

Создание панорамных карт подводного дна на основе больших массивов изображений

Камаев Александр
Вычислительный центр ДВО РАН, Хабаровск, Россия
kamaev_an@mail.ru

Аннотация

Подводное дно является интересным для изучения, но труднодоступным объектом. Одним из способов получения сведений о поверхности дна, о флоре и фауне, а также о состоянии инженерных коммуникаций, проложенных на дне, является составление его панорамных карт. Массивы изображений для составления подобных карт могут быть получены с автономных необитаемых подводных аппаратов АНПА. При этом возникает ряд трудностей, связанных с плохими условиями освещённости, характерными особенностями дна, отсутствием точной gps навигации под водой. В статье рассматриваются все факторы, затрудняющие составление панорамных карт дна и предлагаются пути решения возникающих проблем.

Ключевые слова: Панорама, АНПА, Камера, Особые Точки, Метод Наименьших Квадратов, Bundle Adjustment.

1. ВВЕДЕНИЕ

Рассмотрим АНПА, оснащённый камерой с известной матрицей калибровки $\mathbf{K}=\text{diag}(f_x, f_y, 1)$ [5], жёстко связанной с аппаратом и направленной вниз. Элементы f_x и f_y – фокусные расстояния в пикселях (если пиксели имеют одинаковую высоту и ширину, то $f_x=f_y$). АНПА совершает проходы над дном, непрерывно его фотографируя. На момент окончания миссии получается m фотографий. При этом координаты аппарата $\mathbf{p}_j=(t_{jx}, t_{jy}, t_{jz})$, $j=1, 2, \dots, m$ в момент получения каждой фотографии неизвестны. Известными является курс аппарата γ_j , его глубина d_j и высота над дном h_j . Как правило, эти величины доступны из навигационной системы любого АНПА с известной точностью. Координаты \mathbf{p}_j также могут быть представлены системой навигации АНПА, однако их точность обычно неудовлетворительна для использования в задаче построения панорамы.

В процессе фотографирования дна крен α_j и дифферент β_j колеблются около нулевых значений, часто эти колебания слишком малы для точной регистрации, но могут оказаться существенными при построении панорамы, поэтому данные параметры будем также считать неизвестными.

Задача заключается в определении точных значений векторов $\mathbf{p}_j=(p_{jx}, p_{jy}, p_{jz})$ и $\mathbf{r}_j=(\alpha_j, \beta_j, \gamma_j)$ на основе визуальных особенностей, присутствующих на фотографиях и известных данных γ_j , d_j и h_j . Вектора \mathbf{p}_j и \mathbf{r}_j могут быть использованы для точного позиционирования каждой фотографии в панорамной карте дна.

2. СПЕЦИФИКА ЗАДАЧИ

У рассматриваемой задачи есть ряд специфических особенностей, которые затрудняют её решение. Эти особенности, так или иначе, вытекают из плохих условий съёмки (вода может быть мутной, свет может не проникать на большую глубину), из специфики подводных объектов и из

отсутствия возможности точного определения координат \mathbf{p}_j с помощью GPS, как это возможно делать при аэрофотосъёмке.

Из-за отсутствия света под водой приходится использовать искусственное освещение, что приводит к неравномерной освещённости каждого снимка. Это в свою очередь затрудняет процесс сопоставления. Искусственное освещение и мутная вода способствуют получению слабоконтрастных, нечётких снимков. Также задачу осложняет тот факт, что на дне часто встречается большое количество совершенно одинаковых объектов: ракушки, морские ежи, водоросли и т.д. Всё это приводит к низкой различимости разных визуальных особенностей на снимках.

Отсутствие естественного света делает невозможным съёмку с большой высоты h_j , высота съёмки ограничивается мощностью прожекторов АНПА. Следовательно, зона покрытия каждого снимка оказывается небольшой и для покрытия больших территорий требуются тысячи и даже десятки тысяч снимков. При этом точное положение АНПА \mathbf{p}_j в момент получения каждого снимка остаётся неизвестным. В результате получается задача с огромным количеством изображений и очень слабо выраженными и плохо различимыми визуальными особенностями.

Кроме негативных факторов, усложняющих задачу, присутствует ряд особенностей, делающих решение поставленной задачи возможным. К таким факторам относятся известные величины γ_j , d_j и h_j , а также близкая к плоской траектория движения АНПА. Используя данные особенности, в следующих разделах поэтапно рассмотрим процесс решения задачи построения панорамы.

3. СОПОСТАВЛЕНИЕ ПАР ИЗОБРАЖЕНИЙ

В основе задачи построения панорамы лежит задача определения связи двух изображений. Рассмотрим эту задачу для пары изображений i, j , $j=1, 2, \dots, m$. Для того, чтобы определить как связаны i -е и j -е изображения необходимо на этих изображениях выделить визуальные особенности. Устойчивыми особенностями являются особые точки SIFT [4]. Дескрипторы SIFT [4] устойчивы к плоскому повороту, масштабу, перемещению изображения и к линейному изменению освещённости. Однако эти дескрипторы довольно чувствительны к нелинейным изменениям освещённости, которые имеют место быть при подводной съёмке – рис. 1а.

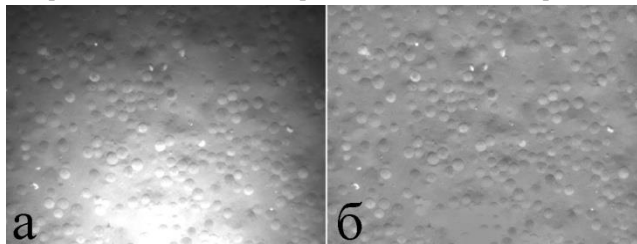


Рис 1: Пример компенсации нелинейной освещённости.

По этой причине до составления дескрипторов, необходимо выровнять освещение на всех снимках, воспользовавшись алгоритмом Single Scale Retinex [6]. Пример применения алгоритма к изображению на рис. 1а показан на рисунке 1б.

Далее в результате сравнения евклидовых расстояний между всеми возможными парами дескрипторов особых точек i -го и j -го изображений получим список пар точек, расстояние между которыми меньше определённого порога. В качестве порога в случае нормализованных дескрипторов можно выбрать расстояние, до 0.5. Среди сопоставленных точек будет очень высокий процент ложных соответствий. Из-за низкой контрастности и чёткости снимков, а также повторяющихся однотипных объектов на дне, различимость дескрипторов становится низкой, и количество ложных соответствий может достигать 90% (рис. 2а).

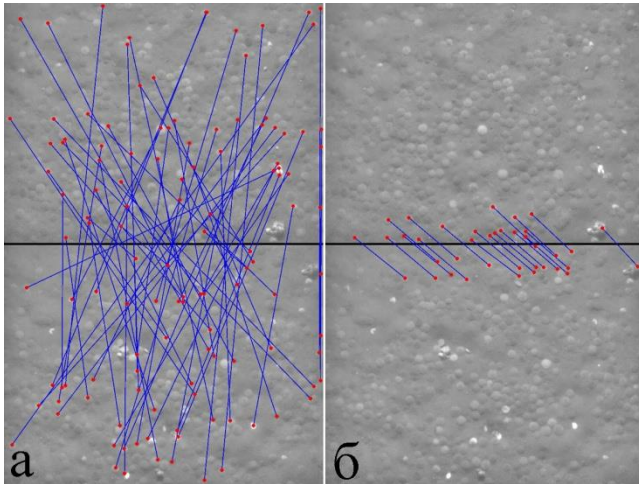


Рис 2: Фильтрация ложных соответствий.

Для фильтрации ложных соответствий обычно используется процедура RANSAC [10], однако, она становится неэффективной, когда количество ложных соответствий превышает 60%. Чтобы отфильтровать ложные соответствия в рассматриваемом случае введём такую модель связи, которая содержит минимум параметров.

Так как перемещение АНПА над дном в первом приближении может считаться плоским, для приблизительной оценки связи изображений ограничимся следующей моделью:

$$\begin{pmatrix} x_i \\ y_i \end{pmatrix} = \frac{h_j}{h_i} \cdot \begin{pmatrix} \cos(\gamma_j - \gamma_i) & \sin(\gamma_j - \gamma_i) \\ -\sin(\gamma_j - \gamma_i) & \cos(\gamma_j - \gamma_i) \end{pmatrix} \cdot \begin{pmatrix} x_j \\ y_j \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix}.$$

Здесь $(x_i, y_i)^T$ и $(x_j, y_j)^T$ – координаты соответствующих точек на i -ом и j -ом изображениях, а $\mathbf{t}_{ij} = (t_x, t_y)$ – вектор смещения центра j -го изображения относительно центра i -го изображения. Неизвестный вектор \mathbf{t}_{ij} имеет масштаб i -го изображения и может быть найден на основе одного точечного соответствия. Если перебрать все точечные соответствия и для каждого построить модель связи, а затем из всех моделей выбрать согласующуюся с наибольшим количеством точек, то выбранная модель и будет истинной, в случае если изображения i и j действительно связаны. Определить, существует ли связь изображений, можно воспользовавшись следующим условием:

$$n_i > \alpha + \beta \cdot n_f$$

Здесь n_i – количество соответствий, удовлетворяющих выбранной модели, а n_f – количество соответствий, находящихся в зоне пересечения изображений, но не

удовлетворяющих модели связи. Коэффициент $\alpha=2$, а $\beta=0.3$ [11]. Невыполнение условия говорит об отсутствии связи.

Если связь изображений подтвердилась, то процедура поиска соответствующих особенностей повторяется заново, при этом учитывается найденная модель связи. В результате повторного поиска, как правило, находятся дополнительные соответствия, найти которые без учёта модели связи было невозможно, а ложные соответствия отфильтровываются – рис. 2б.

4. ПОИСК СВЯЗАННЫХ ПАР ИЗОБРАЖЕНИЙ

Поиск связанных пар изображений – самый затратный по времени этап. Очевидно, что поиск путём полного перебора всех возможных пар и применения к ним алгоритма из раздела 3, подходит лишь для небольшого количества изображений. В условиях, когда большое количество изображений является одной из характеристик решаемой задачи, такой поиск «грубой силой» не применим.

Ключом к организации быстрого поиска связанных пар изображений является процедура поиска похожих дескрипторов в пространстве с большой размерностью. Предположим, что имеется некоторый дескриптор особенности \mathbf{q} , необходимо из множества дескрипторов всех особенностей выбирать дескрипторы, наиболее близкие к дескриптору \mathbf{q} . Это задача называется поиском k ближайших соседей и различные методы её решения, как точные, так и приближенные хорошо изучены. Наиболее известные методы для решения такой задачи в пространствах с высокой размерностью – это «best bin first» [2], локально-чувствительное хеширование (LSH) [7], метрические деревья [3], sr -деревья и гибридные sr -деревья [8]. Гибридные sr -деревья, показавшие наилучшую производительность, были использованы в данной работе, совместно с приоритетной очередью из алгоритма «best bin first».

Из-за того, что дескрипторы особенностей в рассматриваемой задаче имеют низкую различимость (слабо отличаются друг от друга), группы похожих дескрипторов, найденных с помощью гибридных sr -деревьев нельзя напрямую использовать для вычисления связей между изображениями. Но информация о похожих дескрипторах может быть использована для определения групп изображений, которые могут пересекаться. А связи в этих группах могут быть рассчитаны с помощью метода из раздела 3.

Обозначим дескриптор j -ой особенности i -го изображения за \mathbf{v}_{ij} , $i=1, 2, \dots, m$, $j=1, 2, \dots, \lambda(i)$, где $\lambda(i)$ – количество особенностей на i -ом изображении. С помощью гибридного sr -дерева найдём вектор \mathbf{l}_{ij} , состоящий из номеров изображений, содержащих первые k особенностей с дескрипторами, максимально близкими к дескриптору \mathbf{v}_{ij} : $\mathbf{l}_{ij} = (l_1^{ij}, l_2^{ij}, \dots, l_k^{ij})$. Затем составим списки предполагаемых кандидатов на сопоставление для каждого изображения. Для этого введём матрицу $\mathbf{W} = (w_{pb})$, $p, b=1, 2, \dots, m$, где w_{pb} – количество похожих дескрипторов между парой изображений p и b . Изначально, все w_{pb} приравниваются к нулю. Далее перебираются все элементы l_n^{ij} , $i=1, 2, \dots, m$, $j=1, 2, \dots, \lambda(i)$, $n=1, 2, \dots, k$ и компоненты матрицы \mathbf{W} : $w_{i, l_n^{ij}}$ и $w_{l_n^{ij}, i}$ увеличиваются на единицу. В список предполагаемых кандидатов на сопоставление каждого изображения $\mathbf{S}_p = (s_{p1}, s_{p2}, \dots, s_{p\eta})$, $p=1, 2, \dots, m$ помещаются η номеров столбцов наибольших элементов из p -ой строки матрицы \mathbf{W} .

После вычисления списков S_p , процедуру сопоставления 3 нужно выполнять не m раз для каждого p -го изображения, а всего лишь η раз. Скорость и точность алгоритма можно варьировать, изменяя параметр η . Чем η выше, тем больше связанных пар может быть найдено, но тем медленнее будет происходить поиск. Также на скорость и точность алгоритма влияет параметр E – максимальное количество узлов гибридного sp-дерева, которое просматривается при поиске похожих дескрипторов. Влияния параметров E и η на точность и скорость поиска связанных пар изображений отображают графики на рис. 3а и 3б.

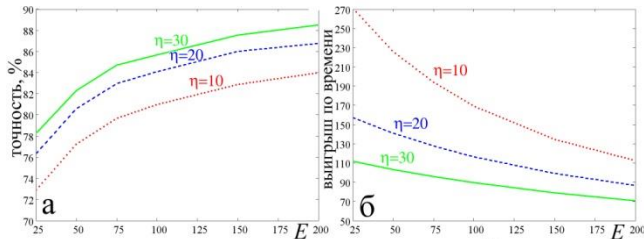


Рис 3: Влияние параметров η и E на процедуру поиска.

В качестве скорости работы алгоритма взят коэффициент выигрыша по времени, который даёт алгоритм в сравнении с поиском «грубой силой», а в качестве точности взят процент связанных пар изображений, которые удалось обнаружить от количества пар, обнаруженных поиском «грубой силой». Графики на рис 3 получены для задачи с $m=3591$ изображениями и с 2362523 дескрипторами.

Ещё одним параметром, влияющим на точность алгоритма, является k – количество похожих дескрипторов, которые нужно находить для каждого дескриптора – запроса. Функция зависимости точности сопоставления от k достигает своего максимума при $k=2$ независимо от значений других параметров. Поэтому целесообразно использовать для k именно это значение.

5. ПРИБЛИЗИТЕЛЬНОЕ ПОЗИЦИОНИРОВАНИЕ

Задача поиска точных значений векторов \mathbf{p}_j и \mathbf{r}_j , $j=1, 2, \dots, m$ по набору известных соответствующих проекций точек пространства является нелинейной, так как связь точки пространства и её проекции является нелинейной. Это означает, что необходимо использовать итерационные методы решения задачи, для которых требуется хорошее начальное приближение. Воспользовавшись тем, что движение АНПА преимущественно плоское, а колебания крена и дифферента аппарата не значительны, для поиска начального приближения будем решать двухмерную задачу, приняв крен и дифферент нулевыми.

После обнаружения всех связанных пар (раздел 4) и вычисления параметров связи \mathbf{t}_{ij} (раздел 3) задача приблизительного позиционирования может быть решена с помощью метода наименьших квадратов (МНК).

Будем считать, что было найдено n связанных пар изображений, а пары изображений с номерами $i(l)$ $j(l)$, где $l=1, 2, \dots, n$ – связаны, т.е. известны векторы $\mathbf{t}_{i(l)j(l)}$. Предполагается, что все m изображений оказались связаны (граф, построенный из вершин – изображений и рёбер – известных связей $i(l)$ $j(l)$ связан). Если это не так, то следует из множества изображений выделить подмножества связанных изображений и решать задачу приблизительного позиционирования для каждого подмножества отдельно.

Найдём положения центра каждого изображения $\mathbf{c}_k=(c_{kx}, c_{ky})$, $k=1, 2, \dots, m$ в системе координат некоторого изображения $q \in \{1, 2, \dots, m\}$. Переведём все векторы $\mathbf{t}_{i(l)j(l)}$ в систему координат q -го изображения, домножив их на масштаб между q -ым и $i(l)$ -ым изображением и повернув на угол $\gamma_{i(l)} - \gamma_q$:

$$\mathbf{d}_l = \frac{h_{i(l)}}{h_q} \cdot \begin{pmatrix} \cos(\gamma_{i(l)} - \gamma_q) & \sin(\gamma_{i(l)} - \gamma_q) \\ -\sin(\gamma_{i(l)} - \gamma_q) & \cos(\gamma_{i(l)} - \gamma_q) \end{pmatrix} \cdot \mathbf{t}_l$$

Здесь \mathbf{d}_l – вектор смещения центра изображения $j(l)$ относительно центра изображения $i(l)$, заданный в системе координат q -го изображения. Каждый вектор \mathbf{d}_l даёт два уравнения относительно неизвестных координат центров изображений \mathbf{c}_k : $\mathbf{d}_l = \mathbf{c}_{j(l)} - \mathbf{c}_{i(l)}$. Совокупность уравнений для $l=1, 2, \dots, n$ составляет систему линейных алгебраических уравнений. Приняв вектор $\mathbf{c}_q=(0, 0)$, разрешив данную систему, воспользовавшись линейным МНК [5].

Отметим, что матрица нормального уравнения МНК $\mathbf{Q}=\mathbf{A}^T\mathbf{A}$, где \mathbf{A} – матрица исходной системы, будет разреженной и к ней применимы методы эффективного решения из [12].

Зная координаты центров изображений \mathbf{c}_k , заданные в системе координат q -го изображения, легко получить планарные координаты АНПА p_{kx}, p_{ky} . Для этого достаточно выполнить следующие умножения:

$$p_{kx} = c_{kx} \cdot \frac{h_k}{f_x}, \quad p_{ky} = c_{ky} \cdot \frac{h_k}{f_y}$$

Поскольку никаких данных о привязке к глобальной системе координат нет, координаты АНПА определяются с точностью до некоторого вектора перемещения.

6. ТОЧНОЕ ПОЗИЦИОНИРОВАНИЕ

Заключительным этапом вычисления векторов \mathbf{p}_j и \mathbf{r}_j является процедура их уточнения. На начало процедуры уточнения приблизительные значения данных векторов известны. Обозначим эти значения через \mathbf{p}_j^0 и \mathbf{r}_j^0 . Значения p_{jx}^0 и p_{jy}^0 , составляющие вектор \mathbf{p}_j^0 были найдены в разделе 5, курс АНПА γ_j^0 известен из навигационной системы, а начальные крен и дифферент примем нулевыми: $\alpha_j^0 = 0, \beta_j^0 = 0$. Значение p_{jz}^0 , как и начальный курс может быть извлечено из навигационной системы: $p_{jz}^0 = d_j$.

Начальные значения \mathbf{p}_j^0 и \mathbf{r}_j^0 могут содержать ошибки, недопустимые при построении панорам, и требуют уточнения. Стандартным способом уточнения является использование метода bundle adjustment [5], работающего, обычно, совместно с алгоритмом Левенберга–Марквардта [9].

Для использования bundle adjustment, дополнительно к векторам параметров камеры $\mathbf{a}_j=(\mathbf{p}_j, \mathbf{r}_j)$ введём вектор точек пространства $\mathbf{b}_i=(x_i, y_i, z_i)$, $j=1, 2, \dots, m, i=1, 2, \dots, n$, где через n обозначим количество точек дна, видимых одновременно на двух и более изображениях. Евклидовы координаты проекции точек \mathbf{b}_i на j -ое изображение обозначим через $\mathbf{x}_i^j = (u_i^j, v_i^j)$, а однородные координаты проекций через $\tilde{\mathbf{x}}_i^j = (\tilde{u}_i^j, \tilde{v}_i^j, \tilde{w}_i^j)$, где $\tilde{w}_i^j = 1$. Перейти от однородных координат к евклидовым можно посредством деления первых двух компонент на третью: $u_i^j = \tilde{u}_i^j / \tilde{w}_i^j, v_i^j = \tilde{v}_i^j / \tilde{w}_i^j$.

Каждая точка \mathbf{b}_i связана с координатами своей проекции на j -ое изображение посредством параметров \mathbf{a}_j :

$$\tilde{\mathbf{x}}_i^j = \mathbf{K} \cdot (\mathbf{R}^T(\mathbf{r}_j) \quad -\mathbf{R}^T(\mathbf{r}_j) \cdot \mathbf{p}_j) \cdot (\mathbf{b}_i \quad 1)^T$$

Здесь $R(\mathbf{r}_j)$ – матрица поворота АНПА, являющаяся комбинацией трёх матриц поворота вокруг координатных осей: $R(\mathbf{r}_j) = R_z(\gamma_j) \cdot R_y(\beta_j) \cdot R_x(\alpha_j)$. Таким образом, можно говорить о некоторой функции проекции f_i^j : $\mathbf{x}_i^j = f_i^j(\mathbf{a}_j, \mathbf{b}_i)$, которая отображает параметры, подлежащие уточнению на координаты проекции i -ой точки на j -ом изображении.

С другой стороны, известны проекции каждой i -ой точки пространства (раздел 3), обозначим их через $\hat{\mathbf{x}}_i^j = (\hat{x}_i^j, \hat{y}_i^j)$.

Метод bundle adjustment заключается в итеративном уточнении параметров \mathbf{a}_j , \mathbf{b}_i , приводящем к уменьшению ошибки репроекции E :

$$E = \sum_{i=1}^n \sum_{j=1}^m |\hat{\mathbf{x}}_i^j - \mathbf{x}_i^j|^2$$

Сам метод и его эффективная реализация рассмотрена в [5]. Использование методик решения разреженных систем линейных уравнений [12] в совокупности с принципом разделения параметров на координаты точек \mathbf{b}_i и параметры камер \mathbf{a}_j описанного в [5] позволяет решать задачу уточнения для десятков тысяч изображений и миллионов точек.

В качестве начальных значений для параметров камеры следует взять \mathbf{p}_j^0 и \mathbf{r}_j^0 , а начальные значения для координат точек \mathbf{b}_i^0 , рассчитываются с использованием DLT метода [5].

7. СШИВКА ИЗОБРАЖЕНИЙ

Вопрос сшивки изображений подводного дна в единую панораму не имеет какой-либо специфики и может быть выполнен любым доступным способом.

Самый простой способ заключается в аппроксимации участка дна, представляемого каждым изображением плоскостью (выбирается плоскость с наименьшей суммой расстояний от этой плоскости до всех точек, видимых на изображении) и проекции этой плоскости на экран, параллельный горизонтальной плоскости дна. Пример карты куска дна, полученного таким образом, представлен на рисунке 4.

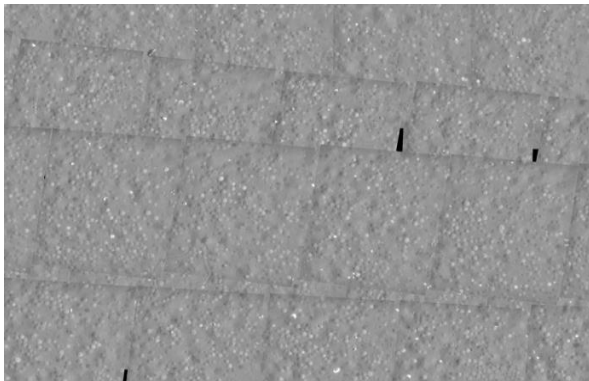


Рис 4: Участок панорамы подводного дна.

Более сложный метод заключается в триангуляции всех точек подводного дна и проецировании каждого треугольника на плоскость карты.

Для устранения швов на границах изображений, если это необходимо, можно использовать любой стандартный метода, например multi-band blending [1].

8. ЗАКЛЮЧЕНИЕ

Метод, представленный в данной статье, позволяет собирать панорамные карты подводного дна из десятков тысяч

изображений, сделанных в условиях плохой освещённости. При этом используются только визуальные особенности на самих изображениях и та информация навигационной системы, которая может быть получена с достаточной точностью большинством АНПА. Для работы метода не требуется знать координаты АНПА, получение которых обычно вызывает затруднения ввиду отсутствия GPS сигнала под водой.

Недостатком описанной техники является её неработоспособность в случае получения изображений без каких-либо деталей. Даже на гладком песчаном дне есть детали, однако, если большое число фотографий было выполнено с большого удаления, когда освещение не достаёт до дна, то в процессе вычисления положения снимков могут возникнуть проблемы. Задача может быть разбита на несколько несвязанных подзадач или вовсе не решена.

9. ССЫЛКИ

- [1] Allene C., Pons J.-P., Keriven R. Seamless image-based texture atlases using multi-band blending. *In Proc. of the ICPR 2008*, pp. 1-4.
- [2] Beis J., Lowe D.G. Shape indexing using approximate nearest-neighbor search in high-dimensional spaces. *In Proc. of CVPR. – Puerto Rico, 1997. – pp. 1000-1006.*
- [3] Ciaccia P., Patella M., Zezula P. M-tree: An efficient access method for similarity search in metric spaces. *In Proc. of the 23-rd VLDB International Conference, 1997. – pp. 426-435.*
- [4] David G. Lowe, Distinctive image features from scale-invariant keypoint. *International Journal of Computer Vision*, 60, 2 (2004), pp. 91-110.
- [5] Hartley R., Zisserman A. Multiple View Geometry in Computer Vision. – Cambridge University Press, 2004.
- [6] Hines G. D., Rahman Z., Jobson D. J., Woodell G. A. Single-Scale Retinex Using Digital Signal Processors. *Global Signal Processing Expo (GSPx)*, (2004).
- [7] Indyk. P., Motwani R. Approximate nearest neighbours: towards removing the curse of dimensionality. *In Proc. of the 30-th ACM Symposium on the STOC, 1998. – pp. 604-613.*
- [8] Liu T, Moore A., Gray A., Yang K. An investigation of practical approximate nearest neighbor algorithms. *In Proc. of the conference on NIPS, 2004. – P. 825-832.*
- [9] Marquard. D. An algorithm for least-squares estimation of nonlinear parameters. *SIAM J. Appl. Math.*, 1963, Vol. 11, pp. 431-441.
- [10] Martin A. Fischler, Robert C. Bolles. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Computations Of the ACM* 24: 381–395, 1981.
- [11] Matthew Brown, David G. Lowe, Automatic panoramic image stitching using invariant features. *International Journal of Computer Vision*, 74, 1 (2007), pp. 59-73.
- [12] Джордж А., Лю Дж. Численное решение больших разреженных систем уравнений: Пер. с англ. – М.: Мир, 1984.

Об авторе

Камаев Александр – аспирант ВЦ ДВО РАН.

Адрес email: kamaev_an@mail.ru.