(12) МЕЖДУНАРОДНАЯ ЗАЯВКА, ОПУБЛИКОВАННАЯ В СООТВЕТСТВИИ С ДОГОВОРОМ О ПАТЕНТНОЙ КООПЕРАЦИИ (PCT)

(19) Всемирная Организация Интеллектуальной Собственности
Международное бюро

(43) Дата международной публикации
19 октября 2023 (19.10.2023)

WIPO | PCT

(10) Номер международной публикации
WO 2023/200357 A1

(51) Международная патентная классификация:
G06T 7/593 (2017.01) *G06T 5/20* (2006.01)

(21) Номер международной заявки: PCT/RU2022/000123

(22) Дата международной подачи:
15 апреля 2022 (15.04.2022)

(25) Язык подачи: Русский

(26) Язык публикации: Русский

(30) Данные о приоритете:
2022110243 15 апреля 2022 (15.04.2022) RU

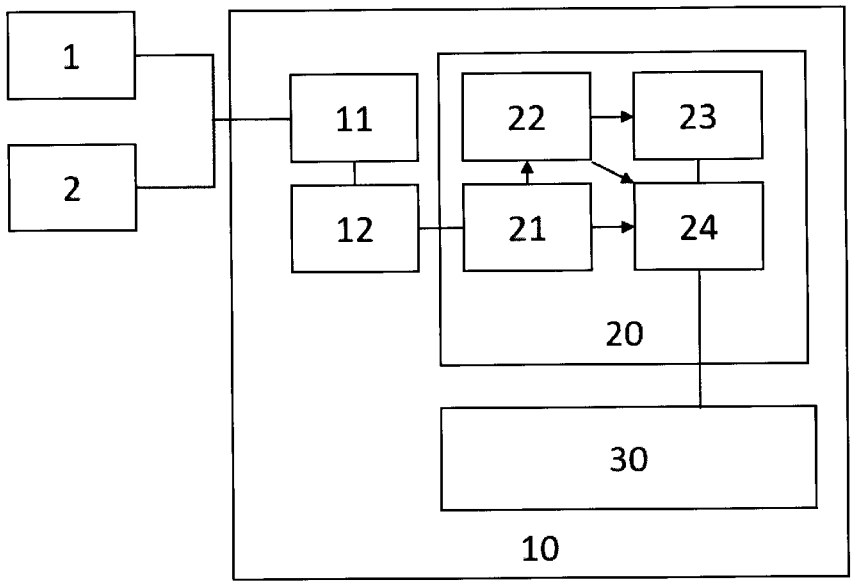
(71) Заявитель: **ОБЩЕСТВО С ОГРАНИЧЕННОЙ ОТВЕТСТВЕННОСТЬЮ "СБЕР АВТОМОТИВ ТЕХНОЛОГИИ"** (**OBSHCHESTVO S OGRANICHENNOJ OTVETSTVENNOST'YU "SBER AVTOMOTIV TEKHNOLOGII"**) [RU/RU]; проспект Андропова, 10А, Москва, 115432, Moscow (RU).

(72) Изобретатели: **МАСЛОВИЧ, Николай Романович (MASLOVICH, Nikolaj Romanovich)**; пр-кт Вернадского, д. 21, корп. 1, кв. 6, Москва, 119331, Moscow (RU). **ЯШУНИН, Дмитрий Александрович (YASHUNIN, Dmitrij Aleksandrovich)**; ул. Красноздвездная, д. 31, кв. 33, г. Нижний Новгород, 603104, g. Nizhnij Novgorod (RU). **ДЕРЕНДЯЕВ, Илья Васильевич (DERENDYAEV, Il'ya Vasil'evich)**; ул. Борисовские пруды, д. 34, корп. 1, кв. 451, Москва, 115408, Moscow (RU).

(74) Агент: **ГЕРАСИН, Борис Валерьевич и др. (GERASIN, Boris Valer'evich** et al.); Публичное акцио-

(54) **Title:** METHOD FOR CONSTRUCTING A DEPTH MAP FROM AN IMAGE PAIR

(54) **Название изобретения:** СПОСОБ ПОСТРОЕНИЯ КАРТЫ ГЛУБИНЫ ПО ПАРЕ ИЗОБРАЖЕНИЙ



ФИГ. 1

(57) **Abstract:** The proposed technical solution relates generally to the field of image data processing, and more particularly to a method and device for constructing a depth map from a pair of images obtained, for example, by means of a stereo camera, using a TUDE device. The technical result is an increase in the accuracy of the values of the depth map. This technical result is achieved using a method for constructing a depth map from an image pair which is implemented using at least one computing device and comprises the steps of: obtaining, from a first camera and a second camera, first and second images containing an image of at least one object; rectifying said first and second images by projecting them onto a single plane; determining, for each pixel of the first and second images, a shift value indicative of the number of pixels by which the most similar pixel of the other image has shifted; generating, for the first and

WO 2023/200357 A1

[продолжение на следующей странице]

нерное общество "Сбербанк России", Правовой департамент, ул. Вавилова, 19, Москва, 117997, Moscow (RU).

(81) Указанные государства *(если не указано иначе, для каждого вида национальной охраны)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Указанные государства *(если не указано иначе, для каждого вида региональной охраны)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), евразийский (AM, AZ, BY, KG, KZ, RU, TJ, TM), европейский патент (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Опубликована:

— *с отчётом о международном поиске (статья 21.3)*

second images, first and second shift maps containing said shift values; generating a depth map of the image on the basis of the values of the shift maps generated in the preceding step.

(57) Реферат: Представленное техническое решение относится, в общем, к области обработки данных изображения, а в частности к способу и устройству построения карты глубины по паре изображений, полученных, например, посредством стерео-камер, с использованием устройства TUDE. Техническим результатом является повышение точности значений карты глубины. Указанный технический результат достигается благодаря осуществлению способа построения карты глубины по паре изображений, выполняемого по меньшей мере одним вычислительным устройством, содержащего этапы, на которых: получают с первой и второй камер первое и второе изображения, содержащие изображение по меньшей мере одного объекта; выполняют процедуру ректификации первого и второго изображений посредством проецирования их в одну плоскость; определяют для каждого пикселя первого и второго изображений значение сдвига, указывающее на количество пикселей, на которое сдвинут наиболее похожий пиксель другого изображения; формируют первую и вторую карты сдвигов для первого и второго изображений, содержащие упомянутые значения сдвига; на основе значений карт сдвигов, сформированных на предыдущем этапе, формируют карту глубины изображения.

СПОСОБ ПОСТРОЕНИЯ КАРТЫ ГЛУБИНЫ ПО ПАРЕ ИЗОБРАЖЕНИЙ.

ОБЛАСТЬ ТЕХНИКИ

[0001] Представленное техническое решение относится, в общем, к области обработки данных изображения, а в частности к способу и устройству построения карты глубины по паре изображений, полученных, например, посредством стерео-камер с использованием устройства TUDE (Transformer-Unet for Depth Estimation, Трансформер-Юнет для Оценки Глубины).

УРОВЕНЬ ТЕХНИКИ

[0002] Существующие аналоги, наиболее близкие к представленному решению, можно условно разделить на два семейства:

1. Классические - использующие для построения карты глубины построчное сравнение окон со значениями яркости пикселей с кадров двух камер;

2. Нейросетевые - использующие для сравнения признаки, выделенные нейронной сетью.

[0003] Несмотря на относительно высокую скорость работы, решения, реализованные на основе классических подходов, имеют ряд недостатков:

- плохо работают в зонах с однородной текстурой (например, изображением автодороги);

- дают много пропусков в предсказаниях и часто ошибаются;

- точность предсказаний сильно зависит от размера объектов.

[0004] Большинство решений, доступных в открытом виде, основываются на методах библиотеки OpenCV. OpenCV (англ. Open Source Computer Vision Library) — библиотека алгоритмов компьютерного зрения, обработки изображений и численных алгоритмов общего назначения с открытым кодом.

[0005] Нейросетевые подходы, такие как AANet [см. AANet: Adaptive Aggregation Network for Efficient Stereo Matching], LEAStereo [см. LEAStereo: Learning Effective Architecture Stereo] значительно превосходят в качестве классические, однако работают медленнее. Это критично для применения в таких областях, как беспилотное вождение, поскольку это напрямую отражается на скорости реакции автопилота на препятствия. Также эти решения не имеют механизма отсечения предсказаний нейросети по пороговому значению уверенности, что приводит к артефактам на краях объектов и в областях объектов, которые видны на одной

камере и не видны на другой. Эти неточности способны негативно влиять на другие алгоритмы, работающие с картой глубины, что ухудшает работу алгоритма автопилота.

[0006] Для безопасного вождения беспилотного автомобиля требуется оперативная реакция на события дорожной сцены, например, внезапный выезд автомобиля со встречной полосы. При высокой скорости движения автомобиля быстрое принятие решений автопилотом становится критичным. Это мешает использованию алгоритмов, которые работают с данными от лидаров, частота работы которых ниже, чем частота работы камеры. Кроме того, полученное с лидара облако точек является сильно разреженным, вследствие чего по нему не всегда можно сделать правильный вывод о наличии и природе объекта.

[0007] Использование двух камер, объединенных в стерео-пару, позволяет быстро и относительно точно получить плотную карту глубины (каждому пикселю с камеры сопоставляется дистанция до объекта), что дает плотное покрытие 3Д точками небольших объектов [см. статью Smolyanskiy, Nikolai and Kamenev, Alexey and Birchfield, Stan «On the Importance of Stereo for Accurate Depth Estimation: An Efficient Semi-Supervised Deep Neural Network Approach», 2018.].

[0008] Предложенное техническое решение работает до пяти раз быстрее аналогичных нейросетевых подходов AANet и LEAStereo, работающих также с парой изображений, при сравнимой точности. Высокая скорость работы алгоритма достигается за счет использования легковесных слоев, работающих в относительно низком пространственном разрешении. Также предложенное решение позволяет фильтровать области с высокой ошибкой определения карты глубины.

РАСКРЫТИЕ ИЗОБРЕТЕНИЯ

[0009] Технической проблемой или задачей, поставленной в данном техническом решении, является создание нового эффективного, простого и надежного метода построения карты глубины по паре изображений.

[0010] Техническим результатом является повышение точности значений карты глубины.

[0011] Указанный технический результат достигается благодаря осуществлению способа построения карты глубины по паре изображений, выполняемого по меньшей мере одним вычислительным устройством, содержащего этапы, на которых:

- получают с первой и второй камер первое и второе изображения, содержащие изображение по меньшей мере одного объекта;

- выполняют процедуру ректификации первого и второго изображений посредством проецирования их в одну плоскость;

- определяют для каждого пикселя первого и второго изображений значение сдвига, указывающее на количество пикселей, на которое сдвинут наиболее похожий пиксель другого изображения;

- формируют первую и вторую карты сдвигов для первого и второго изображений, содержащие упомянутые значения сдвига;

- на основе значений карт сдвигов, сформированных на предыдущем этапе, формируют карту глубины изображения.

[0012] В одном из частных примеров осуществления способа этап определения значения сдвига для каждого пикселя первого и второго изображений содержит этапы, на которых:

- определяют значение яркости (величину освещенности) каждого пикселя первого и второго изображений;

- сопоставляют значения яркости пикселей первого изображения со значениями яркости пикселей второго изображения для определения значения сдвига для каждого пикселя первого и второго изображений, причем при сопоставлении учитывают также значения яркости соседних пикселей.

[0013] В другом частном примере осуществления способа дополнительно выполняют этапы проверки согласованности значений сдвигов пикселей левого и правого изображений.

[0014] В другом частном примере осуществления способа этап формирования карты глубины изображения на основе значений карт сдвигов содержит этапы, на которых:

- на основе значений сдвигов пикселей, содержащихся в картах сдвигов, определяют расстояние от линии, соединяющей центры камер, до каждого пикселя по меньшей мере одного объекта;

- на основе полученных значений расстояний от линии, соединяющей центры камер, до каждого пикселя по меньшей мере одного объекта, формируют карту глубины изображения.

[0015] В другом частном примере осуществления способа упомянутое расстояние определяется по формуле: distance = B×f / D,

где B – размер базы (расстояние между камерами),

f – фокусное расстояние в пикселях,

D – значение сдвига.

[0016] В другом частном примере осуществления способа вычислительное устройство дополнительно оснащено кодировщиком, трансформером и декодером, а этап определения для каждого пикселя первого и второго изображений значения сдвига, указывающее на количество пикселей, на которое сдвинут наиболее похожий пиксель другого изображения, содержит этапы, на которых:

- формируют первый и второй тензоры изображений, содержащие векторные представления (вектора признаков) по меньшей мере одного объекта, причем каждый элемент тензора представляет собой значение яркости соответствующего пикселя;

- нормируют полученные на предыдущем этапе тензоры;

- посредством кодировщика объединяют упомянутые два тензора в один тензор;

- посредством трансформера построчно сравнивают векторные представления по меньшей мере одного объекта, содержащиеся в полученном на предыдущем этапе тензоре, для формирования тензора, содержащего информацию о значениях сдвигов пикселей изображений друг относительно друга;

при этом этап формирования первой и второй карт сдвигов для первого и второго изображения выполняется декодером на основе полученного на предыдущем этапе тензора.

[0017] В другом частном примере осуществления способа кодировщик, трансформер и декодер реализованы на базе нейронных сетей, заранее обученных на тренировочном наборе данных.

[0018] В другом частном примере осуществления способа дополнительно выполняют этапы, на которых:

- на основе карты глубины формируют облако точек в трехмерном пространстве;

- используют облако точек для планирования траектории движения автономного беспилотного транспортного средства

[0019] В другом предпочтительном варианте осуществления заявленного решения представлено устройство построения карты глубины по паре изображений, содержащее по меньшей мере одно вычислительное устройство и по меньшей мере одно устройство памяти, содержащее машиночитаемые инструкции, которые

при их исполнении по меньшей мере одним вычислительным устройством выполняют вышеуказанный способ.

КРАТКОЕ ОПИСАНИЕ ЧЕРТЕЖЕЙ

[0020] Признаки и преимущества настоящего технического решения станут очевидными из приводимого ниже подробного описания технического решения и прилагаемых чертежей, на которых:

[0021] На Фиг. 1 – представлен пример реализации системы обработки изображений.

[0022] на Фиг. 2 – представлены примеры изображений, полученных с левой и правой камер.

[0023] на Фиг. 3 – представлены примеры изображений с заслоненными областями.

[0024] на Фиг. 4 – представлен пример изображений с фильтрацией на карте глубины.

[0025] на Фиг. 5 – представлен пример схемы архитектуры нейросети.

[0026] на Фиг. 6 – представлен пример общего вида вычислительного устройства.

ОСУЩЕСТВЛЕНИЕ ИЗОБРЕТЕНИЯ

[0027] Ниже будут описаны понятия и термины, необходимые для понимания данного технического решения.

[0028] В данном техническом решении под системой подразумевается, в том числе компьютерная система, ЭВМ (электронно-вычислительная машина), ЧПУ (числовое программное управление), ПЛК (программируемый логический контроллер), компьютеризированные системы управления и любые другие устройства, способные выполнять заданную, четко определенную последовательность операций (действий, инструкций).

[0029] Под устройством обработки команд подразумевается электронный блок, вычислительное устройство, либо интегральная схема (микропроцессор), исполняющая машинные инструкции (программы).

[0030] Устройство обработки команд считывает и выполняет машинные инструкции (программы) с одного или более устройств хранения данных. В роли устройства хранения данных могут выступать, но не ограничиваясь, жесткие диски

(HDD), флеш-память, ПЗУ (постоянное запоминающее устройство), твердотельные накопители (SSD), оптические приводы.

[0031] Вычислительное устройство - счётно-решающее устройство, автоматически выполняющее одну какую-либо математическую операцию или последовательность их с целью решения одной задачи или класса однотипных задач (Большая советская энциклопедия. — М.: Советская энциклопедия. 1969 — 1978.).

[0032] Программа - последовательность инструкций, предназначенных для исполнения устройством управления вычислительной машины или устройством обработки команд.

[0033] База данных (БД) - совокупность данных, организованных в соответствии с концептуальной структурой, описывающей характеристики этих данных и взаимоотношения между ними, причем такое собрание данных, которое поддерживает одну или более областей применения (ISO/IEC 2382:2015, 2121423 «database»).

[0034] Сигнал — материальное воплощение сообщения для использования при передаче, переработке и хранении информации.

[0035] Механизм внимания (англ. attention mechanism, attention model) — техника, используемая в рекуррентных нейронных сетях (сокр. RNN) и сверточных нейронных сетях (сокр. CNN) для поиска взаимосвязей между различными частями входных и выходных данных.

[0036] В соответствии со схемой, представленной на фиг. 1, система обработки изображений содержит: первую камеру 1, вторую камеру 2 и устройство 10 построения карты глубины по паре изображений.

[0037] В качестве камер могут быть использованы любые широко известные из уровня техники камеры, расположенные, например, на транспортном средстве, на заданном удалении друг от друга в любом направлении, например, на 10-100 см., и направленные в одну сторону на по меньшей мере один объект. Объектами, на которые направлены камеры, могут быть, например, пешеходы, дорожные заграждения, другие транспортные средства и участники дорожного движения, стены зданий, бордюры, небо, деревья, животные, дорога, тротуары и прочее. Для удобства далее будем считать, что камеры удалены друг от друга по горизонтали. В качестве камер могут быть использованы камеры, работающие как в видимом спектральном диапазоне (см., например, «Видимое излучение»), так и в УФ (см. «Ультрафиолетовое излучение») и ИК (см. «Инфракрасное излучение»)

диапазонах. В камерах могут быть использованы матрицы CMOS (complementary metal-oxide-semiconductor, комплементарная логика на транзисторах металл-оксид-полупроводник, КМОП) с активными чувствительными элементами (Active Pixel Sensor) и CCD (charge-coupled device, прибор с обратной зарядной связью).

[0038] Например, в качестве камеры может быть использована камера модели «LI-IMX390-GW5200-GMSL2-120H» (см, например, https://www.leopardimaging.com/product/autonomous-camera/maxim-gmsl2-cameras/li-imx390-gw5200-gmsl2/li-imx390-gw5200-gmsl2-120h/) или камера модель «acA2040-25gc - Basler ace» (см., например, https://www.baslerweb.com/en/products/cameras/area-scan-cameras/ace/aca2040-25gc/).

[0039] Камеры калибруются известными из уровня техники методами, а информация о положении и углах наклона камер относительно друг друга заносятся разработчиком в устройство 10 построения карты глубины по паре изображений. Также известными методами могут быть скорректированы линзовые искажения на изображениях, получаемых с камер (см. например, процедуру калибровки камер, раскрытую по ссылке в Интернет: https://docs.opencv.org/4.5.2/dc/dbb/tutorial_py_calibration.html).

[0040] Устройство 10 построения карты глубины по паре изображений может быть реализовано на базе по меньшей мере одного вычислительного устройства и содержать: модуль 11 сбора данных, модуль 12 ректификации изображений, модуль 20 формирования карты сдвигов и модуль 30 формирования карты глубины. Упомянутые модули могут быть реализованы на базе программно-аппаратных средств вычислительного устройства, в частности на базе его процессора или микроконтроллера, и оснащены соответствующими интерфейсами связи, логическими элементами, АЦП и ЦАП для обмена сигналами с целью передачи данных, в том числе информации об изображении, раскрытой ниже.

[0041] Соответственно, первое и второе изображения, содержащие по меньшей мере одно изображение объекта, например, левое изображение и правое изображение, с камер 1 и 2 поступают в модуль 11 сбора данных, который может быть оснащен, например, буфером – устройством, обеспечивающим синхронное получение данных с двух камер. Полученные изображения направляются упомянутым модулем 11 в модуль 12 ректификации изображений, который проецирует изображения в одну плоскость. Процедура ректификация пары изображений может осуществляться известными из уровня техники методами, например, раскрытыми в книге Zisserman R. H. A. Multiple view geometry in computer

vision, опубл. в 2004 г., размещенной в Интернет по адресу: https://www.r-5.org/files/books/computers/algo-list/image-processing/vision/Richard_Hartley_Andrew_Zisserman-Multiple_View_Geometry_in_Computer_Vision-EN.pdf, и позволяет получить два кадра с построчным соответствием расположения объектов на них (эпиполярные линии проектируются по горизонтали), в следствие чего повысится точность значений карты глубины.

[0042] Далее первое и второе изображения, прошедшие процедуру ректификации, поступают в модуль 20 формирования карт сдвигов, который выполняет сопоставление пикселей изображений. При сопоставлении пикселей изображения сравниваются значение яркости самого пикселя и значение яркости соседних с ним пикселей. Результатом сопоставления пикселей изображения является карта сдвигов, в которой для каждого сравниваемого пикселя изображения стоит значение сдвига, указывающее на количество пикселей, на которое сдвинут наиболее похожий пиксель изображения с другой камеры. Для каждой камеры получается своя карта сдвигов. Таким образом, получаются первая и вторая карты сдвигов, в частности, для первого и второго изображений, которые направляются в модуль 30 формирования карты глубины.

[0043] Для формирования карты глубины из карт сдвигов упомянутый модуль 30 на основе значений сдвигов пикселей определяет расстояние от линии, соединяющей центры камер, до каждого пикселя по меньшей мере одного объекта на изображении. Это расстояние обратно-пропорционально значению сдвига и вычисляется, например, по формуле:

distance = B×f / D;

где B – размер базы (расстояние между камерами);

f – фокусное расстояние в пикселях (в частности, используются одинаковые камеры и изображения со скорректированными линзовыми искажениями);

D – значение сдвига.

Фокусное расстояние f и размер базы B задаются разработчиком в памяти упомянутого модуля 30, которой он может быть дополнительно оснащен.

[0044] Соответственно, для первой и второй карт сдвигов модулем 30 формируются первая и вторая матрицы, значения которых характеризуют расстояние от точек по меньшей мере одного объекта на изображении до линии, соединяющей центры камер, после чего модуль 30 назначает первую или вторую матрицу в качестве карты глубины, в зависимости от заданного разработчиком программного алгоритма.

[0045] Дополнительно модуль 30 формирования карты глубины может быть выполнен с возможностью фильтрации карт сдвигов посредством проверки согласованности значений сдвигов пикселей левого и правого изображений, по итогу которой модуль 30 сформирует матрицу, содержащую информацию о координатах пикселей и значений, указывающих на то, что значения сдвигов пикселей, содержащихся в первой и второй картах сдвигов, являются согласованными или несогласованными. Алгоритм проверки согласованности значений сдвигов пикселей будет описан более подробно далее в тексте заявки. При формировании карты глубины описанным выше способом несогласованные значения сдвига пикселей при формировании карты глубины не учитываются.

[0046] Далее на основе данных полученной карты глубины упомянутый модуль 30 известными методами формирует облако точек в трехмерном пространстве, выполнив обратную проекцию с камеры в 3Д пространство (см, например, Bostanci, Gazi Erkan & Kanwal, Nadia & Clark, Adrian. (2015). Augmented reality applications for cultural heritage using Kinect. Human-centric Computing and Information Sciences. 5. 1-18. 10.1186/s13673-015-0040-3.). Облако точек (набор точек в трехмерном пространстве) может быть использовано для планирования траектории движения автономного беспилотного транспортного средства широко известными методами.

[0047] В альтернативном варианте реализации заявленного решения карта сдвига может быть сформирована посредством использования нейронной сети, которая состоит из нескольких других нейронных сетей. В данном варианте модуль 20 формирования карты сдвигов дополнительно оснащается модулем 21 нормирования тензоров, кодировщиком 22, трансформером 23 и декодером 24.

[0048] В качестве модуля 21 нормирования тензоров может быть использован по меньшей мере один процессор или микроконтроллер, сконфигурированные в программно-аппаратной части таким образом, чтобы выполнять приписанные модулю 21 ниже функции.

[0049] В качестве кодировщика 22 может быть использована по меньшей мере одна нейросеть, выполняющая выделение признаков на изображении, получая векторные представления объектов в уменьшенном пространственном разрешении. Например, нейросеть может быть реализована в виде стандартной сверточной нейросети ResNet18 (см., например, статью He, Kaiming et al, «Deep Residual Learning for Image Recognition», 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)).

[0050] В качестве трансформера 23 может быть использована по меньшей мере одна нейросеть, выполняющая построчное сопоставление признаков, выделенных предыдущей сетью. Упомянутая нейросеть может быть оснащена механизмом внимания, который оценивает, насколько похожи векторные признаки строки изображения, снятого первой камерой, на векторные признаки соответствующей строки изображения, полученного со второй камеры. Выходные тензоры с картой активации имеют такую же размерность, что и входные данные. Однако вектора тензора несут в себе информацию не о локальных признаках одного кадра, а информацию о сдвиге пикселей одной камеры относительного другой. При обучении трансформер 23 предсказывает тензор, из которого можно получить карту сдвигов в меньшем пространственном разрешении, чем у входного изображения. Трансформер 23 может быть реализован в виде нейросети семейства BERT (Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018)) с двумя слоями и восемью головами.

[0051] В качестве декодера 24 может быть использована по меньшей мере одна нейросеть, декодирующая тензоры признаков, вышедших с кодировщика 22 и трансформера 23, до исходного пространственного разрешения. С кодировщика берутся несколько тензоров признаков разного пространственного разрешения. Декодер 24 может представлять собой часть сети Unet, дополненную механизмом внимания. Выходом декодера 24 является нормированная от 0 до 1 карта сдвигов.

[0052] Изображения в цифровой среде представляются в виде тензоров (или матриц), содержащих значения, характеризующие цвета пикселей или их яркость. Соответственно, первое (например, левое) и второе (например, правое) изображения, прошедшие процедуру ректификации, в виде тензоров поступают в модуль 21 нормирования тензоров. Тензоры первого и второго изображений, T1 и T2 соответственно, могут быть представлены, например, в виде тензоров с размерностями CxHxW, где C - число каналов, H - высота изображения, W - ширина изображения. Например, для изображения с разрешением 512 на 960 пикселей - H=512, W=960 пикселей. Также изображения могут быть как цветными, так и черно-белыми. Цветные изображения содержат 3 канала (C=3, по одному каналу для каждого из трех цветов: красный, зеленый и синий), а черно-белые изображения содержат один канал (C=1). Значения размерностей C, H и W могут быть заданы разработчиком.

[0053] Далее алгоритм построения карты глубины будет раскрыт на примере цветных изображений. В данном варианте реализации технического решения каждый элемент тензора изображения содержит целое число со значениями от 0 до 255 включительно, которое показывает величину освещенности - яркости соответствующего пикселя матрицы камеры. Соответственно, сформированные первый и второй тензоры изображения поступают на вход модулю 21 нормирования тензоров.

[0054] Модуль 21 нормирует входящие тензоры следующим образом. Из значения яркости (величины освещенности) каждого пикселя вычитается величина к1 и затем делится на величину к2. Коэффициенты к1 и к2 задаются разработчиком. Затем упомянутые тензоры объединяются в один тензор NxCxHxW (N=2, C=3, H=512, W=960) для одновременной обработки первого и второго тензоров изображений с помощью кодировщика 22. Размер пакета N определяется упомянутым модулем 21 на основе количества изображений (тензоров), поступивших на вход модуля 21 одновременно или последовательно в заданный разработчиком интервал времени. Поскольку на вход модуля 21 поступило два изображения, то значение размера пакета N будет определено как 2. Кодировщик 22, а также все остальные модули, могут быть реализованы с помощью библиотеки PyTorch (см. https://pytorch.org/). Для реализации могут быть использованы и другие библиотеки, например, TensorFlow (https://www.tensorflow.org/), MxNet (https://mxnet.apache.org). Одновременная обработка данных первого и второго тензоров посредством объединения их в один тензор позволяет ускорить работу алгоритма по сравнению с последовательной обработкой двух исходных тензоров. Например, два тензора T1 и T2 размерности 3x512x960 объединяются в один тензор O путем добавления новой размерности (увеличения ранга тензора). Первая размерность тензора O фактически нумерует объединенные тензоры: O = (T1, T2).

По первой размерности тензора O в индексе 0 содержится тензор T1, в индексе 1 - T2:

O[0] = T1

O[1] = T2

[0055] Пример объединения для случая двух матриц 2x2.

Матрица T1 размером 2x2 1 2

3 4

Матрица T2 размером 2x2 5 6

7 8

Тензор O размером 2x2x2 1 2 5 6
3 4 , 7 8

[0056] Выходом кодировщика 22 при обработке одной пары тензоров изображений, объединенных в один тензор, является тензор, содержащий векторные представления (векторы признаков) по меньшей мере одного объекта в уменьшенном пространственном разрешении, например, в 16 раз, т.е. имеющий размерность 2x256x32x60 (N=2, C=256, H=32, W=60), где первое измерение N соответствует размеру пакета обработки, второе C – размерности вектора признаков каждого элемента, а оставшиеся два измерения H и W – уменьшенному пространственному размеру в 16 раз. Описание процедуры формирования векторного представления по меньшей мере одного объекта из входного тензора изображения раскрыто, например, в статье He, Kaiming et al, «Deep Residual Learning for Image Recognition», 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)). В результате обработки нейронной сетью входного тензора изображения получается новый тензор, который может быть представлен пользователю в виде карты признаков, содержащей информацию о по меньшей мере одном объекте, в том числе информацию о границах объекта. Эта информация закодирована в виде числовых значений. Также кодировщик 22 формирует дополнительные тензоры в заданном разработчиком по меньшей мере одном пространственном масштабе (разрешении), например, три тензора с размерностями 2x128x64x120, 2x64x128x240, 2x64x256x480, которые будут использованы далее декодером 24. Процедура получения тензоров (карт признаков) на разных пространственных масштабах с кодировщика 22 описана в статье He, Kaiming et al, «Deep Residual Learning for Image Recognition», 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[0057] Далее устройство 10 построения карты глубины переходит к этапу построчного сравнения векторных представлений (векторов признаков) по меньшей мере одного объекта для получения информации о сдвиге пикселей друг относительно друга. Векторные представления (вектора признаков) содержат информацию о по меньшей мере одном объекте на изображениях и извлекаются из тензоров, полученных на предыдущем этапе, и, при необходимости, могут быть отображены пользователю с помощью известной операции обращения по индексам. Для построчного сравнения векторных представлений (векторов признаков) объекта на левом и правом изображениях необходимо посредством кодировщика 22 построчно объединить признаки с двух изображений. Для этого

кодировщик 22 объединяет в строку элементы карт признаков, которые соответствуют одной и той же строке входных изображений. Для каждого пакета (тензора) вдоль размерности N кодировщик 22 объединяет строки по размерности W, соответствующей ширине изображения. По индексу с номером 0 расположены данные первого пакета, по индексу 1 - данные второго пакета. Например, пусть полученная карта признаков П имеет размерность NxCxW (2x2x3, N=2, C=2, W=3). Для простоты размерность H пропущена:

Тензор П размером 2x2x3: 1 2 3 7 8 9

4 5 6, 10 11 12

Делая объединение строк по размерности W, получим тензор размерности Cx2W (2x6):

1 2 3 7 8 9

4 5 6 10 11 12

[0058] Таким образом, из тензора NxCxHxW (2x256x32x60, N=2, C=256, H=32, W=60) кодировщик 22 формирует тензор CxHx2W (256x32x120), после чего для пакетной обработки всех строк кодировщиком 22 переставляются размерности тензора (числа содержащиеся в тензоре при этом остаются без изменений), например, тензор приводится к виду Hx2WxC (32x120x256). Алгоритм перестановки размерностей может быть задан разработчиком известными из уровня техники методами. Упомянутый тензор далее передается кодировщиком 22 на вход трансформеру 23, который возвращает тензор такого же размера 32x120x256. Трансформер 23 извлекает из входного тензора векторные представления объекта, определенные для первого изображения, векторные представления объекта, определенные для второго изображения, после чего сравнивает упомянутые векторные представления для определения значений сдвигов пикселей изображений друг относительно друга (см. Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems 30 (2017)), которые также могут называться как значения предсказаний сдвигов пикселей. После обработки трансформер 23 формирует тензор, в который включается полученная информацию о значениях сдвигов пикселей изображений друг относительно друга. Эта информация закодирована в виде числовых значений. Далее полученный тензор преобразуется обратно в тензор с размерностью NxCxHxW (2x256x32x60)

путем разбиения исходного тензора по размерности строк и дальнейшими перестановками размерностей. Алгоритм перестановки размерностей задан разработчиком.

[0059] Соответственно, декодер 24 получает на вход от модуля 21 тензор с нормированными изображениями (2x3x512x960), тензор с выхода трансформера 23 (2x256x32x60), содержащий информацию о значениях сдвигов пикселей изображений друг относительно друга, а также дополнительное подкрепление в виде карт признаков (тензоров) с кодировщика 22 на заданных разработчиком пространственных масштабах (2x128x64x120, 2x64x128x240, 2x64x256x480). Описание работы декодера 24 приведено в статье (Roy, Abhijit Guha, Nassir Navab, and Christian Wachinger. "Recalibrating fully convolutional networks with spatial and channel “squeeze and excitation” blocks." IEEE transactions on medical imaging 38.2 (2018): 540-549). Декодер 24 на основе полученных данных формирует два тензора, содержащих значения сдвигов пикселей, которые могут быть представлены пользователю в виде карт сдвигов. Тензоры соответствуют первому (например, левому) и второму (например, правому) изображениям и имеют заданную разработчиком размерность, например, 512x960. Далее полученные два тензора будем называть картами сдвигов. Затем карты сдвигов направляются в модуль 30 формирования карты глубины.

[0060] Получение сразу двух карт сдвигов необходимо для алгоритма постобработки, выполняемого модулем 30, который проверяет согласованность значений сдвига пикселей для левого и правого изображений. Алгоритм согласования значений сдвига пикселей следующий. Рассматриваем последовательно все пиксели на первом (левом) изображении. Для каждого пикселя по карте сдвигов, сформированной декодером 24 для первого (левого) изображения, модуль 30 находит значение сдвига пикселя. По значению сдвига модуль 30 определяет, какому пикселю на втором (правом) изображении соответствует пиксель на первом (левом) изображении. Аналогично для найденного пикселя на правом изображении с помощью карты сдвигов, сформированной декодером 24, модуль 30 находит соответствующий пиксель на левом изображении.

[0061] В частности, для каждого пикселя первой (левой) карты сдвигов, имеющего координаты x_left, y_left (по горизонтали и вертикали матрицы соответственно) и значение d_left (величина сдвига), модуль 30 находит соответствующий пиксель на второй (правой) карте сдвигов с координатами x_right, y_right, где x_right = x_left +

d_left, y_right = y_left. Далее модуль 30 на основе определенных координат найденного пикселя извлекает из второй карты сдвига значение сдвига, определенное для данного пикселя: d_right. Значение сдвига d_right, определенное для пикселя второго изображения, сравнивается модулем 30 с значением сдвига пикселя d_left, определенное для пикселя первого изображения, причем если разница по модулю между d_left и d_right меньше заданного порогового значения (например, 2), то модуль 30 определяет, что значения сдвигов согласованы, после чего модуль 30 формирует матрицу согласованности значений сдвигов пикселей, размер которой соответствует размеру исходной карты сдвигов, и назначает для пикселя значение (например, указывает значение True в ячейке с координатами x_left, y_left), указывающее на то, что значения сдвигов, определенные для данного пикселя, согласуются. Если разница по модулю между d_left и d_right больше заданного порогового значения, то модуль 30 назначает для пикселя значение (например, False), указывающее на то, что упомянутые значения сдвигов не согласуются. Пороговое значение может быть задано разработчиком модуля 30. Соответственно, если пиксель вернулся в тоже самое место с ошибкой менее 2 пикселей, то считается, что пиксели согласованы. Если ошибка составила более 2 пикселей, то значения сдвигов не согласованы.

[0062] Это необходимо для увеличения точности полученных значений сдвигов за счет фильтрации зон, в которых значения сдвигов не согласуются. Как правило это зоны, которые видны на одной камере и не видны на другой. Примеры показаны на Фиг. 2 и 3. На Фиг. 2 показаны исходные изображения для левой и правой камер. На Фиг. 3 показаны области, которые видны на одной камере и не видны на другой из-за эффекта параллакса (https://en.wikipedia.org/wiki/Parallax). Для таких зон значение сдвига не определено. Наш подход может делать предсказания сдвигов в таких областях, но точность предсказаний здесь получается значительно ниже, чем в случае видимости объекта на двух камерах одновременно. На Фиг. 4 показана исходная карта глубины и карта глубины после фильтрации. Как видно, повысилась четкость границ объектов и были обнаружены области с высокой ошибкой предсказаний сдвигов, что положительно скажется на алгоритмах анализа препятствий, которые будут использовать полученные карты глубины.

[0063] Соответственно, по итогу проверки согласованности значений сдвигов пикселей модуль 30 формирует матрицу, содержащую информацию о координатах пикселей и значений, указывающих на то, что значения сдвигов пикселей, содержащиеся в первой и второй картах сдвигов, являются согласованными или

несогласованными. Далее модуль 30 переходит к этапу формирования известными методами, например, раскрытыми ранее, карты глубины на основе значений карт сдвигов, причем несогласованные значения сдвига пикселей при формировании карты глубины не учитываются.

[0064] Таким образом, за счет того, что изображения, полученные с камер, проходят процедуру ректификации, а карта глубины строится на основе значений карт сдвигов пикселей, определенных для каждого полученного изображения, повышается точность значений карты глубины.

[0065] Для обучения нейронной сети используется общепринятый подход [см Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. Deep learning. MIT press, 2016.]. Берется датасет с изображениями и известными картами глубины. В качестве датасетов, например, можно использовать SceneFlow (https://lmb.informatik.uni-freiburg.de/resources/datasets/SceneFlowDatasets.en.html) и KITTI (http://www.cvlibs.net/datasets/kitti/eval_stereo_flow.php?benchmark=stereo, http://www.cvlibs.net/datasets/kitti/eval_scene_flow.php?benchmark=stereo).

[0066] Нейронная сеть обучается на тренировочном наборе изображений так, чтобы предсказывать размеченную карту глубины.

Ввиду проблем с обучением трансформера 23, обучение всей нейросети ведется в три этапа:

1. Обучаются части нейронной сети, в частности модули: кодировщик 22 и трансформер 23 следующим образом. Изображения с двух камер в виде тензоров поступают на модуль нормировки 21. Полученные нормированные изображения далее обрабатываются кодировщиком 22 и затем трансформером 23. Тензор размерности 2x256x32x60 с выхода трансформера 23 приводится к карте сдвигов с помощью усреднения по измерению канала C=256. Из тензора размерности 2x256x32x60 получается тензор размерности 2x32x60, который соответствует двум картам сдвигов для левого и правого изображений (первая размерность тензора нумерует карты сдвигов). Полученные карты передаются в функцию потерь вместе с известной картой сдвигов, уменьшенной в 16 раз. Этот этап нужен, чтобы предобучить трансформер 23. Без этого этапа трансформер 23 будет обучаться медленно, т. к. преимущественно будут обучаться кодировщик 22, декодер 24, и модель будет оставаться в локальном минимуме.

2. Обучаются все части нейронной сети, модули кодировщик 22, трасформер 23 и декодер 24. Но функция потерь для карты сдвига, полученной с трансформера 23 не вычисляется. В этом режиме обучается преимущественно декодер,

предсказывающий карты сдвига на полном разрешении (таком же как у входного изображения).

3. Обучаются все части нейронной сети, модули кодировщик 22, трасформер 23 и декодер 24. Оптимизируемая функция потерь выглядит как взвешенная сумма значений функций потерь для двух выходов нейросети: с трансформера 23 и с декодера 24. Трансформер 23 предсказывает карты сдвигов в уменьшенном разрешении, декодер 24 — в полном разрешении. Веса равны соответственно 0.1 и 0.001. Это позволяет провести более точную настройку весов всей нейронной сети. Функция потерь для предсказаний трансформера 23 играет роль регуляризации, заключающейся в том, что признаки с трансформера 23 должны нести в себе информацию, достаточную для создания карты сдвигов.

[0067] При обучении известная карта сдвигов нормализуется с помощью выбранного максимального значения сдвига равного 200 пикселей. Если значение больше этого порога, то оно обрезается до 200. Таким образом, все значения входят в отрезок [0;1], что позволяет использовать на выходе нейросети сигмоиду в качестве функции активации и вести обучение с помощью 2D бинарной кросс-энтропии (BCE - Binary Cross Entropy, см ссылки https://towardsdatascience.com/understanding-binary-cross-entropy-log-loss-a-visual-explanation-a3ac6025181a , https://pytorch.org/docs/stable/generated/torch.nn.BCELoss.html).

[0068] Пример вычисления функции BCE. Пусть $p_{i,gt}$ и $p_{i,pred}$ известная и предсказанная нормированная карта сдвигов для пикселя номером i, соответственно. Тогда значении функции потерь L_{bce} равно:

$$L_{bce} = -\frac{1}{N}\sum_{i=1}^{N}[p_{i,gt} * log(p_{i,pred}) + (1 - p_{i,gt}) * log(1 - p_{i,pred})]$$

где N - число пикселей, суммирование идет по всем пикселям, логарифм натуральный.

[0069] В качестве оптимизатора был выбран Adabelief [см Adabelief]. Реализация алгоритма оптимизатора более подробно раскрыта по ссылке: https://github.com/jettify/pytorch-optimizer.

[0070] В общем виде (см. Фиг. 6) вычислительное устройство (200) содержит объединенные общей шиной информационного обмена один или несколько процессоров (201), средства памяти, такие как ОЗУ (202) и ПЗУ (203), интерфейсы

ввода/вывода (204), устройства ввода/вывода (205), и устройство для сетевого взаимодействия (206).

[0071] Процессор (201) (или несколько процессоров, многоядерный процессор и т.п.) может выбираться из ассортимента устройств, широко применяемых в настоящее время, например, таких производителей, как: Intel™, AMD™, Apple™, Samsung Exynos™, MediaTEK™, Qualcomm Snapdragon™ и т.п. Под процессором или одним из используемых процессоров в системе (200) также необходимо учитывать графический процессор, например, GPU NVIDIA с программной моделью, совместимой с CUDA, или Graphcore, тип которых также является пригодным для полного или частичного выполнения способа, а также может применяться для обучения и применения моделей машинного обучения в различных информационных системах.

[0072] ОЗУ (202) представляет собой оперативную память и предназначено для хранения исполняемых процессором (201) машиночитаемых инструкций для выполнения необходимых операций по логической обработке данных. ОЗУ (202), как правило, содержит исполняемые инструкции операционной системы и соответствующих программных компонент (приложения, программные модули и т.п.). При этом, в качестве ОЗУ (202) может выступать доступный объем памяти графической карты или графического процессора.

[0073] ПЗУ (203) представляет собой одно или более устройств постоянного хранения данных, например, жесткий диск (HDD), твердотельный накопитель данных (SSD), флэш-память (EEPROM, NAND и т.п.), оптические носители информации (CD-R/RW, DVD-R/RW, BlueRay Disc, MD) и др.

[0074] Для организации работы компонентов системы (200) и организации работы внешних подключаемых устройств применяются различные виды интерфейсов В/В (204). Выбор соответствующих интерфейсов зависит от конкретного исполнения вычислительного устройства, которые могут представлять собой, не ограничиваясь: PCI, AGP, PS/2, IrDa, FireWire, LPT, COM, SATA, IDE, Lightning, USB (2.0, 3.0, 3.1, micro, mini, type C), TRS/Audio jack (2.5, 3.5, 6.35), HDMI, DVI, VGA, Display Port, RJ45, RS232 и т.п.

[0075] Для обеспечения взаимодействия пользователя с вычислительным устройством (200) применяются различные средства (205) В/В информации, например, клавиатура, дисплей (монитор), сенсорный дисплей, тач-пад, джойстик, манипулятор мышь, световое перо, стилус, сенсорная панель, трекбол, динамики, микрофон, средства дополненной реальности, оптические сенсоры, планшет,

световые индикаторы, проектор, камера, средства биометрической идентификации (сканер сетчатки глаза, сканер отпечатков пальцев, модуль распознавания голоса) и т.п.

[0076] Средство сетевого взаимодействия (206) обеспечивает передачу данных посредством внутренней или внешней вычислительной сети, например, Интранет, Интернет, ЛВС и т.п. В качестве одного или более средств (206) может использоваться, но не ограничиваться: Ethernet карта, GSM модем, GPRS модем, LTE модем, 5G модем, модуль спутниковой связи, NFC модуль, Bluetooth и/или BLE модуль, Wi-Fi модуль и др.

[0077] Дополнительно могут применяться также средства спутниковой навигации в составе устройства (200), например, GPS, ГЛОНАСС, BeiDou, Galileo.

[0078] Конкретный выбор элементов устройства (200) для реализации различных программно-аппаратных архитектурных решений может варьироваться с сохранением обеспечиваемого требуемого функционала.

[0079] Модификации и улучшения вышеописанных вариантов осуществления настоящего технического решения будут ясны специалистам в данной области техники. Предшествующее описание представлено только в качестве примера и не несет никаких ограничений. Таким образом, объем настоящего технического решения ограничен только объемом прилагаемой формулы изобретения.

ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Способ построения карты глубины по паре изображений, выполняемый по меньшей мере одним вычислительным устройством, содержащий этапы, на которых:

- получают с первой и второй камер первое и второе изображения, содержащие изображение по меньшей мере одного объекта;

- выполняют процедуру ректификации первого и второго изображений посредством проецирования их в одну плоскость;

- определяют для каждого пикселя первого и второго изображений значение сдвига, указывающее на количество пикселей, на которое сдвинут наиболее похожий пиксель другого изображения;

- формируют первую и вторую карты сдвигов для первого и второго изображений, содержащие упомянутые значения сдвига;

- на основе значений карт сдвигов, сформированных на предыдущем этапе, формируют карту глубины изображения.

2. Способ по п. 1, характеризующийся тем, что этап определения значения сдвига для каждого пикселя первого и второго изображений содержит этапы, на которых:

- определяют значение яркости (величину освещенности) каждого пикселя первого и второго изображений;

- сопоставляют значения яркости пикселей первого изображения со значениями яркости пикселей второго изображения для определения значения сдвига для каждого пикселя первого и второго изображений, причем при сопоставлении учитывают также значения яркости соседних пикселей.

3. Способ по п. 1, характеризующийся тем, что дополнительно выполняют этапы проверки согласованности значений сдвигов пикселей левого и правого изображений.

4. Способ по п. 1, характеризующийся тем, что этап формирования карты глубины изображения на основе значений карт сдвигов содержит этапы, на которых:

- на основе значений сдвигов пикселей, содержащихся в картах сдвигов, определяют расстояние от линии, соединяющей центры камер, до каждого пикселя по меньшей мере одного объекта;

- на основе полученных значений расстояний от линии, соединяющей центры камер, до каждого пикселя по меньшей мере одного объекта, формируют карту глубины изображения.

5. Способ по п. 4, характеризующийся тем, что упомянутое расстояние определяется по формуле: distance = B×f / D,

где B – размер базы (расстояние между камерами),

f – фокусное расстояние в пикселях,

D – значение сдвига.

6. Способ по п. 1, характеризующийся тем, что вычислительное устройство дополнительно оснащено кодировщиком, трансформером и декодером, а этап определения для каждого пикселя первого и второго изображений значения сдвига, указывающее на количество пикселей, на которое сдвинут наиболее похожий пиксель другого изображения, содержит этапы, на которых:

- формируют первый и второй тензоры изображений, содержащие векторные представления (вектора признаков) по меньшей мере одного объекта, причем каждый элемент тензора представляет собой значение яркости соответствующего пикселя;

- нормируют полученные на предыдущем этапе тензоры;

- посредством кодировщика объединяют упомянутые два тензора в один тензор;

- посредством трансформера построчно сравнивают векторные представления по меньшей мере одного объекта, содержащиеся в полученном на предыдущем этапе тензоре, для формирования тензора, содержащего информацию о значениях сдвигов пикселей изображений друг относительно друга;

при этом этап формирования первой и второй карт сдвигов для первого и второго изображения выполняется декодером на основе полученного на предыдущем этапе тензора.

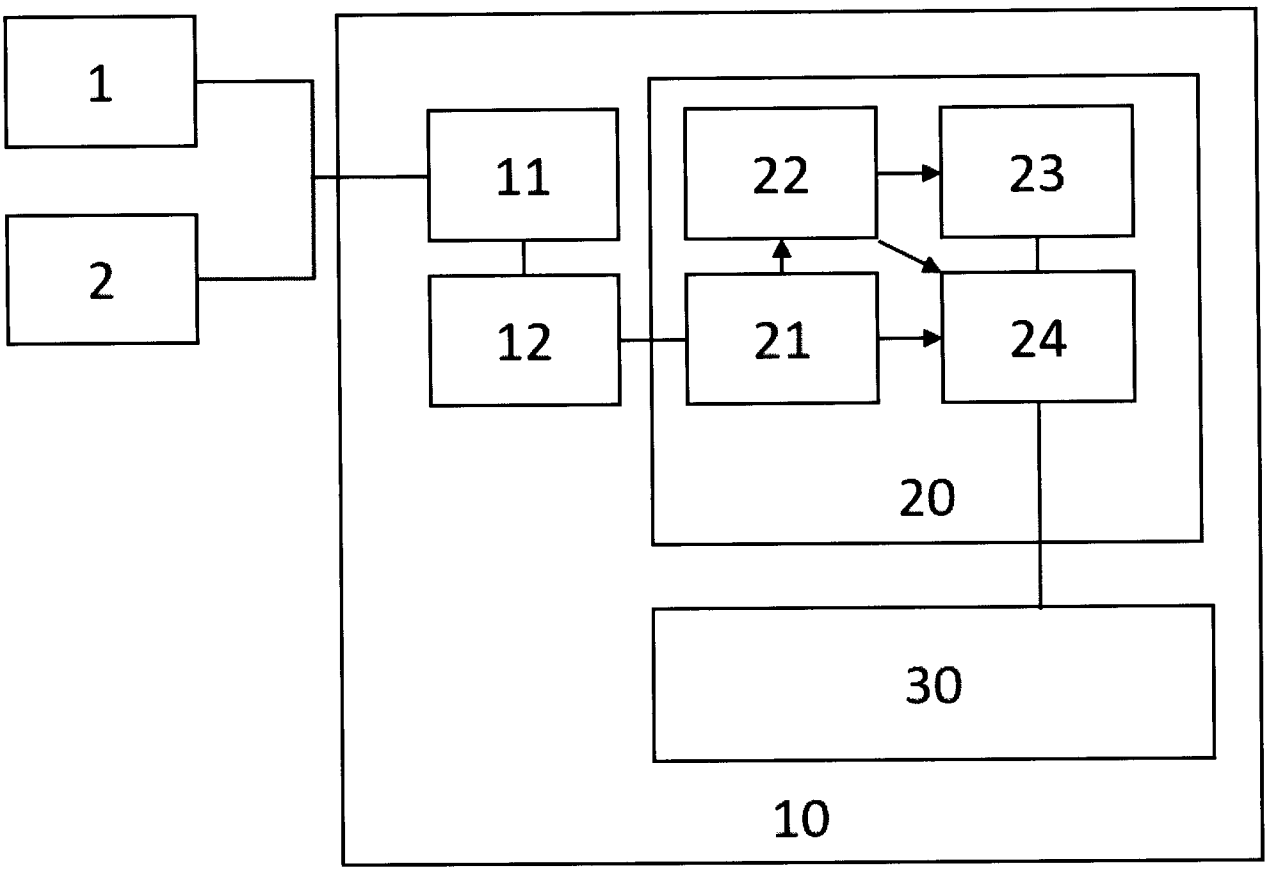
7. Способ по п. 6, характеризующийся тем, что кодировщик, трансформер и декодер реализованы на базе нейронных сетей, заранее обученных на тренировочном наборе данных.

8. Способ по п. 1, характеризующийся тем, что дополнительно выполняют этапы, на которых:

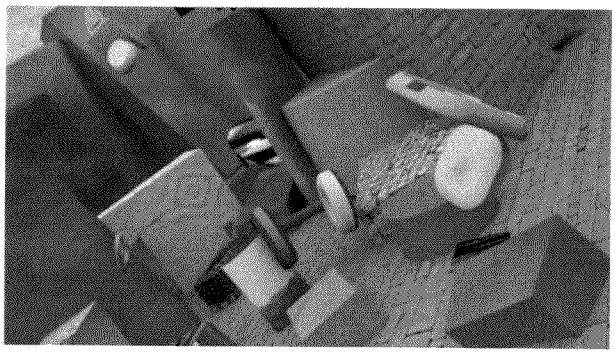
- на основе карты глубины формируют облако точек в трехмерном пространстве;

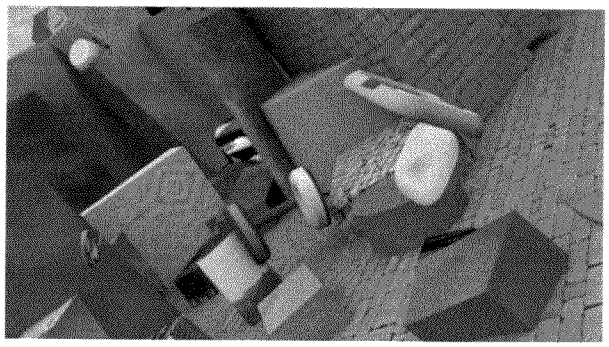
- используют облако точек для планирования траектории движения автономного беспилотного транспортного средства.

9. Устройство построения карты глубины по паре изображений, содержащее по меньшей мере одно вычислительное устройство и по меньшей мере одно устройство памяти, содержащее машиночитаемые инструкции, которые при их исполнении по меньшей мере одним вычислительным устройством выполняют способ по любому из пп. 1-8.

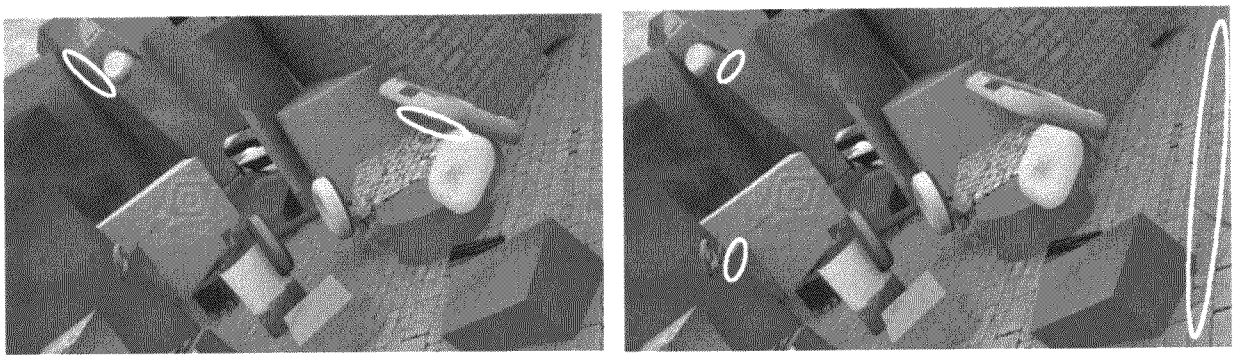


ФИГ. 1

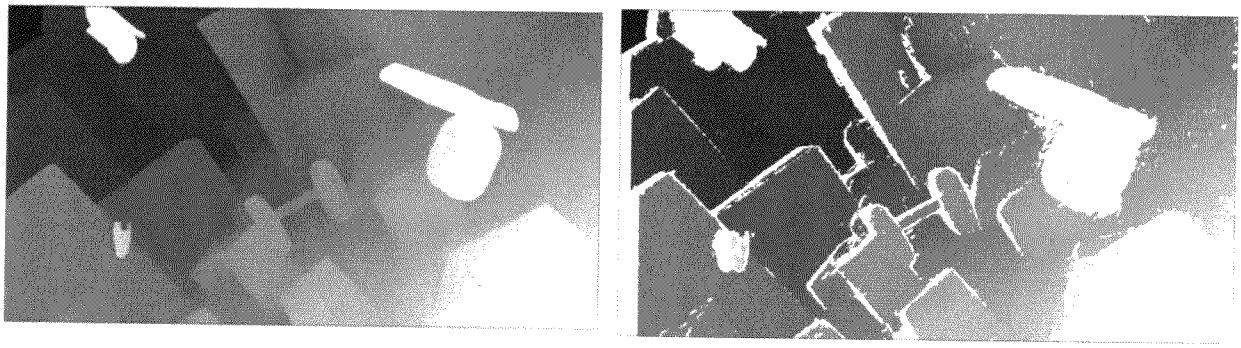




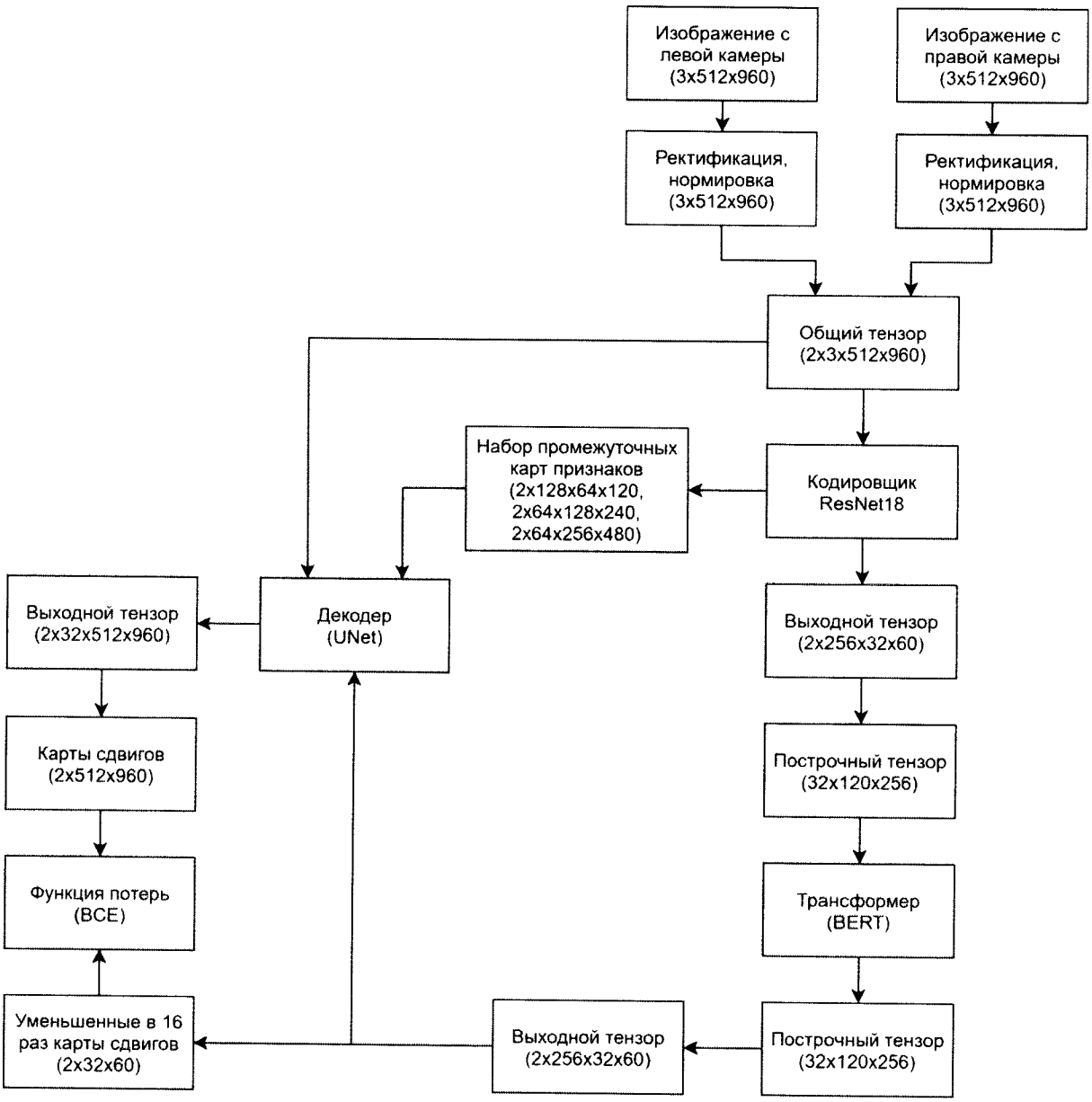
ФИГ. 2



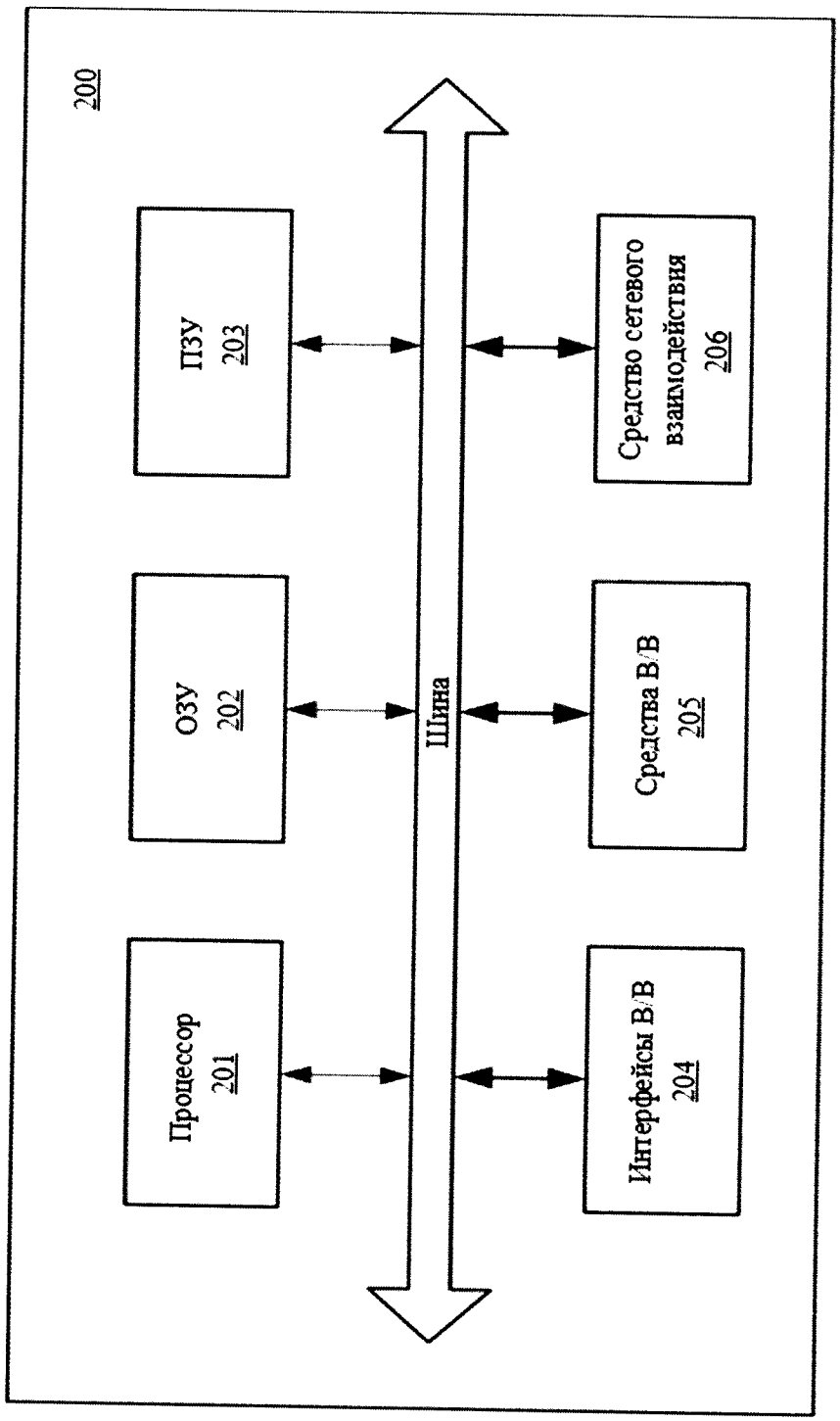
ФИГ. 3



ФИГ. 4


Изображение с левой камеры (3x512x960)
Изображение с правой камеры (3x512x960)
Ректификация, нормировка (3x512x960)
Ректификация, нормировка (3x512x960)
Общий тензор (2x3x512x960)
Набор промежуточных карт признаков (2x128x64x120, 2x64x128x240, 2x64x256x480)
Кодировщик ResNet18
Выходной тензор (2x32x512x960)
Декодер (UNet)
Выходной тензор (2x256x32x60)
Карты сдвигов (2x512x960)
Построчный тензор (32x120x256)
Функция потерь (BCE)
Трансформер (BERT)
Уменьшенные в 16 раз карты сдвигов (2x32x60)
Выходной тензор (2x256x32x60)
Построчный тензор (32x120x256)

ФИГ. 5


200
ПЗУ
203
ОЗУ
202
Процессор
201
Шина
Средство сетевого взаимодействия
206
Средства В/В
205
Интерфейсы В/В
204

ФИГ. 6

INTERNATIONAL SEARCH REPORT

International application No.
PCT/RU 2022/000123

A. CLASSIFICATION OF SUBJECT MATTER

G06T 7/593 (2017.01)
G06T 5/20 (2006.01)

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06T 3/00-3/40, 5/00-5/50, 7/00-7/90

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

PatSearch (RUPTO internal), USPTO, PAJ, K-PION, Esp@cenet, **Информационно-поисковая система ФИПС**

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	NOVIKOV A.I. Algoritmy rekonstruktsi trekhmernykh izobrazheni po posledovatelnosti stereoizobrazheni [onlain] Riazan: Book let, 2018 [retrieved on 2022-12-16]. Naideno in: https://www.elibrary.	1-5, 8-9
Y	ru/item.asp?id=32876453>, p. 91-103	6-7
Y	XINYI LI et al. TransCamP: Graph Transformer for 6-DoF Camera Pose Estimation. ArXiv: 2105.14065vl, 28.05.2021, abstract, §1 Introduction, §4 TransCamP Architecture [onlain] [retrieved on 2022-12-16]. Found in: <doi: 10.48550/arXiv.2105.14065 >	6-7
A	DAEHO KIM et al. Remote proximity monitoring between mobile construction resources using camera-mounted UAVs. Elsevier, tom 99, 21.12.2018 [onlain] [retrieved on 2022-12-16]. Found in: <doi: 10.1016/j.autcon.2018.12.014 >	1-9

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
14 December 2022 (14.12.2022)	12 January 2023 (12.01.2023)
Name and mailing address of the ISA/ RU	Authorized officer
Facsimile No.	Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/RU 2022/000123

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	MAD HU ANAND L. et al. Deep learning for monocular depth estimation from UAV images. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, tom V-2-2020 [onlain] [retrieved on 2022-12-16]. Found in: <doi: 10.5194/isprs-annals-V-2-2020-451-2020 >	1-9
A	ANGELA DAI et al. BundleFusion: Real-Time Globally Consistent 3D Reconstruction Using On-the-Fly Surface Reintegration. ACM Transactions on Graphics, tom 36, № 3, June 2017 [onlain] [retrieved on 2022-12-16]. Found in: <doi:10.1145/3054739>	1-9

ОТЧЕТ О МЕЖДУНАРОДНОМ ПОИСКЕ

Номер международной заявки

PCT/RU 2022/000123

A. КЛАССИФИКАЦИЯ ПРЕДМЕТА ИЗОБРЕТЕНИЯ

G06T 7/593 (2017.01)
G06T 5/20 (2006.01)

Согласно Международной патентной классификации МПК

B. ОБЛАСТЬ ПОИСКА

Проверенный минимум документации (система классификации с индексами классификации)

G06T 3/00-3/40, 5/00-5/50, 7/00-7/90

Другая проверенная документация в той мере, в какой она включена в поисковые подборки

Электронная база данных, использовавшаяся при поиске (название базы и, если, возможно, используемые поисковые термины)

PatSearch (RUPTO internal), USPTO, PAJ, K-PION, Esp@cenet, Информационно-поисковая система ФИПС

C. ДОКУМЕНТЫ, СЧИТАЮЩИЕСЯ РЕЛЕВАНТНЫМИ:

Категория*	Цитируемые документы с указанием, где это возможно, релевантных частей	Относится к пункту №
X	НОВИКОВ А.И. Алгоритмы реконструкции трехмерных изображений по последовательности стереоизображений [онлайн] Рязань: Book Jet, 2018 [найдено 2022-12-16]. Найдено в:<https://www.elibrary.ru/item.asp?id=32876453>, с. 91-103	1-5, 8-9
Y		6-7
Y	XINYI LI и др. TransCamP: Graph Transformer for 6-DoF Camera Pose Estimation. ArXiv: 2105.14065v1, 28.05.2021, реферат, §1 Introduction, §4 TransCamP Architecture [онлайн] [найдено 2022-12-16]. Найдено в: <doi:10.48550/arXiv.2105.14065 >	6-7
A	DAEHO KIM и др. Remote proximity monitoring between mobile construction resources using camera-mounted UAVs. Elsevier, том 99, 21.12.2018 [онлайн] [найдено 2022-12-16]. Найдено в: <doi: 10.1016/j.autcon.2018.12.014 >	1-9

[X] последующие документы указаны в продолжении графы C. [ ] данные о патентах-аналогах указаны в приложении

* Особые категории ссылочных документов:

"A" документ, определяющий общий уровень техники и не считающийся особо релевантным

"D" документ, цитируемый заявителем в международной заявке

"E" более ранняя заявка или патент, но опубликованная на дату международной подачи или после нее

"L" документ, подвергающий сомнению притязание(я) на приоритет, или который приводится с целью установления даты публикации другого ссылочного документа, а также в других целях (как указано)

"O" документ, относящийся к устному раскрытию, использованию, экспонированию и т.д.

"P" документ, опубликованный до даты международной подачи, но после даты испрашиваемого приоритета

"T" более поздний документ, опубликованный после даты международной подачи или приоритета, но приведенный для понимания принципа или теории, на которых основывается изобретение

"X" документ, имеющий наиболее близкое отношение к предмету поиска; заявленное изобретение не обладает новизной или изобретательским уровнем, в сравнении с документом, взятым в отдельности

"Y" документ, имеющий наиболее близкое отношение к предмету поиска; заявленное изобретение не обладает изобретательским уровнем, когда документ взят в сочетании с одним или несколькими документами той же категории, такая комбинация документов очевидна для специалиста

"&" документ, являющийся патентом-аналогом

Дата действительного завершения международного поиска	Дата отправки настоящего отчета о международном поиске
14 декабря 2022 (14.12.2022)	12 января 2023 (12.01.2023)
Наименование и адрес ISA/RU: Федеральный институт промышленной собственности, Бережковская наб., д. 30, корп. 1, Москва, Г-59, ГСП-3, 125993, Российская Федерация тел. +7(499)240-60-15, факс +7(495)531-63-18	Уполномоченное лицо: Кривцова Е. Телефон № 499-240-60-15

ОТЧЕТ О МЕЖДУНАРОДНОМ ПОИСКЕ

Номер международной заявки

PCT/RU 2022/000123

C. (Продолжение). ДОКУМЕНТЫ СЧИТАЮЩИЕСЯ РЕВАЛЕНТНЫМИ

Категория*	Цитируемые документы с указанием, где это возможно, релевантных частей	Относится к пункту №
A	MADHUANAND L. и др. Deep learning for monocular depth estimation from UAV images. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, том V-2-2020 [онлайн] [найдено 2022-12-16]. Найдено в: <doi:10.5194/isprs-annals-V-2-2020-451-2020 >	1-9
A	ANGELA DAI и др. BundleFusion: Real-Time Globally Consistent 3D Reconstruction Using On-the-Fly Surface Reintegration. ACM Transactions on Graphics, том 36, № 3, июнь 2017 [онлайн] [найдено 2022-12-16]. Найдено в: <doi:10.1145/3054739>	1-9