Если влияние этих эффектов на локальную геометрию признака велико, то для аккуратного решения задачи оптимизации пучка можно напрямую минимизировать невязку значений интенсивности в каждом пикселе изображения [26, 39]. Указанный подход требует существенно больше вычислительных ресурсов. Например, вместо двух уравнений невязки относительно центра каждой метки (задача оптимизации пучка, раздел 6), будет $O(\pi\mathcal{R}^2)$ уравнений¹⁰ разностей интенсивностей цвета в пикселях признака. Более того, растет и число параметров: на каждую оптимизируемую 3D-точку добавляется 3D-нормаль¹¹, т.е., минимум пять параметров на точку, вместо трех. Такое решение будет оправдано при работе с широкоугольной моделью камеры, существенно отличающейся от распространенной перспективной модели (раздел 6.1).

Оптические факторы детектирования. Выделение признака по большой окрестности создает трудности, описанные ранее. Кажется, что простым решением будет сокращение размера области. Например, используя признак типа крест, угол или иные пересечения линий на объекте [45]. Но при сильном оптическом размытии или иной потере контрастности (например, при подавлении шума) структура такого признака может быстро разрушаться (рис. 6).

Дело в том, что высокочастотная составляющая сигнала¹², необходимая для отображения четких границ угла, постепенно утрачивается. Таким образом, круг с гладкой структурой контура остается наиболее универсальной меткой для самых точных измерений (даже если нужен лишь контур, а не центр). Круглые метки можно детектировать в реальном времени следующей последовательностью шагов:

- 1) грубая локализация: детектор контрастных точек или «пиков» (англ. «blob») [35, 36];
- 2) выделение контуров [46, 47] или градиентов в окрестности эллипса [48];
- 3) надежная и точная локализация: аппроксимация квадратичной функцией эллипса геометрического места точек контура или поля градиентов.

Стоит заметить, что субпиксельное уточнение при выделении контуров может не требоваться. Исходя из стандартной ошибки оценки среднего значения: шум дискретизации точек контура $e_{edge} \sim \mathcal{U}(-0,5, 0,5)$, $\sigma_{edge} = \frac{1}{\sqrt{12}}$ (в пикселях) значительно усредняется, так что шум определения центра метки (среднее значение случайной величины) кратно меньше и составляет

¹⁰ По два уравнения на каждый пиксель, число которых пропорционально площади наблюдаемой метки.

¹¹ Ориентация единичной нормали задается двумя углами.

¹² В компьютерном зрении, как правило, говорят о частоте пространственного сигнала, в отличие от функции времени в электронике.

$\sigma_{center} = \frac{\sigma_{edge}}{\sqrt{N_{pts}^{edge}}}$, N_{pts}^{edge} – число точек, используемых для определения центра.

Например, для окружности $\mathcal{R} = 25$ пикселей уровень шума дискретизации найденных координат центра σ_{center} будет составлять $1/\sqrt{12 \cdot 2 \cdot \pi \cdot 25} = 0,02$ пикселя.

Количество признаков не компенсирует их низкое качество – общее правило для точной фотограмметрии. Поскольку параметрическая модель луча (или камеры) может быть нелинейной функцией высокого порядка (раздел 6), особенно на краях поля зрения, это неизбежно приводит к проблеме «переобучения».

5.2. Дескриптор признаков

После детектирования для окрестности каждой найденной ключевой точки на изображении выполняется построение локальной характеристики или дескриптора $\mathcal{D}_a = \{\mathbf{d}_a^k\}_{k=1}^{N_{pts}^a}$ (второй шаг фотограмметрического конвейера, стр. 6). Ключевое свойство дескриптора – это устойчивость или инвариантность к геометрическим преобразованиям метки (из-за движения камеры или объекта) и к оптическим искажениям сигнала, например, из-за изменений освещенности, резкости или шума. То есть при самых разнообразных условиях получения изображения одна и та же точка объекта должна иметь неизменный, но при этом уникальный дескриптор. Инвариантность дескриптора, в общем случае, недостижима, поэтому существует большое разнообразие подходов эффективных в том или ином сценарии.

Например, в SfM-задаче для крупногабаритных объектов аффинные искажения практически не меняют углы в структуре признака, но возможны повороты и масштабные изменения, аналогично наблюдениям звездных тел [49]. В близко-ракурсной фотограмметрии для контроля качества средние и малогабаритных объектов, наоборот, могут проявляться заметные проективные искажения признаков (рис. 5). Таким образом, эффективность дескрипторов может существенно различаться в разных условиях [30]. Это важно учитывать при работе с подходами на основе машинного обучения (МО) [25, 31, 32, 34, 38] (при подготовке обучающей базы), или выборе подходящего дескриптора, опираясь на сравнительные результаты обзора [29].

Дескриптор естественных особенностей текстуры может быть вектором вещественных чисел: $\mathbf{d} \in \mathbb{R}^n$ для SIFT [23], SURF [24], SuperPoint [25], DISK [38] или двоичным вектором $\mathbf{d} \in \{0, 1\}^n$ для BRIEF [50], ORB [51], AKAZE [52]. Также большинство искусственных кодовых меток «модулируются» двоичным дескриптором [53–56]: QR-код, ARtag, AprilTag, ArUco, CCTag, Schneider’s Coded Target (SCT) и пр. Кодовые метки SCT [53] (рис. 3) применяются в экспериментальной части. Метки с концентрическими окружностями (CCTag, SCT и т.п.) часто используются в промышленной фотограмметрии. Декодирование меток или определение двоичной последовательности $\mathbf{d} \in \{0, 1\}^n$ существенно упрощается за счет возможности компенсировать аффинные искажения признака: зная центр, угол поворота и величину главных осей можно вычислить пять параметров аффинного преобразования.

Машинное обучение может эффективно применяться для декодирования как отдельных меток, так и произвольных областей на изображении [57]. Для построения дескрипторов можно применять архитектуры на основе сверточных сетей, например, Resnet [58] или Unet [59]. В качестве выходного слоя можно использовать полносвязный побитовый классификатор, например, 12 выходных нейронов с сигмоидальной функцией активации для 12-битной метки. За счет объединения в батч пикселей окрестностей меток с нескольких изображений можно эффективно решать задачу декодирования на GPU в реальном времени.

5.3. Сопоставление дескрипторов

На данном шаге найденные векторы дескрипторов признаков от каждого изображения $\mathcal{D}_a = \{\mathbf{d}_a^k\}_{k=1}^{N_{pts}^a}$, $\mathcal{D}_b = \{\mathbf{d}_b^k\}_{k=1}^{N_{pts}^b}$, $a, b \in [1 \dots N_{im}]$ сопоставляются между собой (**третий шаг** фотограмметрического конвейера, стр. 6) и формируется множество соответствующих индексов: $\mathcal{M}_{a \rightarrow b} = \{(i^k, j^k) \mid i^k \in [1 \dots N_{pts}^a], j^k \in [1 \dots N_{pts}^b]\}_{k=1}^{N_{pairs}^{a \rightarrow b}}$. Для каждого дескриптора исходного множества $\mathcal{D}_a$ с индексом $i \in [1 \dots N_{pts}^a]$ выполняется поиск ближайшего соседа с индексом j^* на множестве $\mathcal{D}_b$ (l_2 -норма разности для примера):

$$(3) \quad j^* = \underset{j \in [1 \dots N_{pts}^b]}{\operatorname{argmin}} \|\mathbf{d}_a^i - \mathbf{d}_b^j\|_2.$$

В простейшем случае строятся два множества $\mathcal{M}_{a \rightarrow b}$, $\mathcal{M}_{a \leftarrow b}$ (стрелкой обозначено направление поиска) с последующей фильтрацией, представленной в разделе 5.4. В общем случае, оптимальное сопоставление более двух дискретных множеств дескрипторов относится к NP -полной транспортной задаче. Методы сопоставления с использованием машинного обучения [31, 32, 34, 60] эффективно аппроксимируют дискретную задачу поиска, объединяя локальные и глобальные свойства дескрипторов.

Если в качестве признака используется пиксель на контуре вместо особой точки (центр маркера, например), строгого соответствия 2D-точек между парой изображений может и не быть. В этом случае координаты соответствующего признака необходимо интерполировать (допуская локальную гладкость или плоскостность поверхности).

В качестве функции расстояния, в зависимости от природы вектора, часто выступает l_2 -норма (евклидово расстояние) [23–25, 38, 61] или для двоичного дескриптора [50–52] – количество идентичных бит (расстояние Хэмминга). В [61] предлагается эффективное преобразование функции расстояния для SIFT-дескриптора $\mathbf{d} = [d^1, \dots, d^{128}]^T$, которое позволяет существенно повысить вероятность найти правильные связи между изображениями. Исходно расстояние между дескрипторами определяется через евклидову норму, т.е. $\operatorname{dist}_E(\mathbf{d}_a, \mathbf{d}_b) = \|\mathbf{d}_a - \mathbf{d}_b\|_2$, однако, учитывая, что $\|\mathbf{d}_a\|_2 = \|\mathbf{d}_b\|_2 = 1$, верным будет:

$$\|\mathbf{d}_a - \mathbf{d}_b\|_2 = \sqrt{2 - 2 \sum_{l=1}^{128} d_a^l d_b^l}.$$

В основе SIFT – гистограмма частот или функция распределения некоторой характеристики окрестности метки. При сопоставлении 3D-признаков облаков точек также применяются гистограммы, например FRFH [62]. Для сравнения схожести распределений вместо l_2 -нормы, лучше использовать f-дивергенцию Хеллингера [61]. В результате l_1 -нормировки и поэлементного корня расстояние между дескрипторами можно привести к соответствующему виду:

$$\text{dist}_H(\mathbf{d}_a, \mathbf{d}_b) = \sqrt{2 - 2 \sum_{l=1}^{128} \sqrt{c_a \ c_b \ d_a^l \ d_b^l}},$$

где c_a, c_b – нормировочные коэффициенты. Таким образом, варьируя функцию расстояния можно существенно увеличить число верных соответствий.

Поскольку количество изображений N_{im} и среднее число найденных на них признаков N_{pts} может быть велико, общая сложность поиска похожих дескрипторов между всеми изображениями нередко оказывается неприемлемо высокой $O(N_{im}^2 \cdot N_{pts}^2)$, так как для надежных результатов нужны десятки наблюдений конкретной 3D-точки, а их общее число N_{pts} может превышать 1000. В этом случае, в первую очередь, применяют ускоряющие структуры в виде случайных деревьев с приближенным поиском ближайшего соседа [63]¹³. Для бинарных дескрипторов функция расстояния отлична от евклидовой, поэтому применяются методы на основе хеш-функций [65]. Приближенный поиск может допускать существенное количество ошибок, но их можно отфильтровать на следующем этапе.

Так как изображения могут не рассматривать одну и ту же область объекта, для ускорения выбора подходящей пары используются VoW-техники [61, 66] (аббревиатура англ. «Bag-of-Words»), это позволяет быстро исключить перекрывающиеся кадры и необходимость сопоставлять все дескрипторы для пары изображений между собой.

5.4. Фильтрация найденных соответствий

Результаты поиска соответствий нужно фильтровать (**четвертый шаг** фотogramметрического конвейера, стр. 6) независимо от методики сопоставления, чтобы минимизировать число ошибок или выбросов (англ. outlier).

Из (3) следует, что направление поиска имеет значение и множества $\mathcal{M}_{a \rightarrow b}$ и $\mathcal{M}_{a \leftarrow b}$ могут быть неидентичны. На этом основывается фильтр «взаимного соответствия» [30–32], т.е. используются взаимно ближайшие дескрипторы или пересечение пар из множеств $\mathcal{M}_{a \rightarrow b}$ и $\mathcal{M}_{a \leftarrow b}$, тогда как остальное отсеивается. Другой популярный фильтр «уникальности дескриптора» [23] – отсеивает пары, если один из дескрипторов близок по метрике расстояния

¹³ Обычные kd-деревья мало эффективны – проблема, известная как «проклятие размерности», когда последовательный поиск оказывается быстрее прохода по дереву из-за распределения информации в многомерных структурах данных [64].

сразу к нескольким сопоставляемым дескрипторам. Еще один распространенный способ – использование многократно повторяющихся дескрипторов, образующих последовательность (например, в анализе видеоданных).

Данные и подобные эвристики сами по себе не очень эффективны и могут удалять большое количество верных соответствий. Это связано с тем, что локальные свойства текстуры объекта, представленные дескрипторами, очень сильно варьируются даже в пределах одного изображения. По сути, задача таких фильтров – отсеив самых грубых ошибок сопоставления и ускорение следующих шагов.

Наиболее эффективные методы фильтрации опираются на глобальный контекст [21, 30, 42], например, проверяют, насколько найденные соответствия удовлетворяют геометрическим свойствам пространства и перспективной модели камеры (раздел 6.1). Если модель камеры существенно отличается от перспективной, например для широкоугольных объективов, или имеет большую дисторсию, то требуются быстрые методы, учитывающие данные искажения [40, 67].

Во-первых, самым универсальным и широко применяемым критерием геометрической консистентности является эпиполярная геометрия – все соответствия на паре изображений должны удовлетворять следующему уравнению эпиполярной прямой:

$$(4) \quad \begin{bmatrix} u_a & v_a & 1 \end{bmatrix} {}^aF_b \begin{bmatrix} u_b \\ v_b \\ 1 \end{bmatrix} = 0,$$

u_a, v_a, u_b, v_b – координаты двух проекций одной 3D-точки на изображениях a и b соответственно, ${}^aF_b \subset \mathbb{R}^{3 \times 3}$ – широко известная фундаментальная матрица, $\text{rank}({}^aF_b) = 2$. Геометрический смысл уравнения (4) в том, что матрица aF_b задает соответствие точка-линия между парой изображений:

$\begin{bmatrix} u_a & v_a & 1 \end{bmatrix} {}^aF_b$ – параметры линии на изображении b , в то же время

${}^aF_b \begin{bmatrix} u_b \\ v_b \\ 1 \end{bmatrix}$ – линия на изображении a . Подставляя восемь пар соответствующих точек в (4), строим систему линейных алгебраических уравнений [68]

относительно неизвестной матрицы aF_b . Если учесть, что $\text{rank}({}^aF_b) = 2$, то, используя семь пар точек, можно построить систему нелинейных уравнений для нахождения aF_b [26]. Таким образом, многократно решая данные уравнения для небольших групп соответствующих точек, в статистических методах RANSAC [69] или PROSAC [70], можно отсеивать значительную часть ошибок сопоставления.

Во-вторых, надежность фильтрации можно увеличить, если известно, что точки находятся на плоскости или перепады глубины на поверхности объекта существенно меньше его габаритов. В этом случае потребуется лишь 4 соответствия для задания модели преобразования 2D-точек при использовании

методов RANSAC или DEGENSAC [71]:

$$(5) \quad \lambda \begin{bmatrix} u_a \\ v_a \\ 1 \end{bmatrix} = {}^a H_b \begin{bmatrix} u_b \\ v_b \\ 1 \end{bmatrix},$$

где ${}^a H_b \subset \mathbb{R}^{3 \times 3}$ – матрица проективного преобразования или гомография между соответствующими проекциями общей 3D-точки, $\det({}^a H_b) \neq 0$.

Методы машинного обучения позволяют выделять особые точки и строить локальные дескрипторы [25, 38]. В [31] текстурные признаки выделяются на практически однородных участках естественной текстуры. В [31, 32, 34] признаки дополняются контекстной информацией из некоторой области изображения, рассчитанной с помощью нейросетевой архитектуры трансформеров или ее модификаций для ускорения вычислений. В результате парного сопоставления получается матрица вероятности соответствия признаков:

$$P^{a,b} \in \mathbb{R}^{N_{pts}^a \times N_{pts}^b}, \quad \sum_{i=1}^{N_{pts}^a} P_{ij}^{a,b} \leq 1, \quad \sum_{j=1}^{N_{pts}^b} P_{ij}^{a,b} \leq 1.$$

Таким образом, можно выбрать соответствия с требуемой достоверностью, исключая точки невидимые с обоих ракурсов. В контексте точной фотограмметрии приведенные решения на основе МО отлично подходят для поиска надежного начального приближения и построения приблизительной позиции камер и облака 3D-точек. Но, к сожалению, они имеют низкую точность локализации признаков – ошибка репроекции в различных задачах нередко превышает один пиксель, а угловая ошибка определения положения свыше 5° , часто обучающая выборка строится на плоских участках сцены, т.е. значительно опирается на (5), а результаты сопоставления все равно требуют фильтрации [30, 32].

Практически вне зависимости от конструкции локальный дескриптор в фотограмметрии используется на предварительном этапе поиска начального приближения в задаче оптимизации пучка при отсутствии априорной информации о геометрии сцены и позиции камер. Когда эти параметры известны с достаточной точностью, сопоставление признаков можно выполнять по эпиполярным линиям, полученным из (4), что значительно повышает количество верных соответствий.

6. Оптимизационная задача фотограмметрии

6.1. Модель камеры и перспектива

Чтобы решить задачу оптимизации в фотограмметрическом конвейере, известную как оптимизация пучка для проекционных лучей, нужно определить ключевой компонент конвейера. «Сердцем» данной технологии, без преувеличения, является математическая модель камеры – функция для отображения

или проекции точек окружающего 3D-пространства на изображение (экран или сенсор). Ключевым свойством любой модели, рассмотренной в обзоре, является прямолинейное распространение света, дифракция или хроматические aberrации пренебрежимо малы. Через любую 2D-точку на сенсоре можно провести луч, который попадет в соответствующую 3D-точку на поверхности объекта. Таким образом, модель камеры задает направление проекционных лучей по 2D-точке сенсора, внутренним параметрам и положению камеры в пространстве.

Перспективная или прямолинейная точечная (англ. «rectilinear pinhole») модель камеры [3, 11, 26, 41, 72, 73] – самая распространенная в компьютерном зрении. Характерная особенность – 3D-прямые в пространстве объектов проецируются в 2D-прямые в пространстве изображения. Данная модель часто предполагается (явно или неявно) в качестве базовой в различных исследованиях. Например, ее используют при поиске гомографии (5), при определении взаимного положения между изображениями [25, 31, 32, 34] или 3D-реконструкции сцены [42]. Указанная модель является начальным приближением для более сложных вариантов параметризации проекционных лучей, рассмотренных далее. Прямолинейная модель камеры определяется проекционной матрицей ${}^aP_o \in \mathbb{R}^{3 \times 4}$, $\text{rank}({}^aP_o) = 3$ и задает следующую взаимосвязь:

$$(6) \quad \lambda \mathbf{q}_a^h = {}^aP_o \mathbf{p}_o^h = \begin{bmatrix} f_u & s_{uv} & u_c & 0 \\ 0 & f_v & v_c & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} {}^aR_o & {}^a\mathbf{t}_o \\ \mathbf{0}^T & 1 \end{bmatrix}_{4 \times 4} \begin{bmatrix} x_o \\ y_o \\ z_o \\ 1 \end{bmatrix} = \begin{bmatrix} K & \mathbf{0} \end{bmatrix}_{3 \times 4} {}^aT_o \begin{bmatrix} x_o \\ y_o \\ z_o \\ 1 \end{bmatrix},$$

$\mathbf{q}_a^h = [u_a, v_a, 1]^T$ – 2D-координаты на изображении a , результат проекции 3D-точки $\mathbf{p}_o^h = [x_o, y_o, z_o, 1]^T$ в СК объекта;

K – верхне-треугольная матрица внутренних параметров камеры $\det(K) \neq 0$ (пять степеней свободы, в пикселях), включает фокусные расстояния f_u, f_v ¹⁴, координаты центра репроекции $\mathbf{c} = [u_c, v_c]^T$ – точки, где оптическая ось объектива пересекает сенсор (рис. 7), а также диагональное искажение сенсора s_{uv} ;

${}^aT_o \in \mathbf{SE3}$ – матрица внешних параметров камеры (шесть степеней свободы) определяет положение объекта относительно камеры в момент получения изображения a и включает поворот ${}^aR_o \in \mathbf{SO3}$ и сдвиг центра СК объекта ${}^a\mathbf{t}_o$;

параметр λ полагают равным z_a , где z_a – z -координата рассматриваемой точки или расстояние от начала координат O_a вдоль оси Z_a до точки (рис. 7)¹⁵.

На основе (6) определим, полезные в будущем, нормированные однородные проекционные координаты экрана $\bar{\mathbf{q}}_a^h$. В пространстве объектов они соответ-

¹⁴ В компьютерном зрении, в отличие от классической оптической модели, выделяют два «фокусных расстояния» для удобства, чтобы учесть в этих величинах возможную неквадратную форму пикселя.

¹⁵ Трехмерное представление поверхности нередко хранится в виде карт глубин относительно оптического центра объектива.

ствуют координатам 3D-точек на плоскости, для которых $Z_a = 1$ перед камерой, или проекции на камеру с внутренними параметрами $K = I_{3 \times 3}$ ($f = 1$):

$$(7) \quad \bar{\mathbf{q}}_a^h = \begin{bmatrix} \bar{u}_a \\ \bar{v}_a \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{x_a}{z_a} \\ \frac{y_a}{z_a} \\ 1 \end{bmatrix} = \frac{1}{z_a} \begin{bmatrix} {}^a R_o & {}^a \mathbf{t}_o \end{bmatrix} \begin{bmatrix} x_o \\ y_o \\ z_o \\ 1 \end{bmatrix}.$$

Схема работы прямолинейной модели камеры представлена на рис. 7. Для наглядности в примере можно положить $f_u = f_v = f$, $s_{uv} = 0$ (часто эти допущения верны и для точных измерений [17]); физический размер пикселя, сенсора, величина фокусного расстояния f для математической постановки значения не имеют, важно соотношение величин; физический сенсор (с отраженной проекцией) располагается за центром СК камеры O_a (оптический центр объектива), его математическую модель принято располагать между центром и объектом [72].

Используя структуру ${}^a P_o$ из (6) и, учитывая, что ${}^a R_o^T = {}^a R_o^{-1}$, можно легко построить уравнение проекционного луча для любой 2D-точки экрана и вычислить 3D-координаты точки $\mathbf{p}_o$ относительно СК объекта:

$$(8) \quad \mathbf{p}_o = \begin{bmatrix} x_o \\ y_o \\ z_o \end{bmatrix} = {}^a R_o^T (K^{-1} \lambda \mathbf{q}_a^h - {}^a \mathbf{t}_o),$$

где известными являются $\mathbf{q}_a^h$ – координаты точки на экране, $\lambda = z_a$ – глубина, а также даны внутренние и внешние параметры камеры.

Центральная задача фотограмметрии – *задача триангуляции* заключается в определении координат 3D-точки $\mathbf{p}_o$ по известным проекциям $\mathbf{q}_a$, $\mathbf{q}_b$. Для прямолинейной модели камеры с учетом (6) может быть построена система линейных алгебраических уравнений:

$$(9) \quad \begin{cases} \lambda_a \mathbf{q}_a^h = [K & \mathbf{0}] {}^a T_o \mathbf{p}_o^h, \\ \lambda_b \mathbf{q}_b^h = [K & \mathbf{0}] {}^b T_o \mathbf{p}_o^h, \end{cases}$$

где ${}^a T_o$, ${}^b T_o$ – матрицы, определяющие взаимное положение объекта и наблюдателя/камеры, K – внутренние параметры одной используемой камеры (для случая, когда камер несколько, матрицы различаются).

В задаче триангуляции K , ${}^a T_o$, ${}^b T_o$ известны и важно наличие параллакса, т.е. $\|{}^a \mathbf{t}_b\|_2 \neq 0$. Система (9) содержит шесть уравнений и пять неизвестных (λ_a , λ_b , $\mathbf{p}_o$). Неизвестные λ_a и λ_b легко выражаются и получается переопределенная система линейных алгебраических уравнений (четыре уравнения и три неизвестных), которая может быть решена в смысле наименьших квадратов относительно 3D-координат $\mathbf{p}_o$, к примеру, с помощью нормальных уравнений.

При решении центральной задачи оптимизации пучка (раздел 6.5) в уравнении (9) известными величинами являются лишь детектированные координаты $\mathbf{q}_a$, $\mathbf{q}_b$ (по два уравнения невязки для каждой наблюдаемой на изображении 3D-точки объекта), все остальные параметры оптимизируются.

6.2. Базовые проекционные модели камер

Прямолинейная модель камеры (6) задает взаимосвязь угла φ между проекционным лучом и оптической осью с координатами точки на экране $\mathbf{q}_a$ (рис. 7):

$$\tan(\varphi) = \frac{r}{f} = \frac{\|\mathbf{q}_a - \mathbf{c}\|}{f} = \|\bar{\mathbf{q}}_a\|.$$

Другими словами, это связь углов между лучом из окружающего пространства (коллинеарен радиус-вектору $\mathbf{p}_a$, входит в объектив камеры) и лучом, падающим на сенсор за объективом, в перспективной математической модели камеры оба луча лежат на одной прямой, проходящей через оптический центр объектива O_a на рис. 7.

В физической модели объектива пучок рассеянного света собирается с участка поверхности и испытывает ряд сложных преломлений, представленные математические модели аппроксимируют данный процесс построения изображения. В зависимости от формы линз и конструкции объектива можно выделить по меньшей мере три базовые проекционные модели [44]:

- 1) прямолинейную или перспективную модель: $r = f \tan(\varphi)$;
- 2) стереографическую модель: $r = 2f \tan(\frac{\varphi}{2})$;
- 3) эквидистантную модель: $r = f\varphi$ (идеальная широкоугольная оптика, поскольку разрешающая способность сенсора не зависит от угла φ).

В популярных широкоугольных камерах Insta360 X4 применяют два объектива с эквидистантной проекционной моделью с углом зрения свыше 180° каждый, чтобы обеспечивать круговой обзор в режиме видео-съемки. Пример двух проекций представлен на рис. 8.

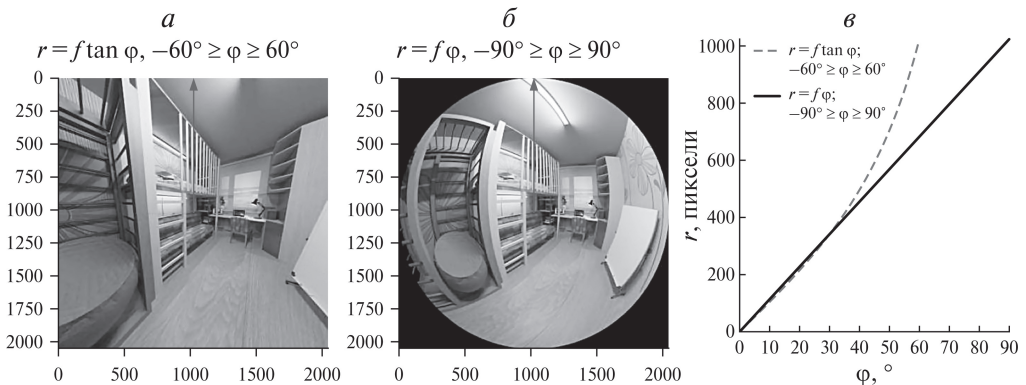


Рис. 8. Реалистичная 3D-модель детской комнаты (рендеринг в Blender 3D [84]) в двух идеальных проекциях без дисторсии: а – прямолинейная, б – эквидистантная.

Приведем далее уравнения для отображения некоторой 3D-точки в СК камеры $\mathbf{p}_a = [x_a, y_a, z_a]^T$ в 2D-точку $\mathbf{q}_a$ на исходном изображении для различных проекционных моделей. Пусть φ , α – сферические координаты проекционного луча (3D-точка на сфере), где φ – угол между оптической осью Z_a и проекционным лучом, α – угол поворота проекционного луча вокруг оси Z_a (рис. 7):

$$(10) \quad \cos(\alpha) = \frac{x_a}{\sqrt{x_a^2 + y_a^2}}, \quad \sin(\alpha) = \frac{y_a}{\sqrt{x_a^2 + y_a^2}}, \quad \tan(\varphi) = \frac{\sqrt{x_a^2 + y_a^2}}{z_a}.$$

В результате координаты точки на изображении можно вычислить через полярные координаты α , $\rho(\varphi)$:

$$(11) \quad \mathbf{q}_a^h = K \begin{bmatrix} \rho(\varphi) \cos(\alpha) \\ \rho(\varphi) \sin(\alpha) \\ 1 \end{bmatrix},$$

$$\rho(\varphi) = \begin{cases} \tan(\varphi) & \text{– для прямолинейной модели,} \\ 2 \tan\left(\frac{\varphi}{2}\right) & \text{– для стереографической модели,} \\ \varphi & \text{– для эквидистантной модели.} \end{cases}$$

Подставляя выражения (10) в (11), можно получить 2D-координаты проекции для любой из приведенных моделей. Нетрудно заметить, что для прямолинейной модели (11) принимает вид (6).

Существуют модели, объединяющие прямолинейную, стереографическую и эквидистантную проекции. В [44] вводится параметр для плавного регулирования модели преломления проекционного луча. В этой работе также подчеркивается важность учета сдвига входного зрачка или оптического центра линзы для широкоугольной оптики. Развивая эту идею, следует выделить и универсальные проекционные модели камер (от англ. «generalized camera») с индивидуальной параметризацией луча [45, 74, 75], когда для каждого пикселя изображения задаются собственные параметры луча. На практике выбирается лишь подмножество таких пикселей, а для остальных параметры интерполируются. Естественно, такие модели легко «переобучаются» и для точных результатов требуется очень много данных.

В фотограмметрическом конвейере использование корректной проекционной модели позволяет существенно снизить величину дисторсии и уменьшить общее число параметров модели, требующих хорошего начального приближения и аккуратной калибровки.

6.3. Дисторсия проекционных моделей

Ввиду различных факторов, например сложности производства линз и объективов, искривления сенсора или погрешности сборки камеры, чаще всего реальная проекция отклоняется от модели, особенно на краях поля зрения.

Это явление называется *дисторсией*. Нередко дисторсией называют отличие между базовыми проекционными моделями, например искривление прямых линий или формы 3D-объекта на рис. 8, поскольку функция дисторсии аппроксимирует данный эффект [67, 76].

Как правило, для прямолинейной модели камеры выделяют радиальную и тангенциальную составляющие дисторсии [3, 4, 17, 72]. Несмотря на «почтенный возраст» и существенное развитие технологий, эта модель работает очень эффективно, что демонстрируется в разделе 7.

Радиальная дисторсия является наиболее существенным фактором искажения прямолинейной проекции, с искривлением прямых в виде «бочки» или «усов». Она аппроксимируется полиномом четной степени, так как функция дисторсии симметрична ввиду центральной симметрии линз:

$$(12) \quad \delta \mathbf{r}(\mathbf{q}) = \mathbf{q}(k_r^1 r^2 + k_r^2 r^4 + k_r^3 r^6 + \dots),$$

$\mathbf{q} = [u, v]^T$ – некоторая 2D-точка на экране (центр дисторсии в $[0, 0]^T$);

k_r^γ – коэффициенты радиальной дисторсии ($\gamma \in \mathbb{N}$);

$\delta \mathbf{r}(\mathbf{q})$ – отклонение наблюдаемой 2D-точки от прямолинейной проекции $\mathbf{q}$ (модель проекции (6)) из-за радиальной дисторсии.

Основной вклад в отклонение от прямолинейной модели, как правило, вносит первое слагаемое $k_r^1 r^2$ правой части (12), тогда как последующие слагаемые часто не используют.

Тангенциальная дисторсия вызвана погрешностями установки системы линз объектива и в отличие от $\delta \mathbf{r}(\mathbf{q})$ создает асимметричное искажение поля сформулирована еще в [2]:

$$(13) \quad \delta \tau(\mathbf{q}) = \begin{bmatrix} k_\tau^2(r^2 + 2u^2) + 2k_\tau^1 uv \\ k_\tau^1(r^2 + 2v^2) + 2k_\tau^2 uv \end{bmatrix} (1 + k_\tau^3 r^2 + k_\tau^4 r^4 + \dots),$$

где k_τ^γ – коэффициенты тангенциальной дисторсии ($\gamma \in \mathbb{N}$).

Коэффициенты $k_\tau^3, k_\tau^4, \dots$, как правило, не учитывают или вовсе игнорируют фактор $\delta \tau(\mathbf{q})$. Но совсем пренебрегать фактором тангенциальной дисторсии не следует, особенно в области точных измерений с фотограмметрией: в основополагающей работе [3] координаты центра дисторсии совпадают с центром репроекции $\mathbf{c}$ из внутренних параметров камеры K . В то же время в работах [67, 76, 77], пренебрегающих тангенциальной дисторсией, отмечается, что $\mathbf{c}$ и координаты центра дисторсии существенно отклоняются. Хотя в [3, 17] указывается, что тангенциальная дисторсия сильно коррелирует с позицией центра репроекции. Вероятно, в [67, 76, 77] именно неучтенная тангенциальная дисторсия искажает оценку $\mathbf{c}$.

Таким образом, точная модель прямолинейной проекции (6) с учетом дисторсии определяет взаимосвязь между 3D-координатами и их 2D-проекциями

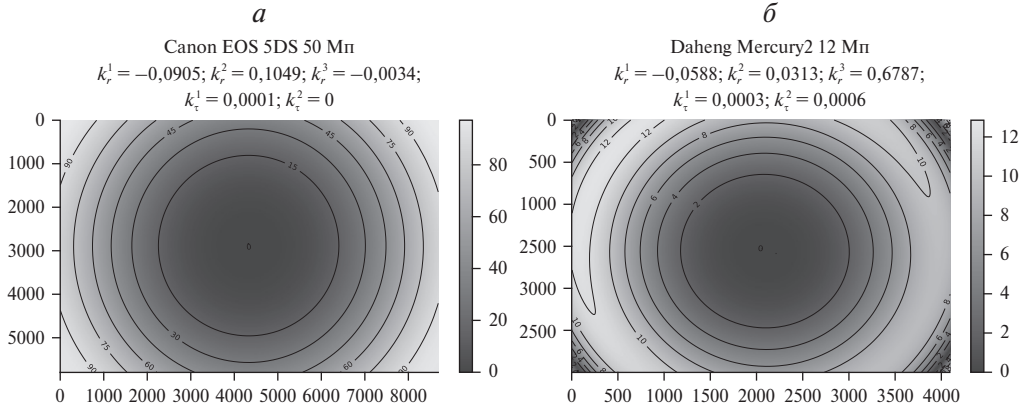


Рис. 9. Амплитуда дисторсии $\|\delta\mathbf{r}(\mathbf{q}) + \delta\boldsymbol{\tau}(\mathbf{q})\|_2 \cdot f$ в каждой точке кадра (в пикселях). Обе оптические системы демонстрируют в основном centrosymmetric отклонение от прямолинейной модели из-за радиальной дисторсии.

следующим образом:

$$(14) \quad K^{-1}\mathbf{q}_a^h + \delta(K^{-1}\mathbf{q}_a^h) = \frac{1}{z_a} \begin{bmatrix} {}^aR_o & \mathbf{a}\mathbf{t}_o \end{bmatrix} \begin{bmatrix} x_o \\ y_o \\ z_o \\ 1 \end{bmatrix} = \bar{\mathbf{q}}_a^h,$$

K – матрица внутренних параметров прямолинейной модели камеры¹⁶,
 $\mathbf{q}_a^h$ – 2D-координаты найденного признака на исходном изображении,
 $\bar{\mathbf{q}}_a^h$ – нормированные 2D-координаты проекции соответствующей 3D-точки из (7),

$\delta(\mathbf{q}^h) = \begin{bmatrix} \delta\mathbf{r}(\mathbf{q}) + \delta\boldsymbol{\tau}(\mathbf{q}) \\ 0 \end{bmatrix}_{3 \times 1}$ – совокупная дисторсия для однородных координат.

Фундаментальная формула (14) распространена в фотограмметрической литературе [2–4, 11, 17], где полиномиальная функция корректирует дисторсию наблюдаемой точки $\mathbf{q}_a$. Это отличается от формулировки в популярной категории «нефотограмметрических» работ, например [40, 41, 67, 72], в них полиномиальная функция, наоборот, добавляет эффект дисторсии к нормированным координатам:

$$(15) \quad \mathbf{q}_a^h = K[\bar{\mathbf{q}}_a^h + \delta(\bar{\mathbf{q}}_a^h)].$$

В разделе 7 приводится сравнение эффективности моделей, основанных на выражениях (14) и (15). Наглядный пример двух откалиброванных камер (моделей дисторсий) представлен на рис. 9.

Для камеры Daheng Mercury2 заметна асимметрия – вклад тангенциальной дисторсии. Более того, в углах кадра наблюдается немонотонность

¹⁶ По сути, K^{-1} выполняет нормировку пространства координат изображения.

(уменьшение) величины радиального искажения и очень большое значение k_r^3 (рис. 9,б). Все это – результат неустойчивости решения оптимизационной задачи из-за нехватки наблюдений (маркеров), особенно в области с сильными искажениями. Другими словами, дисторсия аппроксимируется полиномом высокого порядка, и он некорректно экстраполирует величину искажения [17].

Как следует из представленных результатов, учитывать дисторсию необходимо, так как ее величина очень существенна. Поэтому модели проекции прямолинейной камеры с дисторсией (14) и (15) чрезвычайно распространены и встречаются в массе работ по фотограмметрии, SfM/SLAM, визуальной одометрии, дополненной реальности и пр. В данной работе в экспериментальной части рассматривается прямолинейная модель с дисторсией (14) и (15).

6.4. Автокалибровка

В фотограмметрии выполняется совместная оптимизация 3D-координат наблюдаемых признаков, внутренних параметров камер и их позиций (раздел 6.5). Для выполнения нелинейной процедуры оптимизации сначала нужно найти начальное приближение для искомых параметров, данную задачу выполняет процедура автокалибровки [26, 78] (**пятый шаг** фотограмметрического конвейера, стр. 6). В методах 3D-реконструкции с МО [42] также требуется такая процедура. Для решения задачи предварительной калибровки и позиционирования камер в профессиональной фотограмметрии «Hexagon DPA Pro» применяются специальные инструментальные средства – группы кодовых маркеров, размещенные на заранее известной геометрической структуре в виде «креста». Такой подход имеет существенные неудобства. Представленные далее алгоритмы позволяют избежать необходимости размещения в сцене специальных структур.

В наиболее популярном сценарии фотограмметрии при использовании одной камеры, описываемой прямолинейной моделью (стр. 23), перемещаемой относительно объекта, можно выделить следующие оптимизируемые величины:

- 1) 3D-координаты наблюдаемых признаков на поверхности объекта:
 $\{\mathbf{p}_o^i \in \mathbb{R}^3\}_{i=0}^{N_{pts}};$
- 2) внутренние параметры камеры (элементы K) и коэффициенты дисторсии:
 $\{f, \mathbf{c}, k_r^1, k_r^2, k_r^3, k_\tau^1, k_\tau^2, \dots\};$
- 3) позиции камеры в разные моменты времени:
 $\{^aT_o \in \mathbf{SE3}\}_{a=2}^{N_{im}},$ фиксируя первую позицию $^1T_o = I_{4 \times 4}$ в качестве СК объекта¹⁷.

¹⁷ В общем случае, задача фотограмметрии решается с точностью до **SE3**-трансформации с произвольным масштабом получаемого облака точек [26], таким образом можно зафиксировать параметры одной позиции.

Дисторсией на шаге автокалибровки часто можно пренебречь – найденные параметры будут корректны для центральной области изображения. Точная оценка параметров дисторсии будет выполнена на следующем шаге. Для центра репроекции $\mathbf{c}$ – центр изображения, как правило, является хорошим начальным приближением.

Если дисторсия проекционной модели слишком велика и не позволяет надежно провести автокалибровку или проекционная модель отличается от прямолинейной из-за большого угла зрения, то потребуются итеративный процесс коррекции этих эффектов через приведение к прямолинейной модели и уточнение параметров. В большинстве случаев именно прямолинейная модель проекции используется в алгоритмах автокалибровки [78, 79]. В [79] автокалибровка сводится к определению взаимного положения камер aR_b , ${}^a\mathbf{t}_b$, $a, b \in [1 \dots N_{im}]$, по найденным ранее соответствиям, с использованием (4) и структуры фундаментальной матрицы:

$$(16) \quad {}^aF_b = K^{-T} {}^aE_b K^{-1} = K^{-T} [{}^a\mathbf{t}_b]_{\times} {}^aR_b K^{-1},$$

$$[\mathbf{t}]_{\times} = \begin{bmatrix} 0 & -t_3 & t_2 \\ t_3 & 0 & -t_1 \\ -t_2 & t_1 & 0 \end{bmatrix} - \text{кососимметрическая матрица векторного произ-}$$

ведения. Сдвиг между камерами ${}^a\mathbf{t}_b$ может быть определен с точностью до масштаба (вектор задает лишь направление) – это теоретическое ограничение автокалибровки. Для нахождения aR_b , ${}^a\mathbf{t}_b$ достаточно пяти невязок из (4), полученных из такого же числа соответствующих точек между парой изображений, так как aF_b имеет пять степеней свободы [26]. Минимальное число соответствий позволяет эффективно применять RANSAC [69] для повышения устойчивости найденного решения. В [79] для автокалибровки с двумя видами требуются внутренние параметры камеры (матрица K). При этом фокусное расстояние f сложно оценить заранее, так как оно может существенно варьироваться в зависимости от поля зрения камеры. Если наблюдаемые соответствия (для двух видов или позиций камеры) находятся на плоском участке поверхности объекта, то внутренние параметры оценить невозможно [26, 79]. Таким образом, наиболее надежный способ из [79] – это использовать три и более кадра, полученные с разных позиций.

В рамках данной работы был реализован альтернативный подход [78] на основе факторизации матрицы квадратичной формы¹⁸. Задача сводится к решению нелинейной системы уравнений (детали решения подробно изложены в [78])

$$(17) \quad K^{-1} P_b \Omega P_b^T K^{-T} = \lambda I_{3 \times 3}$$

относительно элементов матрицы квадратичной формы $\Omega = H H^T \in \mathbb{R}^{4 \times 4}$ и матрицы внутренних параметров K , λ – произвольный масштабный коэффициент, $P_b = K [{}^bR_a | {}^b\mathbf{t}_a]_{3 \times 4} H^{-1}$ – известная матрица проективного преобразования (6) (позиция b , например), которая была получена из факторизации

¹⁸ В англоязычной литературе матрица именуется «absolute quadric».

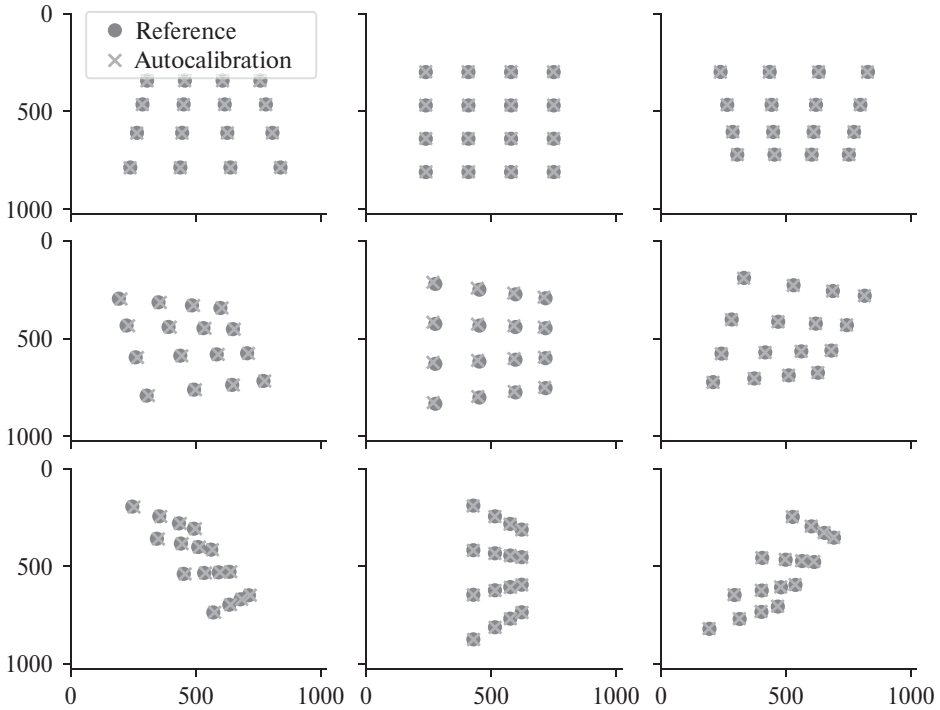


Рис. 10. Модельные изображения (1024x1024 пикселя) 9 наклонных видов для плоского объекта (круглые метки) и проекция найденных 3D-координат (кресты) – результат работы алгоритма автокалибровки [78], выполняющий оценку внутренних и внешних параметров камеры, 3D-координат точек объекта. Среднеквадратическая ошибка репроекции по (6) 3,35 пикселя при амплитуде шума входных наблюдений один пиксель.

фундаментальной матрицы aF_b , а aF_b , в свою очередь, найдена по соответствиям между изображениями a , b . Таким образом, зная соответствия между несколькими видами (не менее трех), можно определить относительное положение камеры bR_a , bt_a и ее параметры K .

На следующем шаге неизвестные 3D-координаты точек вычисляются через уравнение триангуляции (9), далее уже следует задача оптимизации пучка. На рис. 10 представлен результат работы алгоритма автокалибровки для плоского объекта по соответствиям от девяти видов. Разнообразные методы оценки относительного положения и параметров камер без начального приближения также доступны в библиотеке с открытым исходным кодом [80].

6.5. Оптимизация пучка

Измерения, получаемые фотограмметрией, – это 3D-координаты наблюдаемых 2D-признаков, а также пространственное положение объектов (соответствующих облаков точек) относительно камер. Для этого требуется совместная оптимизация множества параметров, приведенных выше. Централь-

ной задачей фотограмметрии является минимизация суммы квадратов невязок или ошибок репроекции (**шестой шаг** фотограмметрического конвейера, стр. 6):

$$(18) \quad \theta^* = \underset{\theta}{\operatorname{argmin}} \mathcal{L}(\theta),$$

где θ – совокупный вектор оптимизируемых параметров (раздел 6.4); $\mathcal{L}(\theta)$ – минимизируемая целевая функция.

Геометрический смысл задачи – сведение пучка проекционных лучей в соответствующих 3D-точках путем минимизации ошибки репроекции, отсюда, вероятно, и происходит название «оптимизация пучка».

Слагаемые целевой функции получены из проекционной модели камеры. Модель в лучшем случае подбирается под объектив камеры, например (6) для прямолинейной камеры без дисторсии (специальный дизайн объектива, к примеру), (11) для широкоугольной камеры, (15) или (14) для точной прямолинейной модели с радиальной и тангенциальной дисторсией. Целевая функция с ошибкой репроекции, построенной с (14), имеет следующий вид:

$$(19) \quad \mathcal{L}(\theta) = \sum_{a=1}^{N_{im}} \sum_{i=1}^{N_{pts}} e_{a,i}(\theta) + w_{scale} \sum_{(i,j,d) \in S} (\|\mathbf{p}_o^i - \mathbf{p}_o^j\|_2 - d)^2,$$

$$e_{a,i}(\theta) = \left\| K(\theta)^{-1} \mathbf{q}_a^{m_a(i)} + \delta_\theta \left(K(\theta)^{-1} \mathbf{q}_a^{m_a(i)} \right) - \frac{1}{z_a} {}^aT_o(\theta) \mathbf{p}_o^i \right\|_2^2,$$

где $e_{a,i}(\theta)$ – квадрат l_2 -нормы 2D-невязки или ошибки репроекции для одной точки на изображении;

$\mathbf{p}_o^i$ (компонент вектора θ) – искомая 3D-точка на поверхности объекта (например, метка на рис. 3), символ h опущен для краткости;

$m_a(i)$ – отображение i -й 3D-точки на объекте в индекс ее проекции на изображении a (полагая, для краткости, что все точки объекта видны на изображении);

$\mathbf{q}_a$ – однородные 2D-координаты признака в пикселях, исходные наблюдения на изображении a ;

$S = \{(i, j, d) | i, j \in [1 \dots N_{pts}], d \in \mathbb{R}^+\}_{i=1}^{N_{scales}}$ – масштабные эталоны (N_{scales} штук) в виде заранее известных точек и расстояний между ними, например, кодированные метки на двух масштабных жезлах на рис. 3;

${}^aT_o(\theta)$ – позиция СК камеры включает матрицу поворота ${}^aR_o(\theta)$ и начало отсчета ${}^a\mathbf{t}_o$ (компонент вектора θ) при построении изображения a ;

w_{scale} – вес слагаемого погрешности измерения длины эталонов. Вес нужен затем, чтобы сбалансировать различное число уравнений двух типов. К тому же, ошибка репроекции может иметь разные единицы измерения: для (15) пиксели и нормированные координаты для (14).

Для лучшей численной устойчивости желательно нормировать координаты изображения, чтобы $\mathbf{q}_a \in [-1, 1]^2$. Это распространенная практика, но при этом можно упустить важный нюанс – сначала нужно перенести начало от-

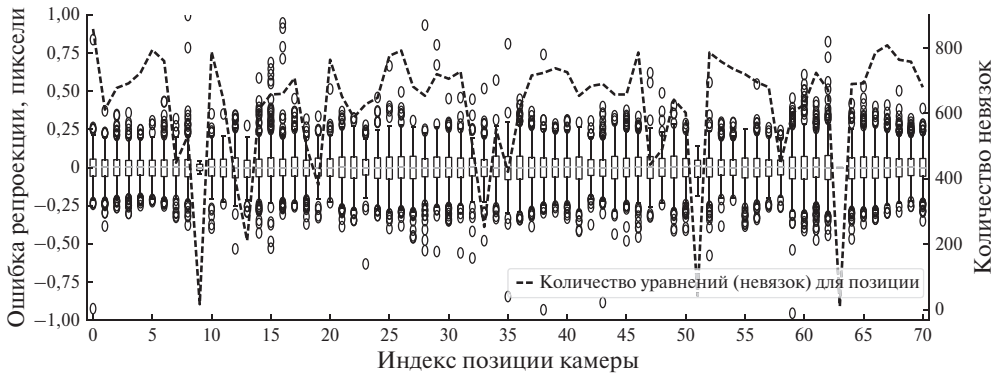


Рис. 11. Распределение ошибки репроекции 3D-точек из (19) (левая ось Y) в пикселях по 71 позиции камеры в виде «ящичков с усами», охватывающих 99 % точек. Кругок обозначает выброс (менее 1 % точек). Пунктирная линия (правая ось Y) обозначает число невязок для позиции – по две на каждую наблюдаемую точку.

счета из центра первого пикселя в крайний угол изображения, чтобы отвязать систему координат от исходных пикселей изображения. Таким образом, результирующая калибровка камеры K не будет привязана к определенному разрешению.

Следует отметить ряд ключевых аспектов оптимизационной задачи (18):

- 1) сравнительно большое число оптимизируемых параметров: в экспериментальной сцене 71 позиция камеры (шесть степеней свободы на каждую), и 430 меток на объекте, т.е. не менее $71 \cdot 6 + 430 \cdot 3 = 1716$ параметров;
- 2) разреженность системы уравнений (якобиана): каждое слагаемое в (19) – ошибка репроекции включает лишь одну позицию, и далеко не все точки объекта наблюдаются в каждом кадре;
- 3) наличие выбросов или точек с очень высокой ошибкой репроекции: на рис. 11 видно, что почти в каждой позиции встречаются точки с невязкой, многократно превышающей среднеквадратическую ошибку σ_{repr} .

Таким образом, необходимо использовать оптимизационные методы устойчивые к шуму и выбросам (даже после аккуратного детектирования и фильтрации признаков). Как видно из рис. 11, выбросы возникают чаще, чем можно прогнозировать, опираясь на нормальное или гауссовское распределение, что подчеркнуто в [39]. В ней предлагается заменить в (19) $e_{a,i}(\theta)$ на $\rho(e_{a,i}(\theta))$, где $\rho(x)$ – неотрицательная скалярная функция «стабилизатор» (от англ. «robustifier» или «robust kernel»): $\rho(x) > 0, \forall x \in \mathbb{R}, \rho'(0) = 0$, которая подавляет градиенты целевой функции в случае выбросов за счет $\rho'(x) < 1$. В экспериментальной части для данной цели использовалась функция Хьюбера:

$$(20) \quad \rho_t(x) = \begin{cases} x, & \text{если } x < t^2, \\ t^2 (2\sqrt{x/t^2} - 1), & \text{иначе,} \end{cases}$$

где $t = 0,05$ пикселя – нормировочный коэффициент.

При невязке в пределах t функция-стабилизатор не оказывает влияния на целевую функцию, однако, в области выбросов существенно уменьшает их вклад в целевую функцию (19).

Для надежного решения задачи оптимизации пучка (18) подходят методы Ньютона с «затуханием», где на каждой итерации алгоритма, выполняется балансировка направления шага [39]. Алгоритм оптимизации Левенберга–Маркварда из данной категории применялся в экспериментальной части. Эффективная реализация данного метода оптимизации, включающая стабилизирующие функции и операции с разреженными матрицами, доступна в [81].

7. Экспериментальная часть

Цель эксперимента – выделить ключевые факторы процесса оптимизации (18), влияющие на точность реконструируемого 3D-облака точек. Для съемки использовалась камера из профессионального фотограмметрического набора «Hexagon DPA Pro» (HDP) – Canon EOS 5DS с черно-белым сенсором $w \times h = 8700 \times 5800$ (50 Мп) и объективом $f = 28$ мм. Стоит заметить, что для точных измерений могут применяться различные профессиональные модели зеркальных фотокамер [28].

Тестовая сцена и пример полученного 3D-облака из 430 точек представлены на рис. 3. Сцена включает два жестких объекта – стальной сварочный стол с наклеенными метками (габарит ~ 2000 мм) и углепластиковая калибровочная плита (габарит ~ 800 мм, метки нанесены производителем). Два масштабных жезла или эталона длины 1390 и 790 мм позволяют задать метрические единицы измерения для восстановленного облака точек, а также обеспечить избыточную проверку точности. Съемка выполнялась тщательным образом со стабилизацией и контролем температурных условий (20°C). Было собрано два практически идентичных набора данных с интервалом в один день (74 и 79 изображений). Обработка изображений за каждый день выполнялась как в программном пакете HDP, так и в собственной программной разработке авторов на основе языка Python и математических библиотек с открытыми исходным кодом Jax [82], SciPy [83], OpenCV [72].

Процесс построения облака 3D-точек с одновременным получением параметров проекционной модели камеры (калибровкой) включает следующие шаги:

- 1) съемка тестового объекта фотокамерой с разных позиций (дистанция между камерой и объектом 1,5–2 м), пример на рис. 4;
- 2) вычисление параметров круглых меток (кодовых и обычных) на изображениях и сопоставление (группировка) кодовых меток по двоичному коду, раздел 5;
- 3) автокалибровка, результат – грубая калибровка камеры (без учета дисторсии (12)) и ее позиции в разные моменты времени, раздел 6.4;
- 4) приблизительная оптимизация пучка, результат – облако 3D-координат кодовых меток и калибровка камер 6.5;

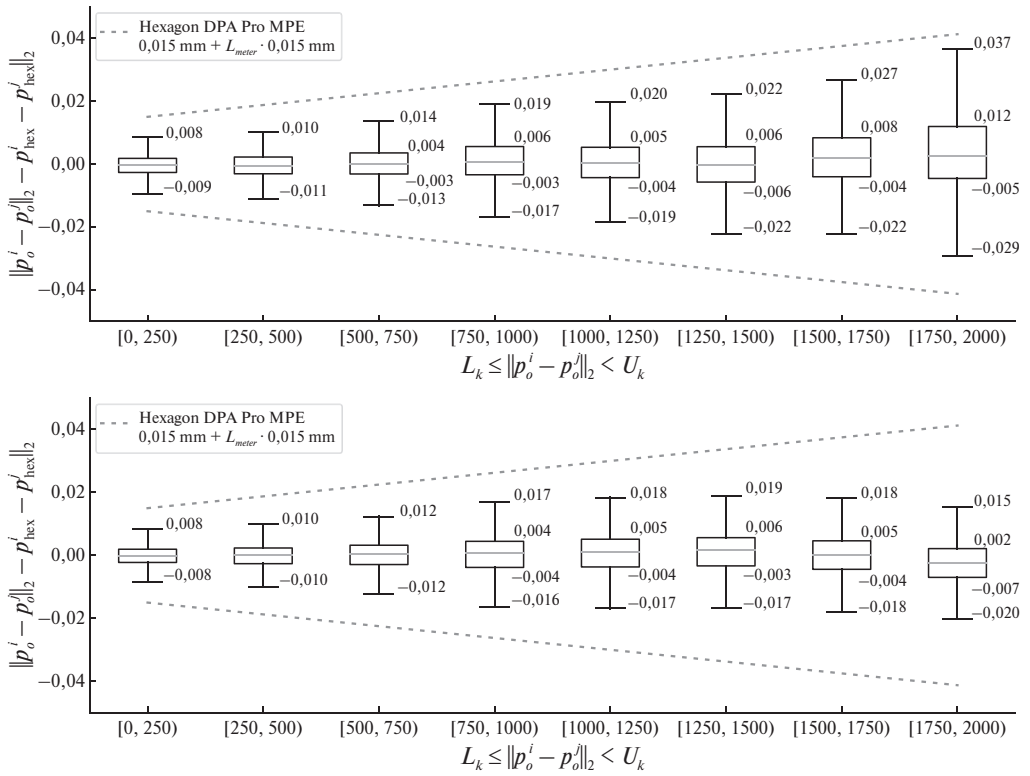


Рис. 12. Сравнение результатов фотограмметрической реконструкции тестовой сцены от «Hexagon DPA Pro» и текущего подхода для двух экспериментов (данные за два дня – верхний и нижний графики). Единицы измерения в **миллиметрах**. На оси абсцисс указаны диапазоны измеряемых расстояний.

- 5) сопоставление (группировка) всех меток по уравнению фундаментальной матрицы (4), полученной из калибровки камеры [26];
- 6) точная оптимизация пучка, результат – облако 3D-координат всех меток и аккуратная калибровка камер.

Далее рассмотрим результаты – реконструированные облака 3D-точек, полученные для различных конфигураций оптимизационной задачи (18). Обозначим проекционную модель Брауна [3] (14), как Браун; со стабилизацией невязки (20), как Браун & $\rho_{0,05}$. Для случая без учета тангенциальной дисторсии (13) Браун & $\rho_{0,05}$ & k_{τ}^0 . Модель (15) обозначим как OpenCV¹⁹ и соответственно предыдущим определениям введем OpenCV & $\rho_{0,05}$, OpenCV & $\rho_{0,05}$ & k_{τ}^0 .

На рис. 12 представлено сравнение реконструкции от HDP и схемы Браун & $\rho_{0,05}$. Каждый график включает $k = 8$ распределений («ящики с усами») разностей расстояний между соответствующими парами точек двух облаков: $\|\mathbf{p}_o^i - \mathbf{p}_o^j\|_2 - \|\mathbf{p}_{hex}^i - \mathbf{p}_{hex}^j\|_2$, $\forall i, j \in [1 \dots N_{pts}]$, где $\mathbf{p}_{hex}$ – точки, по-

¹⁹ Именно такая формулировка дана в [72].

Повторяемость расстояний для облаков 3D-точек, полученных одним алгоритмом по данным от разных дней (меньше-лучше)

Алгоритм	σ_{table} , МКМ	σ_{plate} , МКМ	e_{bar} , МКМ
HDP	46,5	7,2	−2,9 ²⁰
Браун	80,4	22,5	138,7
Браун & $\rho_{0,05}$	18,0	7,5	−6,9
Браун & $\rho_{0,05}$ & k_{τ}^0	41,1	7,5	3,3
OpenCV	41,8	7,6	−154,2
OpenCV & $\rho_{0,05}$	39,8	7,6	−9,0
OpenCV & $\rho_{0,05}$ & k_{τ}^0	41,2	7,7	4,3

лученные из HDP, $\mathbf{p}_o$ – результат оптимизации (18) с моделью Браун & $\rho_{0,05}$. Распределение отклонений дано для определенного интервала $[L_k, U_k)$ расстояний между точками. Красная линия обозначает предельную погрешность «Hexagon DPA Pro» для измерения длины. Следует подчеркнуть, что HDP – сертифицированная измерительная система, по стандарту VDI / VDE 2634, часть 1. Это гарантирует объемную точность измерений в указанных красной линией пределах. Таким образом, модель Браун & $\rho_{0,05}$ дает результаты в пределах погрешности HDP.

В HDP в задаче оптимизации пучка использовались оба масштабных эталона. Для всех результатов, полученных авторами, в целевой функции (19) включалась невязка длины только большего из эталонов (1390мм, $N_{scales} = 1$), эталон длиной 790 мм использовался для дополнительной верификации. Увеличивая w_{scale} в (19), можно без труда добиться нулевой погрешности длины эталона ценой проективного искажения результирующего облака точек.

В таблице приводится сравнение отклонений расстояний между парами точек для двух облаков аналогично рис. 12, но при этом алгоритм общий, а данные от разных дней. Таким образом, можно оценить повторяемость результатов – ключевую характеристику любой измерительной системы и фотограмметрии в частности. В сцене присутствовали два жестких объекта – стальной сварочный стол и карбоновая калибровочная плита, повторяемость оценивается отдельно для каждого подмножества точек. Столбцы σ_{table} и σ_{plate} показывают корень среднеквадратической ошибки (англ. «root mean squared error») для точек стола и плиты соответственно. Также представлена наибольшая ошибка измерения длины масштабного жезла (790 мм) за два дня e_{bar} .

Как видно из таблицы, повторяемость измерений небольшой калибровочной плиты σ_{plate} почти идентична (за одним исключением) и не зависит от выбора конфигурации алгоритма. Для стола отличие может быть существенным – наилучшая повторяемость достигается при оптимизации с подавлением выбросов: Браун & $\rho_{0,05}$ и OpenCV & $\rho_{0,05}$ дают лучшие результаты в своих группах. К тому же без стабилизации в конфигурациях Браун и OpenCV наблюдаются существенные ошибки в измерениях эталона e_{bar} . Также резуль-

таты подтверждают необходимость учета тангенциальной дисторсии (13) в задаче оптимизации пучка даже для высококлассного объектива. Особенно отметим, что Браун & $\rho_{0,05}$ существенно превосходит НДР по σ_{table} ! В итоге следует подчеркнуть, что основными факторами высокоточных результатов являются аккуратность выделения координат текстурных признаков и учет ошибок выделения этих признаков в оптимизационной задаче (18).

8. Заключение

В рамках данной работы был представлен широкий обзор алгоритмов, составляющих фотограмметрический конвейер. Продемонстрированы ключевые факторы, влияющие на точность определения выходного облака 3D-точек. Представлена математическая формулировка оптимизационной задачи фотограмметрии для различных моделей камер. В экспериментальной части подтверждены высокие показатели точности для описанных математических моделей на реальном устройстве и сценарии применения.

Авторы выражают свою искреннюю признательность сотрудникам и руководству компании АО «Цифровая сборка» (г. Санкт-Петербург) в помощи с организацией эксперимента. Благодаря профессионализму инженеров компании удалось собрать качественные данные для проверки различных гипотез, а также получить результаты, не уступающие уровню мировых лидеров в области фотограмметрии. Авторы готовы предоставить исходные данные (изображения, полученные облака точек) по запросу.

СПИСОК ЛИТЕРАТУРЫ

1. *Финстервальдер С.* Геометрические основы фотограмметрии (оригинал на немецком) // Ежегодник Немецкого математического общества. 1897. Т. 2. № 6. С. 1–41.
2. *Brown D.* Decentering Distoriton of Lenses // Photogrammetric Engineering. 1966. V. 32. No. 3. P. 444–462.
URL: <https://api.semanticscholar.org/CorpusID:117271607>
3. *Brown D.* Close-Range Camera Calibration // Photogrammetric Engineering. 1971. V. 37. No. 8. P. 855–866.
URL: <https://api.semanticscholar.org/CorpusID:56490170>
4. *Fryer J., Brown D.* Lens distortion for close-range photogrammetry // Photogrammetric Engineering and Remote Sensing. 1986. V. 52. P. 51–58.
5. *Drobyshev F.V.* Soviet stereophotogrammetric instruments // Photogrammetria. 1960. V. 17. P. 60–68.
6. *Дубиновский В.Б., Буров Ю.Л., Бергер Н.Я., Портнова О.В.* Строгий способ построения фотограмметрических сетей при обновлении топографических карт // Известия высших учебных заведений. Геодезия и аэрофотосъемка. 1990. № 6. С. 68–72.
7. *Алчинов А.И.* Фирма «ТАЛКА-ТДВ» и цифровая фотограмметрическая станция «ТАЛКА» // Геопрофи. 2005. № 1. С. 10–11.

8. *Luhmann T.* Close range photogrammetry for industrial applications // *ISPRS Journal of Photogrammetry and Remote Sensing*. 2010. V. 65. No. 6. P. 558–569. ISPRS Centenary Celebration Issue.
URL: <https://www.sciencedirect.com/science/article/pii/S0924271610000584>
9. *Bösemann W.* Industrial Photogrammetry – Accepted Metrology Tool or Exotic Niche // *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 2016. V. XLI-B5. P. 15–24.
URL: <https://isprs-archives.copernicus.org/articles/XLI-B5/15/2016/>
10. *Алтухов В.Г.* Исследование точности фотограмметрии как метода определения объема объекта // *Автоматика и программная инженерия*. 2020. Т. 32. № 2. С. 69–74.
11. *Leizea I., Herrera I., Puerto P.* Calibration Procedure of a Multi-Camera System: Process Uncertainty Budget // *Sensors*. 2023. V. 23. No. 2.
URL: <https://www.mdpi.com/1424-8220/23/2/589>
12. *Balanji H.M., Turgut A.E., Tunc L.T.* A novel vision-based calibration framework for industrial robotic manipulators // *Robot. Comput.-Integr. Manuf. USA*. 2022. V. 73. No. C. 15 p. <https://doi.org/10.1016/j.rcim.2021.102248>
13. *Puerto P., Leizea I., Herrera I., et al.* Analyses of Key Variables to Industrialize a Multi-Camera System to Guide Robotic Arms // *Robotics*. 2023. V. 12. P. 10–22.
14. *Li Z., Li S., Qin W., et al.* A Robust Camera Self-calibration Method Based on Circular Oblique Images // *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 2024. V. X-1-2024. P. 131–136.
URL: <https://isprs-annals.copernicus.org/articles/X-1-2024/131/2024/>
15. *Ulrich M., Steger C., Butsch F., et al.* Vision-guided robot calibration using photogrammetric methods // *ISPRS Journal of Photogrammetry and Remote Sensing*. 2024. V. 218. P. 645–662.
URL: <https://www.sciencedirect.com/science/article/pii/S0924271624003757>
16. *Визильтер Ю.В., Горбачевич В.С., Рубис А.Ю., Выголов О.В.* Сравнение изображений по форме с использованием диффузной морфологии и диффузной корреляции // *Компьютерная оптика*. 2015. Т. 39. № 2. С. 265–274.
17. *Fraser C.S.* Automatic camera calibration in close-range photogrammetry // *Photogrammetric Engineering and Remote Sensing*. 2013. V. 79. P. 381–388.
URL: <https://api.semanticscholar.org/CorpusID:19013771>
18. *Lebedev M.A., Stepaniants D.G., Komarov D.V., et al.* A real-time photogrammetric algorithm for sensor and synthetic image fusion with application to aviation combined vision // *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 2014. V. XL-3. P. 171–175.
URL: <https://isprs-archives.copernicus.org/articles/XL-3/171/2014/>
19. *Knyaz V., Knyaz V., Zheltov S., et al.* Multi-sensor Data Analysis for Aerial Image Semantic Segmentation and Vectorization // *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 2024. V. XLVIII-1-2024. P. 291–296.
URL: <https://isprs-archives.copernicus.org/articles/XLVIII-1-2024/291/2024/>
20. *Kobzev A., Chibunichev A.* Aerial Triangulation Using Different Time Images of Urban Areas Obtained from Unmanned Aerial Systems // *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*. 2024. V. XLVIII-2/W5-2024. P. 87–93.
URL: <https://isprs-archives.copernicus.org/articles/XLVIII-2-W5-2024/87/2024/>

21. *Schonberger J.L., Frahm J.-M.* Structure-from-Motion Revisited // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. P. 4104–4113.
22. *Turki H., Ramanan D., Satyanarayanan M.* Mega-NeRF: Scalable Construction of Large-Scale NeRFs for Virtual Fly-Throughs. 2022.
URL: <https://arxiv.org/abs/2112.10703>
23. *Lowe D.* Distinctive Image Features from Scale-Invariant Keypoints // Int. J. Comput. Vision. USA. 2004. V. 60. No. 2. P. 91–110.
URL: <https://doi.org/10.1023/B:VISI.0000029664.99615.94>
24. *Bay H., Tuytelaars T., Van Gool L.* SURF: Speeded Up Robust Features // Computer Vision – ECCV 2006 / Ed. by A. Leonardis, H. Bischof, A. Pinz. Berlin, Heidelberg: Springer Berlin Heidelberg. 2006. P. 404–417.
25. *DeTone D., Malisiewicz T., Rabinovich A.* SuperPoint: Self-Supervised Interest Point Detection and Description // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). 2018. P. 337–348.
26. *Hartley R., Zisserman A.* Multiple View Geometry in Computer Vision. 2 edition. New York, NY, USA: Cambridge University Press. 2003.
URL: <https://www.bibsonomy.org/bibtex/24867b22ef159cc28fc1bbe1476a0ec57/daill>
27. *Sola J., Deray J., Atchuthan D.* A micro Lie theory for state estimation in robotics. 2021. URL: <https://arxiv.org/abs/1812.01537>
28. *Rieke-Zapp D., Tecklenburg W., Peipe J., et al.* Evaluation of the geometric stability and the accuracy potential of digital cameras - Comparing mechanical stabilisation versus parameterisation // ISPRS Journal of Photogrammetry and Remote Sensing. 2009. V. 64. No. 3. P. 248–258. Theme Issue: Image Analysis and Image Engineering in Close Range Photogrammetry.
URL: <https://www.sciencedirect.com/science/article/pii/S0924271608001068>
29. *Tareen S.A.K., Saleem Z.* A comparative analysis of SIFT, SURF, KAZE, AKAZE, ORB, and BRISK // 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET). 2018. P. 1–10.
30. *Jin Y., Mishkin D., Mishchuk A., et al.* Image Matching Across Wide Baselines: From Paper to Practice // Int. J. Comput. Vision. USA. 2021. V. 129. No. 2. P. 517–547. <https://doi.org/10.1007/s11263-020-01385-0>
31. *Sun J., Shen Z., Wang Y., et al.* LoFTR: Detector-Free Local Feature Matching with Transformers // CVPR. 2021. V. 1. P. 8918–8927.
32. *Lindemberger P., Sarlin P.-E., Pollefeys M.* LightGlue: Local Feature Matching at Light Speed // 2023 IEEE/CVF International Conference on Computer Vision (ICCV). V. 1. 2023. P. 17581–17592.
33. *Vizilter Y., Zheltov S., Lebedev M.* Image and shape comparison via morphological correlation // The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. 2021. V. XLIV-2/W1-2021. P. 207–211.
34. *Sarlin P.-E., DeTone D., Malisiewicz T., et al.* SuperGlue: Learning Feature Matching With Graph Neural Networks // 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019. P. 4937–4946.
URL: <https://api.semanticscholar.org/CorpusID:208291327>
35. *Harris C.G., Stephens M.J.* A Combined Corner and Edge Detector // Alvey Vision Conference. 1988. URL: <https://api.semanticscholar.org/CorpusID:1694378>

36. *Shi J., Tomasi C.* Good Features to Track // Proceedings / CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2000. V. 600.
37. *Rosten E., Drummond T.* Machine Learning for High-Speed Corner Detection // Computer Vision – ECCV 2006 / Ed. by A. Leonardis, H. Bischof, A. Pinz. Berlin, Heidelberg: Springer Berlin Heidelberg. 2006. P. 430–443.
38. *Tyszkiewicz M.J., Fua P., Trulls E.* DISK: Learning local features with policy gradient. 2020. URL: <https://arxiv.org/abs/2006.13566>
39. *Triggs B., McLauchlan P.F., Hartley R.I., et al.* Bundle Adjustment – A Modern Synthesis // Proceedings of the International Workshop on Vision Algorithms: Theory and Practice. ICCV’99. Berlin, Heidelberg: Springer-Verlag. 1999. P. 298–372.
40. *Fitzgibbon A.* Simultaneous linear estimation of multiple view geometry and lens distortion // Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. V. 1. 2001. P. I–125.
41. *Zhang Z.* A Flexible New Technique for Camera Calibration // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2000. V. 22. P. 1330–1334. MSR-TR-98-71, Updated March 25, 1999. URL: <https://www.microsoft.com/en-us/research/publication/a-flexible-new-technique-for-camera-calibration/>
42. *Wang J., Karaev N., Rupprecht C., et al.* VGGsFM: Visual Geometry Grounded Deep Structure From Motion // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024. P. 21686–21697.
43. *Gurdjos P., Sturm P., Wu Y.* Euclidean Structure from $N \geq 2$ Parallel Circles: Theory and Algorithms // Computer Vision – ECCV 2006 / Ed. by A. Leonardis, H. Bischof, A. Pinz. Berlin, Heidelberg: Springer Berlin Heidelberg. 2006. P. 238–252.
44. *Gennery D.* Generalized Camera Calibration Including Fish-Eye Lenses // International Journal of Computer Vision. 2006. V. 68. P. 239–266.
45. *Schops T., Larsson V., Pollefeys M., et al.* Why Having 10 000 Parameters in Your Camera Model Is Better Than Twelve // 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Los Alamitos, CA, USA: IEEE Computer Society. 2020. P. 2532–2541. <https://doi.ieeecomputersociety.org/10.1109/CVPR42600.2020.00261>
46. *Canny J.* A Computational Approach To Edge Detection // Pattern Analysis and Machine Intelligence, IEEE Transactions on. 1986. V. PAMI-8. P. 679–698.
47. *Shen Z., Zhao M., Jia X., et al.* Combining convex hull and directed graph for fast and accurate ellipse detection // Graphical Models. 2021. V. 116. P. 101–110. URL: <https://www.sciencedirect.com/science/article/pii/S1524070321000151>
48. *Ouellet J.-N., Hebert P.* Precise ellipse estimation without contour point extraction // Mach. Vis. Appl. 2009. V. 21. P. 59–67.
49. *Mortari D., Junkins J., Samaan M.* Lost-in-Space Pyramid Algorithm for Robust Star Pattern Recognition // Spaceflight Mechanics 2005. 2001. V. 120. P. 10–20.
50. *Calonder M., Lepetit V., Strecha C., et al.* BRIEF: Binary Robust Independent Elementary Features // Computer Vision – ECCV 2010 / Ed. by K. Daniilidis, P. Maragos, N. Paragios. Berlin, Heidelberg: Springer Berlin Heidelberg. 2010. P. 778–792.
51. *Rublee E., Rabaud V., Konolige K., et al.* ORB: An efficient alternative to SIFT or SURF // 2011 International Conference on Computer Vision. 2011. P. 2564–2571.

52. *Fernandez Alcantarilla P.* Fast Explicit Diffusion for Accelerated Features in Non-linear Scale Spaces // British Machine Vision Conference (BMVC) at Bristol, UK. 2013.
53. *Schneider C.T., Sinnreich K.* Optical 3-D Measurement Systems for Quality Control in Industry // XVIIth ISPRS Congress Technical Commission V: Close-Range Photogrammetry and Machine Vision. 1992. V. 29. P. 56–59.
54. *dos Santos Cesar D. B., Gaudig C., Fritsche M., et al.* An evaluation of artificial fiducial markers in underwater environments // OCEANS 2015 – Genova. 2015. P. 1–6.
55. *Calvet L., Gurdjos P., Griwodz C., et al.* Detection and Accurate Localization of Circular Fiducials under Highly Challenging Conditions // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. P. 562–570.
56. *Tushev S., Sukhovilov B., Sartasov E.* Robust Coded Target Recognition in Adverse Light Conditions // 2018 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM). 2018. P. 1–6.
57. *Lin T.-Y., Dollar P., Girshick R., et al.* Feature Pyramid Networks for Object Detection // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017. P. 936–944.
58. *He K., Zhang X., Ren S., et al.* Deep Residual Learning for Image Recognition // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. P. 770–778.
59. *Ronneberger O., Fischer P., Brox T.* U-net: Convolutional networks for biomedical image segmentation // Medical image computing and computer-assisted intervention – MICCAI 2015: 18th international conference, Munich, Germany, October 5–9, 2015, proceedings, part III 18 / Springer. 2015. P. 234–241.
60. *Caetano T.S., Cheng L., Le Q.V., et al.* Learning Graph Matching // 2007 IEEE 11th International Conference on Computer Vision. 2007. P. 1–8.
61. *Arandjelovic R., Zisserman A.* Three things everyone should know to improve object retrieval // 2012 IEEE Conference on Computer Vision and Pattern Recognition. 2012. P. 2911–2918.
62. *Rusu R.B., Blodow N., Beetz M.* Fast Point Feature Histograms (FPFH) for 3D registration // 2009 IEEE International Conference on Robotics and Automation. 2009. P. 3212–3217.
63. *Muja M., Lowe D.* Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration // VISAPP 2009 – Proceedings of the Fourth International Conference on Computer Vision Theory and Applications, Lisboa, Portugal, February 5–8. V. 1. 2009. P. 331–340.
64. *Bishop C.* Pattern Recognition and Machine Learning (Information Science and Statistics). Berlin, Heidelberg: Springer-Verlag, 2006.
65. *Lu Q., Josephson W., Wang Z., et al.* Multi-probe LSH: efficient indexing for high-dimensional similarity search // Proceedings of the 33rd International Conference on Very Large Data Bases. VLDB '07. VLDB Endowment. 2007. P. 950–961.
66. *Nister D., Stewenius H.* Scalable Recognition with a Vocabulary Tree // 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06). V. 2. 2006. P. 2161–2168.

67. *Hartley R., Kang S.B.* Parameter-Free Radial Distortion Correction with Centre of Distortion Estimation // Proceedings / IEEE International Conference on Computer Vision. IEEE International Conference on Computer Vision. V. 2. 2005. P. 1834–1841.
68. *Hartley R.* In defense of the eight-point algorithm // IEEE Transactions on Pattern Analysis and Machine Intelligence. 1997. V. 19. No. 6. P. 580–593.
69. *Fischler M.A., Bolles R.C.* Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography // Readings in Computer Vision / Ed. by M. A. Fischler, O. Firschein. San Francisco (CA): Morgan Kaufmann. 1987. P. 726–740.
URL: <https://www.sciencedirect.com/science/article/pii/B9780080515816500702>
70. *Chum O., Matas J.* Matching with PROSAC – progressive sample consensus // 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05). V. 1. 2005. P. 220–226.
71. *Chum O., Werner T., Matas J.* Two-view geometry estimation unaffected by a dominant plane // 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05). V. 1. 2005. P. 772–779.
72. *Kaehler A., Bradski G.* Learning OpenCV, 2nd Edition. O’Reilly Media, Inc... 2014.
73. *Chibunichev A., Govorov A., Chernyshev V.* Research of the Camera Calibration Using Series of Images with Common Center of Projection // The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. 2019. V. XLII-2/W18. P. 19–22.
URL: <https://isprs-archives.copernicus.org/articles/XLII-2-W18/19/2019/>
74. *Grossberg M., Nayar S.* A general imaging model and a method for finding its parameters // Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001. V. 2. 2001. P. 108–115.
75. *Tecklenburg W., Luhmann T., Hastedt H.* Camera Modelling with Image-variant Parameters and Finite Elements // Optical 3D-Measurement Techniques V. 2001. P. 328–335.
76. *Hughes C., McFeely R., Denny P., et al.* Equidistant ($f\theta$) fish-eye perspective with application in distortion centre estimation // Image and Vision Computing. 2010. V. 28. No. 3. P. 538–551.
URL: <https://www.sciencedirect.com/science/article/pii/S0262885609001863>
77. *Bukhari F., Dailey M.N.* Automatic Radial Distortion Estimation from a Single Image // Journal of Mathematical Imaging and Vision. 2013. V. 45. No. 1. P. 31–45.
78. *Triggs B.* Autocalibration from Planar Scenes // Proceedings of the 5th European Conference on Computer Vision. V. 1 of ECCV’98. Berlin, Heidelberg: Springer-Verlag. 1998. P. 89–105.
79. *Nister D.* An efficient solution to the five-point relative pose problem // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2004. V. 26. No. 6. P. 756–770.
80. *Larsson, Viktor and contributors.* PoseLib – Minimal Solvers for Camera Pose Estimation. 2020. URL: <https://github.com/vlarsson/PoseLib>
81. *Agarwal S., Mierle K., The Ceres Solver Team.* Ceres Solver. 2023.
URL: <https://github.com/ceres-solver/ceres-solver>
82. *Bradbury J., Frostig R., Hawkins P., et al.* JAX: composable transformations of Python+NumPy programs. 2018. URL: <http://github.com/jax-ml/jax>

83. *Virtanen P., Gommers R., Oliphant T.E., et al.* SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python // Nature Methods. 2020. V. 17. P. 261–272.
URL: <https://rdcu.be/b08Wh>
84. *Blender Online Community.* Blender – a 3D modelling and rendering package. Blender Foundation. Stichting Blender Foundation, Amsterdam. 2024.
URL: <http://www.blender.org>

Статъя представена к публикации членом редколлегии Б.М. Миллером.

Поступила в редакцию 26.01.2025

После доработки 04.07.2025

Принята к публикации 08.07.2025