

Грабарь Д. М., Иванов Ю. С.
D. M. Grabar, Y. S. Ivanov

ИССЛЕДОВАНИЕ МЕТОДОВ И АЛГОРИТМОВ РАСПОЗНАВАНИЯ ОБРАЗОВ НА ИЗОБРАЖЕНИЯХ

INVESTIGATION OF METHODS AND ALGORITHMS FOR PATTERN RECOGNITION IN IMAGES

Грабарь Даниил Михайлович – магистр кафедры «Управление инновационными процессами и проектами» Комсомольского-на-Амуре государственного университета (Россия, Комсомольск-на-Амуре). E-mail: grabardm@ml-dev.ru.

Daniil M. Grabar – Master's Degree Student, Management of Innovative Processes and Projects Department, Komsomolsk-na-Amure State University (Russia, Komsomolsk-on-Amur). E-mail: grabardm@ml-dev.ru.

Иванов Юрий Сергеевич – кандидат технических наук, доцент кафедры «Промышленная электроника» Комсомольского-на-Амуре государственного университета (Россия, Комсомольск-на-Амуре); 681013, Хабаровский край, г. Комсомольск-на-Амуре, ул. Ленина, д. 27. E-mail: ivanov_ys@icloud.com.

Yurii S. Ivanov – PhD in Engineering, Associate Professor, Industrial Electronics Department, Komsomolsk-na-Amure State University (Russia, Komsomolsk-on-Amur); 681013, Khabarovsk territory, Komsomolsk-on-Amur, 27 Lenin str. E-mail: ivanov_ys@icloud.com.

Аннотация. В статье проведены исследование методов и разработка алгоритмов распознавания образов на изображениях в коллаборативных роботизированных системах. Проведён обзор существующих подходов и алгоритмов, включая методы машинного обучения, обработки изображений и компьютерного зрения. Подготовлен сравнительный анализ проведённых экспериментов с использованием различных наборов данных Crowd Pose, COCO, Pose Track. Точность работы алгоритмов вычислялась с помощью метрики AP. Описаны алгоритмы калибровки мультикамерных систем по шаблону. В качестве объединения локальных декартовых систем координат в глобальную декартову систему координат использовалась триангуляция данных с использованием прямого линейного преобразования.

Summary. The article investigates methods and develops algorithms for pattern recognition in images in collaborative robotic systems. The review of existing approaches and algorithms, including methods of machine learning, image processing and computer vision, is carried out. A comparative analysis of the conducted experiments using various datasets Crowd Pose, COCO, Pose Track has been prepared. The accuracy of the algorithms was calculated using the AP metric. Algorithms of calibration of multi-cameral systems according to a template are described. As a combination of local Cartesian coordinate systems into a global Cartesian coordinate system, data triangulation using a direct linear transformation was used.

Ключевые слова: калибровка, ключевые точки, мультикамерная система, распознавание образов.

Key words: calibration, key points, multi-camera system, pattern recognition.

Работа выполнена при поддержке Российского научного фонда № 22-71-10093 «Разработка и синтез перспективных мультимодальных адаптивных алгоритмов и методов управления поведением коллаборативных робототехнических систем с учётом нестандартных ситуаций и экстремальных условий в недетерминированной среде».

УДК 004.855.5

Введение. В концепции Индустрия 4.0 [1] роботы стали обычным явлением, как и развитие многих технологий. Наиболее распространённой областью, помимо робототехники, стала область ИИ (искусственный интеллект). Основным недостатком Индустрии 4.0 выступил тот факт, что всё производство построено без вмешательства человека в технологический процесс. Выполнение

специфических задач, НИР и ОКР на производстве, где требуется непосредственное взаимодействие робот-человек, стало невозможным. Для решения текущей проблемы была предложена концепция Индустрия 5.0 [2]. Стратегические методы интеграции людей с роботами необходимы для получения максимальной эффективности технологических процессов, так были созданы совершенно новые роботы, которые получили название «коботы» [3]. Важной особенностью кобота является совместная работа человека и робота. Хотя роботы гораздо более точны и эффективны, чем люди, излишняя автоматизация бесполезна. Роботы неспособны мыслить и негибки, что делает их чрезмерное использование вредным. Коботы позволяют безопасно взаимодействовать с внешней средой за счёт встроенных датчиков. При этом внутренняя сенсорика не позволяет оценить всю обстановку вокруг робота, что может пагубно повлиять на время исполнения и потреблённую электроэнергию заранее выстроенной программы.

Постановка задачи. Задачу повышения эффективности распознавания объектов на видеоданных с мультикамерной системы для предотвращения нештатных ситуаций в экстремальных условиях работы с роботизированными коллаборативными комплексами сформулируем следующим образом. По поступающему видеопотоку с установленных по разным сторонам камер наблюдения необходимо распознать и локализовать в пространстве, а также построить динамическую модель предсказаний движений человека в зоне работы коллаборативного робота.

Алгоритмы обнаружения и локализации ключевых точек в мировой системе координат разбиваются на решение нескольких подзадач:

- выполнение предварительной калибровки многокамерной системы и привязка её к системе координат коллаборативного робота;
- выполнение обнаружения ключевых точек человека на отдельной камере;
- выполнение процесса триангуляции (реконструкции) обнаруженных точек.

Исследование существующих методов и алгоритмов распознавания образов на изображениях. Прежде чем переходить к процессу калибровки и триангуляции (реконструкции) точек, необходимо определиться с базовой моделью глубокого обучения по распознаванию образов на изображениях. На сегодняшний день большой популярностью пользуются модели, которые основываются на графическом подходе и пространственно-временных ассоциациях, использовании region-proposal слоёв, а также скрытых марковских моделях.

Основной структурой данных в графическом методе построения нейронных сетей является сам граф. Граф – тип структуры данных, состоящий из узлов и вершин. Графические нейронные сети – это специализированные типы нейронных сетей, которые могут работать с графическим форматом данных. Существует несколько типов графических нейронных сетей: графические свёрточные нейронные сети [4] (GCNN, от англ. Graph Convolutional Neural Networks), сети автоматического кодирования графов [5], рекуррентные графические нейронные сети (RGNN, от англ. Recurrent Graph Neural Networks) [6].

Методы построения нейронных сетей, использующие region-proposal (RPN, от англ. Region Proposal Network) слои, представляют собой полностью свёрточную сеть, которая одновременно предсказывает границы объекта и оценки в каждой позиции. RPN проходит сквозное обучение для создания высококачественных предсказаний по регионам изображения.

Скрытые марковские модели (СММ) используют статистические свойства сигналов и учитывают непосредственно их пространственные характеристики, поэтому они широко применяются в задачах распознавания как речи, так и изображений. Элементами СММ являются множество наблюдаемых символов, множество различных состояний, набор различных наблюдаемых символов, вектор начальных состояний, матрица переходных вероятностей, матрица вероятностей наблюдаемых символов.

В сравнительном анализе использовались следующие модели: Mask R-CNN [7], AlphaPose [8], HigherHRNet-W48 [9], SPPE [10], OpenPose [11], Assoc. Emb [12], PersonLab [13], MultiPoseNe [14], HigherHRNe [9], openSVAI [15], MIPAL [16], Miracle [17], MSRA/FlowTrack [18], STAF [19] и OpenPifPaf [20]. Вышеописанные модели основываются на нескольких методах. Часть из них относится к методам обнаружения ключевых точек, основанных на графовом подходе и простран-

ственно-временных ассоциациях. Другая часть основывается на использовании region-proposal слоёв и марковских процессов принятия решений.

Для сравнения работы алгоритмов использовались следующие наборы данных:

– Crowd Pose: в наборе [10] предлагается набор поз толпы. Индекс изображения представляет собой степень перекрытия между ограничивающими рамками людей. Набор данных разделён на 10 000 изображений, предназначенных для обучения: 2000 – для проверки и 8000 – для тестового набора.

– COCO: стандартный набор данных для прогнозирования ключевых точек человека [21]. COCO содержит 56 599 разнообразных обучающих изображений с аннотированными ключевыми точками; набор для проверки и тестирования содержит соответственно 5000 и 20 288 изображений.

– Pose Track 2017 и 2018: наборы данных содержат короткие видеоряды с аннотированными и отслеживаемыми позами людей в различных ситуациях. Набор данных Pose Track 2018 содержит 593 сцены обучения, 170 сцен проверки и 375 тестовых сцен.

Для оценки набора данных Crowd Pose и COCO используется метрика оценки сходства ключевых точек объекта (OKS, от англ. Object Keypoint Similarity) [10, 21] для присвоения ограничивающей рамки каждой ключевой точке. Аналогично обнаружению, метрика накладывается между базовым значением и предсказанными ограничивающими рамками для вычисления стандартных показателей обнаружения средней точности (AP, от англ. Average Precision) и средней полноты (AR, от англ. Average Recall). В Crowd Pose тестовый набор разбит на разные уровни изображений: AP_{easy} , AP_{medium} и AP_{hard} . Набор AP_{easy} содержит изображения с индексом в диапазоне от 0 до 0.1, AP_{medium} – от 0.1 до 0.8, AP_{hard} – от 0.8 до 1.

Для оценки Pose Track 2017 и 2018 используется точность отслеживания нескольких объектов (MOTA, от англ. Multi Object Tracker Accuracy). Метрика объединяет ложноположительные (False Positive) результаты, ложноотрицательные (False Negative) результаты и переключатели идентификаторов в единую метрику.

В табл. 1-3 приведены сравнительные показатели выбранных нейросетевых моделей.

Наилучший результат в распознавании показывает модель OpenPifPaf. Модель лучше справляется со статическими изображениями, чем с динамическими. Высокую точность обеспечивают блоки PIF (от англ. Part Intensity Field) и PAF (от англ. Part Association Field), которые совместно обнаруживают и формируют пространственно-временные ассоциации ключевых точек за один этап, что делает её первым алгоритмом обнаружения и отслеживания позы в реальном времени [20].

На вход нейронной сети подаётся изображение размером $H \times W$ с тремя цветовыми каналами. Кодировщик на основе нейронной сети создаёт поля PIF и PAF с каналами 17×5 и 19×7 . После декодер преобразует поля PIF и PAF в оценки позы, содержащие по 17 суставов. Каждое соединение представлено координатами x и y . В качестве кодера используется ResNet. Блоки PIF и PAF представляют собой свёртки размером 1×1 , за которыми следуют свёртки субпикселей. Эти блоки обучены обнаруживать и связывать ключевые точки.

Таблица 1

Сравнительная характеристика работы различных моделей на наборе данных Crowd Pose

Модель	Метрика						
	AP	$AP^{0.5}$	$AP^{0.75}$	AP_{easy}	AP_{medium}	AP_{hard}	FPS
Mask R-CNN	57.2	83.5	60.3	69.4	57.9	45.8	2.9
AlphaPose	61.0	81.3	66.0	71.2	61.4	51.1	10.9
HigherHRNet-W48	65.9	86.4	70.6	73.3	66.5	57.9	-
SPPE	66.0	84.2	71.5	75.5	66.3	57.4	10.1
HigherHRNet-W48+	67.6	87.4	72.6	75.8	68.1	58.9	-
OpenPifPaf	70.5	89.1	76.1	78.4	72.1	63.8	13.7

Таблица 2

Сравнительная характеристика работы различных моделей на наборе данных COCO

Модель	Метрика			
	AP	AP _{medium}	AP _{easy}	t[мс]
OpenPose	61.8	57.1	68.2	100
Assoc. Emb	65.5	60.6	72.6	166
PersonLab	68.7	64.1	75.5	-
MultiPoseNet	69.6	65.0	76.3	43
HigherHRNet	70.5	66.6	75.8	>1000
OpenPifPaf	71.9	68.5	77.4	69

Таблица 3

Сравнительная характеристика работы различных моделей на наборе данных Pose Track

Модель	Метрика		
	MOTA	AP	FPS
openSVAI	54.5	63.1	-
MIPAL	54.9	67.8	-
Miracle	57.4	70.9	-
MSRA/FlowTrac	61.4	74.0	0.7
STAF	53.8	70.3	3
MIPAL	54.5	68.8	-
MSRA/FlowTrack	57.8	74.6	0.7
HRNet	57.9	74.9	-
LightTrack	58.0	66.6	-
OpenPifPaf	60.6	71.5	12.2
KeyTrack	61.2	74.0	1.0

Калибровка камеры. Одним из важных пунктов корректного определения ключевых точек в пространстве служит калибровка камеры. Под калибровкой камеры понимается определение её внутренних параметров, в данном случае матрицы камеры K . Модель камеры можно обобщить, включив в неё радиальную дисторсию и другие артефакты, если приложение нуждается в призонных измерениях. В большинстве случаев простой модели достаточно. Стандартный способ калибровки камеры – многократное фотографирование рабочей области в виде шахматной доски. Такой метод используется в библиотеке OpenCV, которая применяется в данной работе [22].

Рассмотрим простой метод калибровки. Большинство параметров можно задать, приняв простые предположения, так что основную сложность представляет определение фокусного расстояния. Для описываемого метода необходим плоский прямоугольный калибровочный объект, рулетка или линейка и плоская поверхность, при этом выполним ряд условий:

- измерить стороны прямоугольного калибровочного объекта (обозначим их как dX и dY);
- положить камеру и калибровочный объект на плоскую поверхность так, чтобы обратная сторона камеры была параллельно объекту и объект находился примерно в центре поля зрения камеры;
- измерить расстояния между камерой и калибровочным объектом (обозначим его как dZ);
- сделать снимок и проверить, что стороны объекта совмещаются со строками и столбцами матрицы изображения;
- измерить ширину и высоту объекта в пикселях (обозначим их как dx и dy).

Из подобия треугольников вытекает следующая формула, дающая компоненты фокусного расстояния:

$$f = \frac{dx}{dX} * dZ, f = \frac{dy}{dY} * dZ.$$

Камера использовалась с точечной диафрагмой.

Процесс калибровки и размещения камер представлен на рис. 1-2.

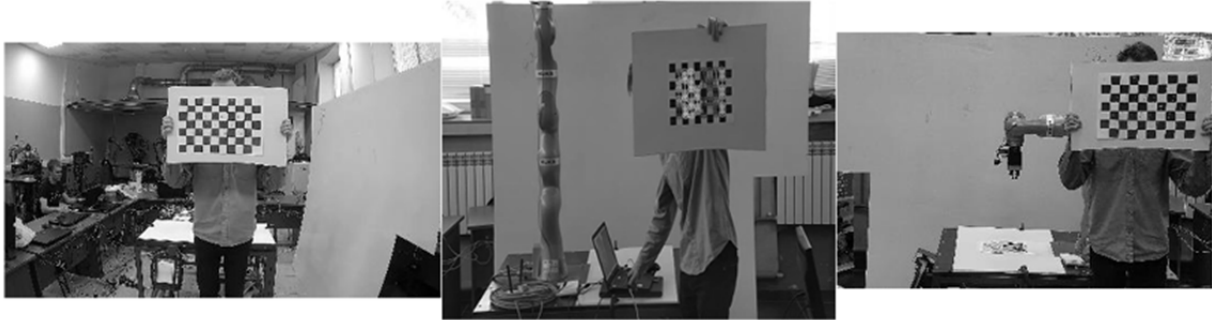


Рис. 1. Калибровка камеры по шаблону

После калибровки камеры формируется матрица **A**:

$$\mathbf{A} = \begin{bmatrix} \frac{f}{w} & 0 & c_{x_0} \\ 0 & \frac{f}{h} & c_{y_0} \\ 0 & 0 & 1 \end{bmatrix},$$

где c_{x_0} и c_{y_0} – естественные координаты главной точки относительно начала координат камеры; w и h – соответственно масштабы вдоль осей x и y ; f – расстояние от оптического центра.

Кроме матрицы **A**, формируется вектор искажений **г**. С помощью уравнения Эйлера получаем повороты углов по осям x (тангаж) и y (рысканье).

Для триангуляции точек в пространстве недостаточно матрицы поворотов и информации о камере. Необходимо определить исходную точку в мировой системе координат и её ориентацию. Простым способом служит выбор позиции камеры № 1 в качестве источника мирового пространства. В общем случае система координат камеры № 1 определяется как находящаяся за экраном камеры точка (см. рис. 3).

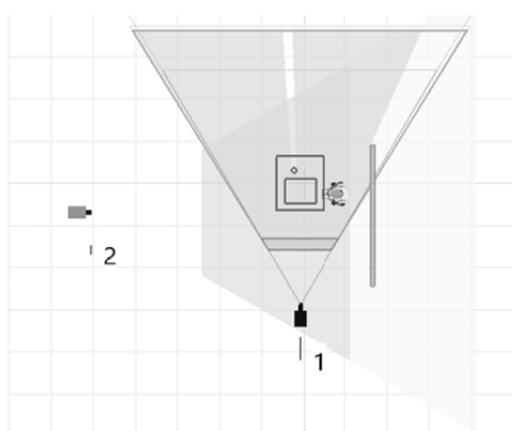


Рис. 2. Наблюдаемая сцена

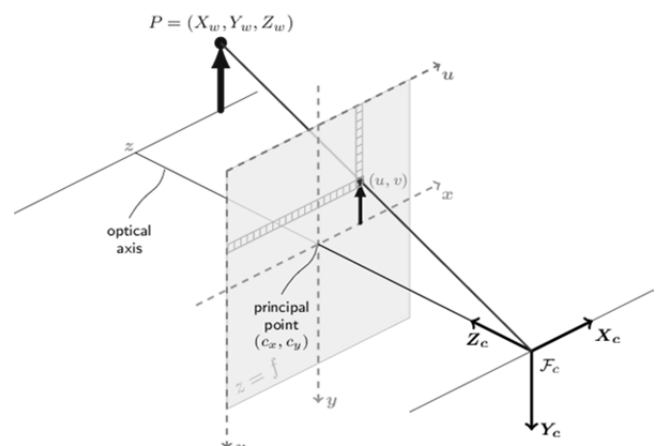


Рис. 3. Процесс триангуляции

Процесс триангуляции (реконструкции) распознанных точек. После того как были определены основные алгоритмы по распознаванию образов на изображении, а также откалиброваны камеры необходимо объединить распознанные точки в одном 3D-пространстве. Для объединения используются прямые линейные преобразования (DLT, от англ. Direct Linear Transforms). Метод DLT основан на разложении по сингулярным значениям (SVD, от англ. Singular Value Decomposition) [23].

DLT – метод вычисления матричного уравнения вида $\mathbf{Ax} = 0$, где \mathbf{A} – некоторая матрица и \mathbf{x} – векторная неизвестная, которые необходимо найти. Предположим, у нас есть трёхмерная точка в реальном пространстве с координатами, заданными как $\mathbf{X} = [x, y, z, w]$. Предположим, мы наблюдаем эту точку с помощью двух камер, которые имеют пиксельные координаты $\mathbf{U}_1 = [u_1, v_1, 1]$ для камеры № 1 и $\mathbf{U}_2 = [u_2, v_2, 1]$ для камеры № 2. Используя матрицу проекции камеры \mathbf{P}_1 , мы можем записать \mathbf{U}_1 как

$$\vec{U}_1 = \alpha \mathbf{P}_1 \vec{x}.$$

В задаче триангуляции (триангуляция – процесс определения точки в трёхмерном пространстве с учётом её проекций на два или более изображений) не известен вектор \mathbf{x} . Можно определить координаты пикселей, а также предположить, что мы нашли проекционную матрицу с помощью калибровки камеры. Задача состоит в нахождении неизвестных в \mathbf{x} . Поскольку \mathbf{U}_1 и $\mathbf{P}_1 \mathbf{x}$ являются параллельными векторами, их перекрёстное произведение должно быть равно нулю:

$$\begin{bmatrix} u_1 \\ v_1 \\ 1 \end{bmatrix} \times \begin{bmatrix} \vec{P}_1 \vec{x} \\ \vec{P}_2 \vec{x} \\ \vec{P}_3 \vec{x} \end{bmatrix} = \begin{bmatrix} v_1 \vec{P}_3 \vec{x} - \vec{P}_2 \vec{x} \\ \vec{P}_1 \vec{x} - u_1 \vec{P}_3 \vec{x} \\ u_1 \vec{P}_2 \vec{x} - v_1 \vec{P}_1 \vec{x} \end{bmatrix} = \begin{bmatrix} v_1 \vec{P}_3 - \vec{P}_2 \\ \vec{P}_1 - u_1 \vec{P}_3 \\ u_1 \vec{P}_2 - v_1 \vec{P}_1 \end{bmatrix} \vec{x} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Вышеприведённое уравнение формирует уравнение вида $\mathbf{Ax} = 0$. Но третья строка представляет собой линейную комбинацию первых двух строк, которая затем формирует только 2 системы уравнений, чего недостаточно для решения 3 неизвестных в \mathbf{x} . Это ожидаемо, поскольку невозможно определить 3D-координату с одного вида камеры. Поскольку используется две камеры, расширим матрицу, чтобы иметь больше строк. Фактически просто добавляется больше строк для любого количества просмотров:

$$\mathbf{A} \vec{x} = \begin{bmatrix} v_1 \vec{P}_3 - \vec{P}_2 \\ \vec{P}_1 - u_1 \vec{P}_3 \\ v_2 \vec{P}_3 - \vec{P}_2 \\ \vec{P}_1 - u_2 \vec{P}_3 \\ \vdots \end{bmatrix} \vec{x} = 0.$$

При триангуляции камеры нам задаётся \mathbf{A} и требуется определить \mathbf{x} . В этой настройке используем DLT для определения \mathbf{x} .

Необходимо получить нетривиальное решение уравнения вида $\mathbf{Ax} = 0$. Поскольку возможен некоторый шум, запишем уравнение в виде $\mathbf{x} = \mathbf{w}$ и решим уравнение для таких \mathbf{x} , что \mathbf{w} было сведено к минимуму. Первым шагом является определение SVD-разложения \mathbf{A} :

$$\mathbf{A} \vec{x} = \mathbf{USV}^T \vec{x}.$$

Наша цель – минимизировать \mathbf{w} для некоторых \mathbf{x} . Это можно сделать, взяв скалярное произведение:

$$\vec{w}^T \vec{w} = (\vec{x}^T \mathbf{VSU}^T) \cdot (\mathbf{USV}^T \vec{x}) = \vec{x}^T \mathbf{VS}^2 \mathbf{V}^T \vec{x},$$

где \mathbf{U} и \mathbf{V} – ортонормированные матрицы, а \mathbf{S} – диагональная матрица. Записи на диагонали \mathbf{S} уменьшаются, так что последняя запись на диагонали является минимальным значением. Это га-

рантируется разложением SVD. Используем свойство, что V является ортонормированной матрицей, если мы просто выбираем x – один из векторов столбцов V^T :

$$\vec{v}_i^T V S^2 V^T \vec{v}_i = s_i^2.$$

Минимальное значение получается, если мы выбираем вектор последнего столбца V как x . Таким образом, было решено уравнение вида $x = w$ с присутствием шума. Если шума нет, метод SVD всё равно будет работать.

После проведения прямого линейного преобразования появляется возможность размещения распознанных образов на изображениях в 3D-пространстве. На рис. 4 приведены полученные результаты.



Рис. 4. Триангуляции распознанных точек в 3D-пространстве

Из рис. 4 видно, как модель OpenPifPaf справляется с распознанными точками. С помощью DLT преобразования отстраивается 3D-модель по ключевым точкам в пространстве.

Выводы. В работе проведён сравнительный анализ методов распознавания образов на изображениях. Методы, основывающиеся на использовании region-proposal слоёв и марковских процессов принятия решений, в сравнении с методом, основывающимся на графическом подходе и пространственно-временных ассоциациях, демонстрируют низкие метрические показатели. Тестирование проводилось на трёх разных наборах данных: Crowd Pose; COCO; Pose Track. В качестве оценки адекватности работы модели использовались следующие метрические показатели: для набора данных Crowd Pose и COCO – AP; для набора данных Pose Track – MOTA.

На основе выбранного нейросетевого алгоритма распознавания ключевых точек на теле человека удалось построить динамическую модель предсказаний движений человека в зоне работы коллаборативного робота. Процесс был реализован при помощи использования откалиброванной мультикамерной системы и триангуляции данных распознанных ключевых точек на теле человека.

В результате анализа был выявлен наилучший алгоритм (OpenPifPaf) по метрическим показателям, который основан на графическом подходе и пространственно-временных ассоциациях. Полученный результат позволил сформировать дальнейшее видение разработки и синтеза мультимодальных адаптивных алгоритмов при работе с коллаборативными робототехническими системами с учётом нештатных ситуаций и экстремальных условий в недетерминированной среде.

ЛИТЕРАТУРА

1. ГОСТ Р 59799-2021. Умное производство. Модель эталонной архитектуры индустрии 4.0 (RAMI 4.0) = Smart manufacturing. Reference architecture model industry 4.0 (RAMI 4.0): издание официальное: утверждён и введён в действие Приказом Федерального агентства по техническому регулированию и метрологии от 25 октября 2021 г. № 1301-ст: введён впервые: дата введения 2022-04-30 / подготовлен Федеральным государственным автономным образовательным учреждением высшего образования «Национальный исследовательский университет «Высшая школа экономики». – М.: ФГАОУВО «Национальный исследовательский университет «Высшая школа экономики». – 36 с.

2. Индустрия 5.0: Основы создания системной тетрады киберсоциальных экосистем / А. В. Бабкин, И. В. Либерман, П. М. Клачек, Е. В. Шкарупета // Вестник АГТУ. Серия: Экономика. – 2023. – № 1. – С. 171-174.
3. Галин, Р. Р. Эффективное функционирование коллаборативной робототехнической системы в едином пространстве / Р. Р. Галин, С. Б. Камешева // Известия КБНЦ РАН. – 2021. – № 1 (99). – С. 5-14.
4. Chen Q., Wu R. CNN is all you need // arXiv preprint arXiv:1712.09662. – 2017.
5. Makhzani A. et al. Adversarial autoencoders // arXiv preprint arXiv:1511.05644. – 2015.
6. Zhang Y., Huang D., Wang Y. PC-RGNN: Point cloud completion and graph neural network for 3D object detection // Proceedings of the AAAI conference on artificial intelligence. – 2021. – Т. 35. – № 4. – P. 3430-3437.
7. He K. et al. Mask r-cnn // Proceedings of the IEEE international conference on computer vision. – 2017. – P. 2961-2969.
8. Fang H. S. et al. Rmpe: Regional multi-person pose estimation // Proceedings of the IEEE international conference on computer vision. – 2017. – P. 2334-2343.
9. Cheng B. et al. Higherhrnet: Scale-aware representation learning for bottom-up human pose estimation // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. – 2020. – P. 5386-5395.
10. Li J. et al. Crowdpose: Efficient crowded scenes pose estimation and a new benchmark // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. – 2019. – P. 10863-10872.
11. Cao Z. et al. Realtime multi-person 2d pose estimation using part affinity fields // Proceedings of the IEEE conference on computer vision and pattern recognition. – 2017. – P. 7291-7299.
12. Newell A., Huang Z., Deng J. Associative embedding: End-to-end learning for joint detection and grouping // Advances in neural information processing systems. – 2017. – Т. 30.
13. Papandreou G. et al. Personlab: Person pose estimation and instance segmentation with a bottom-up, part-based, geometric embedding model // Proceedings of the European conference on computer vision (ECCV). – 2018. – P. 269-286.
14. Kocabas M., Karagoz S., Akbas E. Multiposenet: Fast multi-person pose estimation using pose residual network // Proceedings of the European conference on computer vision (ECCV). – 2018. – P. 417-433.
15. Guo C. et al. OpenWSI: a low-cost, high-throughput whole slide imaging system via single-frame autofocusing and open-source hardware // Optics Letters. – 2020. – Т. 45. – № 1. – P. 260-263.
16. Chen Y. et al. Cascaded pyramid network for multi-person pose estimation // Proceedings of the IEEE conference on computer vision and pattern recognition. – 2018. – P. 7103-7112.
17. Xiao B., Wu H., Wei Y. Simple baselines for human pose estimation and tracking // Proceedings of the European conference on computer vision (ECCV). – 2018. – P. 466-481.
18. Raaj Y. et al. Efficient online multi-person 2d pose tracking with recurrent spatio-temporal affinity fields // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. – 2019. – P. 4620-4628.
19. Sun K. et al. Deep high-resolution representation learning for human pose estimation // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. – 2019. – P. 5693-5703.
20. Kreiss S., Bertoni L., Alahi A. Pifpaf: Composite fields for human pose estimation // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. – 2019. – P. 11977-11986.
21. Lin T. Y. et al. Microsoft coco: Common objects in context // Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13. – Springer International Publishing, 2014. – P. 740-755.
22. Solem J. E., Programming Computer Vision with Python // O'Reilly Media Inc., 1005 Gravenstein Highway North Sebastopol, CA 95472, 2016. P. 84-117.
23. Direct Linear Transforms (DLT) // Temugeb. – URL: https://temugeb.github.io/computer_vision/2021/02/06/direct-linear-transforms.html (дата обращения: 10.03.2023). – Текст: электронный.