УДК 004.932.2 DOI 10.18522/2311-3103-2026-2-286-298

К.И. Морев

СИСТЕМА СОПОСТАВЛЕНИЯ ИЗОБРАЖЕНИЙ С ИСПОЛЬЗОВАНИЕМ ИНТУИЦИОНИСТСКИХ НЕЧЕТКИХ МНОЖЕСТВ

Представлена полностью обучаемая система для решения задачи сопоставления двух изображений. Все основные элементы системы являются обучаемыми, т.е. их финальный вид соответствует целевому набору данных по которому проводилось обучение. Факт обучаемости системы, методы, используемые при обучении, и архитектура системы позволяют использовать систему для решения большого числа разнообразных задач компьютерного зрения. Система состоит из сверточной нейронной сети, служащей одновременно для выделения особых точек и их дескрипторов, а также обучаемого модуля сопоставления выделенных особых точек на основе их описания и взаимного расположения в наблюдаемой сцене. Используемая сверточная нейронная сеть обрабатывает полноразмерные изображения и вычисляет как местоположение точек интереса с пиксельной точностью, так и связанные с ними дескрипторы за один полный цикл вычислений. Сопоставление особых точек – отдельный шаг и выполняется после выполнения полного цикла вычислений нейронной сети, но на основе результатов этих вычислений. В процессе обучения модели вычисления положений особых точек и их дескрипторов используется метод формирования обучающей выборки, называемый гомографическая адаптация – подход, способствующий повышению повторяемости и точности обнаружения особых точек. Процесс обучения модели выделения особых точек заключается в получении новых весов в процессе дообучения базового детектора, представляющего инициализирующие веса модели. Финальная модель выделения особых точек, обученная на универсальном наборе изображений MS-COCO с использованием гомографической адаптации, многократно превосходит исходный базовый детектор по параметрам количества, надежности и повторяемости особых точек, а также превосходит любой другой традиционный детектор уголков основанный на классических подходах.

Машинное обучение; сверточная нейронная сеть; сопоставление изображений; особые точки изображения; сопоставление особых точек; интуиционистская нечеткая логика.

K.I. Morev

IMAGE MATCHING SYSTEM WITH USING INTUITIONISTIC FUZZY SETS

This paper presents a fully learnable system for solving the problem of matching two images. All the main elements of the system are trainable, i.e. their final form corresponds to the target dataset on which the training was carried out. The fact that the system is trainable, the methods used in training and the architecture of the system allow using the system to solve a large number of various computer vision problems. The system consists of a convolutional neural network that serves both to extract key points and their descriptors, as well as a trainable matcher of the extracted key points based on their description and mutual arrangement in the observed scene. The used convolutional neural network processes full-size images and calculates both the location of interest points with pixel accuracy and the descriptors associated with them in a single forward pass. Matching key points is a separate step and is performed after the forward pass of the neural network. In the process of training the model for calculating the positions of key points and their descriptors, a method for forming a training sample is used, called homographic adaptation - an approach that helps to increase the repeatability and accuracy of detecting key points. The process of training the feature point detection model consists of obtaining new weights in the process of additional training of the base detector, which represents the initialization weights of the model. The final feature point detection model, trained on the universal MS-COCO image set using homographic adaptation, repeatedly outperforms the original base detector in terms of the number, reliability and repeatability of feature points, and also outperforms any other traditional corner detector based on classical approaches.

Machine learning; convolutional neural network; image matching; image local features; feature point matching; intuitionistic fuzzy logic.

Введение. Первым шагом в задачах компьютерного зрения с геометрическим контекстом, таких как одновременная локализация и картографирование [1], восстановление структуры из движения [2], калибровка камеры или сопоставление изображений является извлечение точек интереса из изображений [3]. Точка интереса – это двумерная область на изображении, хорошо локализуемая и не изменяющая своих характеристик при различных условиях освещения и ракурсах [4]. Большинство алгоритмов раздела компью-

терного зрения, известного как «геометрия множества ракурсов», основаны на предположении, что точки интереса могут быть надежно извлечены и сопоставлены на изображениях. Однако в составе реальных систем компьютерного зрения входными данными для алгоритмов являются необработанные изображения извлечение из которых идеализированных точек интереса и их описаний на самом деле является сложной задачей. Многими исследователями было показано, что сверточные нейронные сети превосходят вручную разработанные методы практически во всех задачах, использующих изображения в качестве входных данных. В частности, полностью сверточные нейронные сети, которые выделяют двумерные ключевые точки, хорошо изучены для различных таких задач как оценка позы человека, обнаружение объектов и построение карты помещения [5]. Обучение таких моделей основано на обучении с учителем и большом наборе истинных данных о связи между элементами сцены на различных изображениях, а также о взаимосвязях внутренних элементов одного изображения, другими словами, на размеченных данных.

Используемая в данной работе нейронная сеть выделения особых точек SuperPoint [6] предлагает методики обучения с применением техник самообучения. Вместо использования вручную размеченных данных и строгого контроля, в процессе обучения создается большой набор данных псевдоистинных точек интереса на реальных изображениях, сформированных некоторым базовым детектором точек интереса, и не требующих масштабной человеческой работы по разметке.

Для получения базового детектора и генерации псевдоистинных точек интереса, сначала обучается полностью сверточная нейронная сеть на миллионах примеров из созданного синтетического набора данных. Синтетический набор данных состоит из простых геометрических фигур с однозначно известным расположением точек интереса. Полученный обученный детектор будет являться базовым для инициализации весов итогового обучаемого детектора. Очевидно, что по сравнению с классическими детекторами точек интереса на разнообразном наборе реальных текстур и узоров изображений, базовый детектор пропускает многие потенциально надежные точки интереса. Для устранения этого недостатка, используется многомасштабный метод синтезированных преобразований – гомографическая адаптация. Гомографическая адаптация предназначена для самоконтролируемого обучения детекторов точек интереса. Адаптация заключается в многократной деформации входного изображения, для имитации наблюдения сцены с множества различных ракурсов и при различных масштабах.

После обнаружения надежных и повторяемых точек интереса, для решения более высокоуровневых семантических задач, например, сопоставления изображений, требуется назначение каждой точке вектора фиксированного размера в качестве дескриптора. Используемая модель SuperPoint устроена таким образом, что все выделенные карты признаков идут одновременно и в модуль выделения особых точек, и в модуль вычисления их дескрипторов. Далее вычисленные особые точки и их описания подаются на вход модулю сопоставления, который на основе нечетких функций принадлежности решает задачу назначения соответствий между точками. Результирующая система показана на рис. 1.

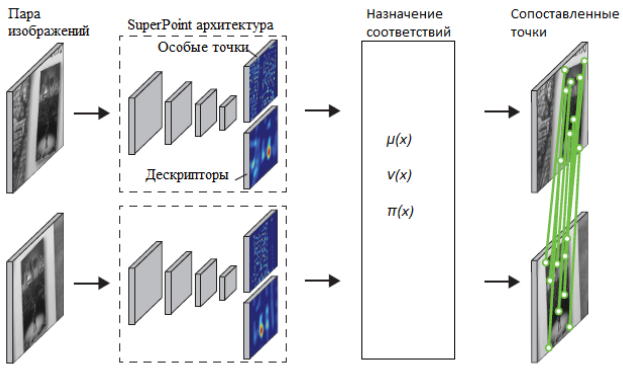

Рис. 1. Общая схема обучаемой системы сопоставления изображений

Схожие подходы. Традиционные детекторы точек интереса были тщательно изучены ранее [7, 8]. Детектор углов FAST был первой попыткой решения задачи выделения особых точек с помощью машинного обучения, а масштабно-инвариантное преобразование признаков, или SIFT, до сих пор, является наиболее известным дескриптором локальных признаков в компьютерном зрении, основанным на традиционных подходах. Используемая архитектура SuperPoint вдохновлена достижениями в применении глубокого обучения в обнаружении точек интереса и обучении дескрипторов. В способности выявлять и сопоставлять внутренние паттерны в изображениях, SuperPoint подобен UCN [9] и в некоторой степени DeepDesc [10]. Однако оба упомянутых алгоритма не выполняют обнаружения точек интереса, а только вычисление их дескрипторов.

Другая альтернатива LIFT – сверточная сеть, реализующая близкую к традиционной цепочке «обнаружить, затем описать» последовательность на основе патчей. Конвейер LIFT включает обнаружение точек интереса, оценку ориентации и вычисление дескрипторов, но дополнительно требует контроля со стороны и разметки при формировании данных для обучения.

Примером обучения без учителя выступил Quad-Networks. Однако эта архитектура основана на патчах (входные данные представляют собой небольшие фрагменты изображения) и является относительно мелкой двухслойной сетью. Система обнаружения точек интереса TILDE использовала принцип, аналогичный принципу гомографической адаптации. Однако этот подход не использует мощь больших сверточных нейронных сетей для выделения признаков.

Архитектура SuperPoint. SuperPoint – архитектура полностью сверточной нейронной сети, которая работает с полноразмерным изображением и выполняет обнаружение точек интереса, а также вычисление для обнаруженных точек дескрипторов, за один прямой проход (рис. 2).

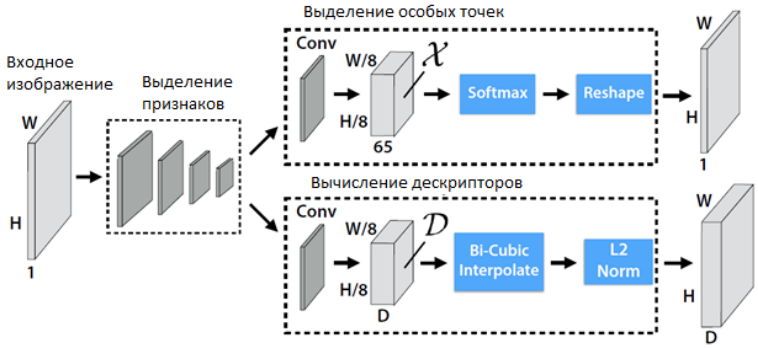

Рис. 2. Схема архитектуры SuperPoint.

Модель SuperPoint имеет один общий блок выделения признаков или «кодер». На рис. 2 «кодер» обведен мелким пунктиром и подписан «Выделение признаков». «Кодер» занимается обработкой входного изображения с целью уменьшения его размерности и выделения признаков. После «кодера» архитектура разделяется на две «головы». «Голова» – отдельная часть архитектуры сети, отвечающая за решение своей конкретной задачи, иначе «голову» можно назвать «декодер». На рис. 2 пунктиром выделены две «головы» архитектуры, вычисляемые после блока выделения признаков. «Голова», отвечающая за выделение особых точек на изображении, обведена жирным пунктиром и подписана «Выделение особых точек». «Голова», отвечающая за формирование описаний для выделенных особых точек, аналогично обведена жирным пунктиром и подписана «Вычисление дескрипторов». Схема на рис. 2 наглядно демонстрирует наличие общего «кодера» для обеих «голов», что отличает SuperPoint от традиционных систем, которые сначала обнаруживают точки интереса, а затем вычисляют дескрипторы и не могут совместно использовать один общий набор признаков [11].

Общий «кодер». SuperPoint использует кодер в стиле VGG [12] для уменьшения размерности изображения. Кодер составляют сверточные слои, блок понижения пространственной размерности и нелинейные функции активации. Выходное изображение «кодера» состоит из пикселей, полученных путем серии сверток и выбора максимума результата свертки в области 2x2 пикселя. Выбор максимума в области 2x2 пикселя, или «макс-пулинг» ядром размера 2x2, понижает размерность изображения в два раза, т.к. из каждой ячейки 2x2 исходного изображения выбирается только 1 пиксель. Три последовательных (по разу в каждом новом представлении внутри «кодера») операции «макс-пулинг» уменьшают размерность итогового изображения в 8 раз по каждой из осей. Кодер преобразует входное изображение I в промежуточный тензор B с меньшей пространственной размерностью, но большим количеством каналов так, что

$$I \in R^{H \times W}$$

и

$$B \in R^{H_c \times W_c \times F}.$$

Для обнаружения особых точек решается задача назначения каждому пикселю выходных данных сети вероятности «принадлежности» к классу особая точка. Эта вероятность далее восстанавливается для соответствующего пикселя на входных данных. Стандартная конструкция сети для попиксельной классификации включает в себя пару «кодер-декодер», где пространственное разрешение уменьшается с помощью объединения или пошаговой свертки, а затем повышается до полного разрешения с помощью операций повышения размерности, например, как это делается в SegNet. Однако, слои, повышающие размерность, как правило, требуют большого объема вычислений и могут вносить нежелательные артефакты [13], поэтому после получения признаков для выделения особых точек используется специальный декодер без масштабирования, чтобы сократить объем вычислений модели.

«Голова» детектора из входного набора признаков вычисляет представление χ

$$\chi \in R^{H_c \times W_c \times 65}. \quad (1)$$

И формирует тензор

$$R^{H \times W}, \quad (2)$$

где 65 каналов соответствуют локальным, неперекрывающимся областям сетки 8x8 пикселей и дополнительный канал со слабым откликом «без точек интереса». После применения функции «softmax» для каждого канала и удаления канала со слабым откликом «без точек интереса» выполняется изменение формы тензора таким образом, что представление из выражения (1) преобразуется в (2):

$$R^{H_c \times W_c \times 65} \rightarrow R^{H \times W}.$$

«Голова» для дескрипторов вычисляет представление D

$$D \in R^{H_c \times W_c \times D},$$

и формирует тензор размером

$$R^{H \times W \times D}.$$

Для вывода плотной карты L2-нормализованных дескрипторов фиксированной длины используется модель, похожая на UCN, чтобы вывести разреженную сетку дескрипторов (например, один на каждые 8 пикселей). Формирование дескрипторов выполняется для разреженной сетки, а не плотной, для уменьшения требований памяти при обучении и сокращения времени вычисления дескрипторов во время выполнения [14]. Далее в декодере выполняется бикубическая интерполяция разреженных дескрипторов для вычисления плотной сетки и затем L2-нормализует полученные дескрипторы для формирования векторов единичной длины.

Гомографическая адаптация. SuperPoint при обучении требует начальные веса базового детектора точек интереса и большой набор неразмеченных изображений (например, MS-COCO). Работая в парадигме обучения без учителя, сначала генерируется набор псевдоистинных точек интереса для каждого изображения в имеющемся наборе, а далее

используется традиционный механизм обучения. В основе описываемого метода лежит процесс, который применяет случайные матрицы гомографии для преобразования копий входного изображения с последующим выделением особых точек на измененных копиях и репроекции их на оригинальное изображение. Объединение результатов репроецированных особых точек – процесс, который называется гомографической адаптацией (рис. 3).

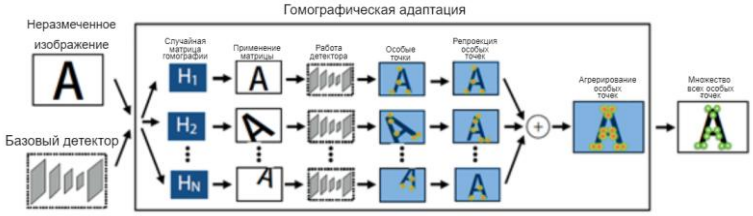

Рис. 3. Гомографическая адаптация

Используются матрицы гомографии, потому что они дают точные преобразования исходного изображения в другое изображение в случае имитации движения камеры с вращением вокруг оптической оси или для сцен с большими расстояниями до объектов и плоских сцен. В дополнение, поскольку проекция мира в кадр – плоская, то матрица гомографии является хорошей моделью, описывающей изменения, возникающие при наблюдении одной точки объемного мира с разных ракурсов.

Пускай преобразование $f_\theta(.)$ представляет собой функцию выделения идеальных точек интереса, которую мы хотим аппроксимировать, I – входное изображение, x – полученные точки интереса, а H – случайная матрица гомографии. Таким образом

$$x = f_\theta(I).$$

Идеальный оператор точки интереса должен быть ковариантным относительно матрицы гомографии. Функция $f_\theta(.)$ ковариантна H, если выход преобразуется вместе с входом. Другими словами, ковариантный детектор будет удовлетворять для любой H:

$$Hx = f_\theta(H(I)), \tag{3}$$

перемещая операторы гомографии вправо, получаем:

$$x = H^{-1} f_\theta(H(I)). \tag{4}$$

На основании выражения (4) можно сформировать следующий вывод: для произвольного детектора особых точек $f_\theta(.)$, и двух изображений I и $I`=H(I)$, связанных матрицей гомографии $H$, справедливо что выделенные на I особые точки $x=f_\theta(I)$ и выделенные на $I`$ особые точки $x`=f_\theta(I`)$ будут идентичны и отличаться только положением (в соответствии с преобразованием между изображениями H). В этом и заключается свойство ковариантности детектора $f_\theta(.)$ относительно матрицы H.

Однако на практике детектор не будет идеально ковариантным – разные матрицы гомографии в выражении (3) выше приведут к разным особым точкам x. Основная идея гомографической адаптации заключается в проведении суммирования всех особых точек по достаточно большой выборке случайных H (см. рисунок 3). Результирующее агрегирование по выборкам, таким образом, приводит к новому и улучшенному детектору $F(.)$:

$$F(I, f_\theta) = \frac{1}{N_h} \sum_{i=1}^{N_h} H^{-1} f_\theta(H(I)).$$

Сопоставление особых точек. После выполнения процедуры выделения и описания особых точек на двух изображениях следует процедура их сопоставления. Сопоставление особых точек заключается в поиске для каждой выделенной особой точки (далее – ОТ) ее же отображения на другом изображении сцены [15, 16]. В общем случае, точки на изображениях сцены связаны некоторым линейным преобразованием M, взаимно трансформирующим один набор точек в другой [17]. В рамках представленной работы исследуется группа линейных геометрических преобразований плоскости, ограниченная только смещением, поворотом, масштабом.

При решении задачи сопоставления найденных ОТ используется обучаемый модуль сопоставления, задача которого – выделить из набора всех возможных соответствий между точками только истинные. Другими словами – из множества всех соответствий выделить подмножество истинных соответствий. При разделении подмножеств, в качестве признаков используются интуиционистские степени принадлежности и частично теория интуиционистских нечетких множеств (далее - ИНМ). Процесс формирования значений принадлежности, непринадлежности и неопределенности раскрывается ниже.

Для сопоставления, пользуясь методом полного перебора, сравниваются дескрипторы (описания) окрестностей всех точек [18]. Результатом сравнения дескрипторов является расстояние между описаниями ОТ [19]. Это расстояние формируется как Эвклидово расстояние в признаковом пространстве между двумя векторами дескрипторов. Так получаем количественную оценку разницы окрестностей для всех ОТ двух изображений. Полученное множество пар соответствий фильтруется по относительному порогу:

$$f(dif) = 1 - \frac{dif}{desc\_vec\_len},$$

$$\begin{cases} match = true, \text{когда } f(dif) \geq thr \\ match = false, \text{когда } f(dif) < thr \end{cases},$$

где dif – вычисленное расстояние, $desc\_vec\_len$ – длина вектора дескриптора, thr – выбираемый в процессе работы (обучение) порог в пределах [0; 1], а match – истинность проверяемого соответствия. Таким образом из всего множества соответствий выбираются пары, точки в которых схожи по дескрипторам (относительно обучаемого порогового значения *thr*).

Значение функции непринадлежности *v*(*x*) пары ко множеству истинных соответствий определяется разностью описаний точек. Чем больше величина этой метрики, тем выше степень несоответствия, и эта величина интерпретируется как степень ложности соответствия. Графическое представление функции непринадлежности показано на рис. 4, при этом максимальное значение разницы между векторными описаниями равняется 2, что соответствует максимально возможному расстоянию между векторами нормированной длины.

$$v(x) = 1 - \frac{\sqrt{\sum_{k=1}^{n}(a_k - b_k)^2}}{2},$$

где x – оцениваемое соответствие, a и b – дескрипторы ОТ из соответствия x.

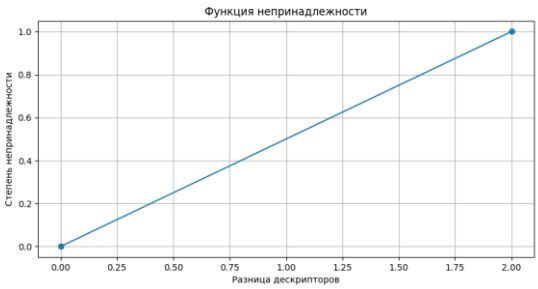

Рис. 4. Функция непринадлежности, v(x)

Далее, для прошедших порог соответствий, вычисляется структура сопоставленных точек: случайно взятое соответствие назначается опорным и от точки опорного соответствия на первом (из двух сопоставляемых изображений) назначается дистанция (в пикселях) и угол поворота относительно всех остальных сопоставленных точек на первом изображении. Аналогично и для второго изображения.

Дистанция между точками определяется как Эвклидово расстояние [20]:

$$D_{ij} = \sqrt[2]{(X_i - X_j)^2 + (Y_i - Y_j)^2},$$

где D_{ij} обозначает дистанцию, i и j – индексы разных точек, а X и Y представляют собой координаты центров особых точек в системе координат изображения. В системе координат изображения дистанция измеряется в пикселях. Угол поворота между двумя точками вычисляется как угол между вектором, который соединяет опорную точку с оцениваемой, и положительным направлением оси Ox (направление осей Ox, Oy изображено на рисунке 5 векторами белого цвета). Таким образом, в результате построения структуры особых точек, для каждой прошедшей фильтрацию особой точки имеется дистанция и угол поворота относительно опорной точки в изображении.

На рис. 5 отображён процесс построения структуры. Соответствия, прошедшие фильтрацию по заданному порогу сходства дескрипторов, выделены жёлтыми линиями. Синим цветом изображены линии построения структуры сопоставленных ОТ. Бирюзовыми точками обозначены центры сопоставленных особых точек. Построение структуры в каждом изображении начинается от опорной точки, отмеченной буквой «O» красного цвета и крестом, и заключается в определении дистанций и углов поворота до всех остальных сопоставленных точек, прошедших фильтрацию по порогу.

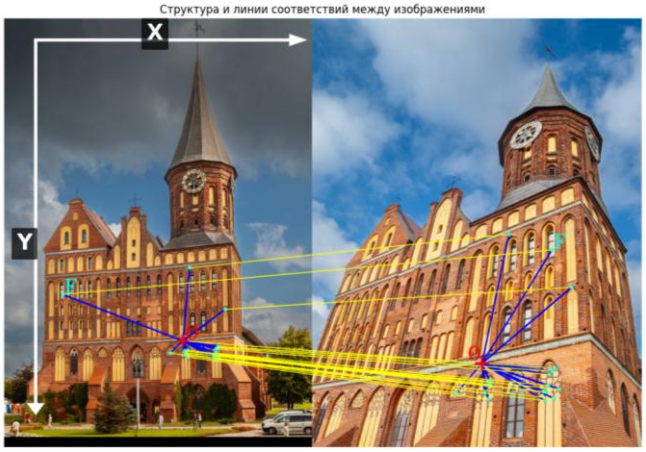

Рис. 5. Пример определения структуры ОТ на изображениях

Стоит отметить, что на рис. 5 есть одно ложное соответствие. Точки, соединяемые этим соответствием обозначены буквой «F» бирюзового цвета. В данном случае ложное соответствие соединяет похожие по признаковому описанию особые точки (на первом и втором изображениях особая точка «F» – это кусок оконной рамы), однако соединенные особые точки изображений отображают различные точки реального 3-Д пространства.

Для каждой пары соответствующих точек, которые прошли фильтрацию, известны дистанция и угол поворота относительно опорной точки в изображении. При предположении, что истинное соответствие связывает одну и ту же точку сцены на двух различных изображениях, можно считать, что искомое преобразование M действует по универсальному правилу. В результате все истинные соответствия формируют кластер точек, которые перемещаются в пространстве сходным образом. Это означает, что структура точек сохраняется, а их дистанции и углы поворота изменяются пропорционально. В идеальном случае все истинные соответствия образуют кластер, для которого выполняется следующее условие:

$$\begin{cases} \dfrac{M_1 dist_a}{M_1 dist_b} = \dfrac{M_2 dist_a}{M_2 dist_b} = \cdots = \dfrac{M_n dist_a}{M_n dist_b}, \\ M_1 angle_a - M_1 angle_b = \cdots = M_n angle_a - M_n angle_b, \end{cases}$$

где $M_i dist_a$ и $M_i dist_b$ – значения дистанций ОТ для сопоставляемых изображений соответственно из i-го соответствия, а $M_i angle_a$ и $M_i angle_b$ – значения углов поворота ОТ.

Анализируя изменения углов и дистанций для истинных соответствий, можно определить угол поворота сцены и масштаб её изображения [22]. Изменение масштаба сцены вычисляется как отношение дистанций особых точек первого изображения к дистанциям особых точек второго изображения, тогда как поворот определяется как разница углов поворота особых точек точки первого и второго изображений.

Тем не менее, на практике данные часто бывают сильно зашумлены. Шум возникает из-за особенностей работы детекторов ОТ, ошибок субпиксельного позиционирования центров ОТ и неточностей при сравнении соседних областей [23, 24]. При предположении, что шум подчиняется нормальному распределению, за истинные соответствия принимаются те, у которых изменения дистанции и угла поворота укладываются в установленные границы:

$$\begin{cases} \delta_{dist} \pm \sigma_{\text{dist}} \\ \delta_{angle} \pm \sigma_{\text{angle}} \end{cases},$$

где δ – среднее значение, а σ – среднеквадратичное отклонение (далее – СКО) для распределения изменений дистанций (*dist*) и угла поворота (*angle*).

То есть, после вычисления изменений дистанции и угла поворота для всех проверяемых соответствий, определяется среднее значение δ и стандартное отклонение σ для всего распределения. Если изменение дистанции или угла поворота для конкретного соответствия отклоняется от среднего более чем на установленный порог, такое соответствие признается ложным. В случае, когда относительное отклонение меньше порогового значения, оно рассматривается как степень принадлежности к истинным соответствиям. На основе этого формируется функция принадлежности или степень истинности, которая имеет следующий вид:

$$\mu(\text{x}) = \max\left(1 - \frac{|dist - \delta_{dist}|}{\sigma_{\text{dist}}}, 1 - \frac{|angle - \delta_{angle}|}{\sigma_{\text{angle}}}\right),$$

где *dist* – изменение дистанции ОТ, *angle* – изменение ее угла поворота, δ_{scale} и δ_{rotate} – средние значения распределения изменений дистанции и угла поворота соответственно, а σ_{dist} и σ_{angle} – соответствующие СКО. На рис. 6 отображена функция принадлежности $\mu(x)$.

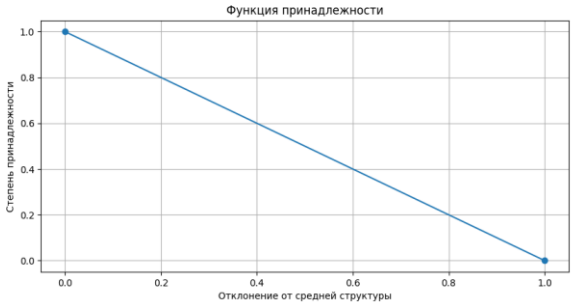

Рис. 6. Функция принадлежности, μ(x).

Таким образом, для каждого соответствия вычисляется степень его истинности и ложности. Однако согласно свойствам интуиционистских нечетких множеств, сумма значений функций принадлежности и непринадлежности не должна превышать единицу. При использовании приведённых выше функций оценки соответствия и несоответствия могут возникать ситуации, когда это условие нарушается (рис. 7, левая часть).

Для решения этой проблемы использовался подход, предлагаемый Атанасовым К., описанный в [25]. Подход заключается в нормализации имеющихся значений функций принадлежности и непринадлежности. В результате нормализации сумма значений функций не превышает 1. Графическое представление нормализации изображено на рис. 7 и описывается переходом от единичного квадрата, изображенного на рис. 7 слева, к классической треугольной интерпретации интуиционистского нечеткого множества (см. рис. 7 справа).

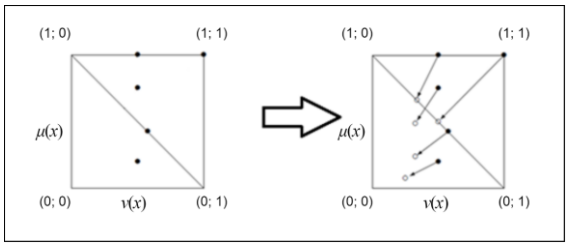

Рис. 7. Интерпретация ИНМ при заданных функциях μ(x) и ν(x). Слева некорректная форма ИНМ, в которой сумма μ(x) и ν(x) может превышать 1. Справа преобразованное в корректную форму ИНМ

Нормализация осуществляется при помощи непрерывного биективного преобразования:

$$F(\nu,\mu)=\begin{cases} <0{,}0>, \text{при } \nu=\mu=0 \\ <\dfrac{\mu^2}{\mu+\nu},\dfrac{\mu\nu}{\mu+\nu}>, \text{при } \nu,\mu\in[0{,}1] \text{ и } \nu\geq\mu \text{ и } \nu+\mu\neq 0 \\ <\dfrac{\mu\nu}{\mu+\nu},\dfrac{\mu^2}{\mu+\nu}>, \text{при } \nu,\mu\in[0{,}1] \text{ и } \nu\leq\mu \text{ и } \nu+\mu\neq 0 \end{cases}$$

где μ и ν – заданные значения функций принадлежности и непринадлежности соответственно. Заданные значения преобразуются для представления в корректной форме ИНМ.

Далее, применяя приведенное выше преобразование к полученным значениям степени принадлежности и непринадлежности, вычисляются оценки соответствий между парами точек. В конечном итоге, итеративно назначая опорным каждое соответствие, определяются пары взаимосоответствующих точек, сохранивших структуру в наибольшем количестве итераций [26]. Так определяются истинные соответствия, а также параметры изменения масштаба, угла поворота и смещения сцены в кадре. Например, на рисунке 5 есть ложное соответствие и соединенные им точки помечены буквой «F» бирюзового цвета.

Следующим шагом, для осуществления нечеткого вывода, необходимо обработать полученные оценки соответствий для определения их истинности и ложности. Для этого формируется решающее правило в пространстве признаков на основе интуиционистских функций принадлежности. Решающее правило формируется в процессе обучения нейронной сети. Архитектура используемой сети относительно простая. Сеть содержит три слоя: входной, скрытый и выходной [27]. На вход сети поступает трехмерный вектор признаков, далее сигнал поступает на полносвязный скрытый слой. Третий слой сети – один выходной нейрон. Выходное значение сети варьируется в диапазоне [0; 1]. Если значение больше 0,5 – соответствие принадлежит ко множеству истинных, иначе – не принадлежит.

Для обучения использовался набор данных реальных изображений. При этом, разметка (истинные соответствия) была сформирована синтезированием из действительных данных: подобно методу геометрической адаптации. Набор данных состоит из пар изображений, где второе изображений из пары сформировано применением заранее известных преобразований матрицей гомографии к исходному (первому в паре) изображению. Формирование разметки возможно за счет использования матриц гомографии: каждое поступающее на вход модуля сопоставления соответствие проверялось на истинность по следующему правилу:

$$\begin{cases} match = true, \text{когда } Euq(A*H,B)\leq 2, \\ match = false, \text{иначе} \end{cases}$$

где $match$ – оцениваемое соответствие, A – координаты ОТ в исходном изображении, B – координаты ОТ в преобразованном изображении, H – матрица гомографии используемая при преобразовании, $Euq()$ – значение Эвклидова расстояния (дистанции) между ОТ; 2 – порог в пикселях, отличен от нуля, т.к. при на разных изображениях образ ОТ может меняться и положение ее центра может субпиксельно изменяться.

Результаты экспериментов. Для проведения экспериментов был сформирован тестовый набор данных, на основе открытого MegaDepth [27]. MegaDepth – это набор изображений архитектурных достопримечательностей и улиц различных городов Земли. Разметка была сформирована на основе подхода гомографической адаптации: к исходным изображениям были применены случайно сгенерированные матрицы гомографии и дополнительно искажения (дисторсия, изменения яркости) имитирующие аберрации оптической системы. Положения особых точек на новых изображениях были известны, т.к. были известны преобразования порождающие новые изображения [28]. Применение обратного преобразования позволяет оценить истинность сопоставленных точек, в соответствии с выражением (3). Таким образом, было сформировано примерно 62 тыс. тестовых пар изображений.

В эксперименте участвовали две цепочки сопоставления: SuperPoint+NN и SuperPoint+Fuzzy [29]. Обе цепочки используют ОТ вычисленные описанной выше архитектрой SuperPoint. Разница в модулях сопоставления вычисленных особых точек

SuperPoint+NN сопоставляет вычисленные ОТ по правилу выбора ближайшего соседа (nearest neighbor) с использованием подхода ratio check с порогом 4/5. Таким образом истинным соответствие считается, если два ближайших соседа для ОТ различаются больше чем на 20%.

SuperPoint+Fuzzy это предлагаемая в этой работе система. Принцип работы Fuzzy-сопоставителя описан выше.

В качестве оценок для сравнения были выбраны метрики Precision и Recall, вычисляемые по назначенным истинными соответствиям. Результаты эксперимента приведены в табл. 1.

Таблица 1

Результаты экспериментальной проверки.

Метод	Precision	Recall
SuperPoint+NN	24.3	**58.2**
SuperPoint+Fuzzy	**48.2**	52.7

Экспериментальная проверка показала положительный эффект от использования SuperPoint+Fuzzy цепочки: параметр Precision увеличился в два раза, что говорит о значительном снижении ложно выбранных соответствий в качестве истинных. Незначительное снижение показателя Recall обусловлено повышением «требовательности» цепочки к качеству сопоставления и отбрасывания некоторого количества пограничо плохих, по критериям цепочки, соответствий.

Заключение. Представленная в данной работе обучаемая система для сопоставления изображений представляет собой современный и высокоэффективный инструмент, способный решать широкий спектр задач в области компьютерного зрения. Использование методов машинного обучения позволяет системе не только адаптироваться к различным типам и качеству визуальных данных, но и значительно улучшать точность сопоставления за счет выявления сложных закономерностей и особенностей изображений, недоступных традиционным алгоритмам.

В ходе исследования было показано, что такая система успешно справляется с задачами обнаружения признаков, описанием выделенных особенностей и последующим сопоставлением изображений, обеспечивая устойчивость к изменениям масштаба, поворота, освещенности и другим факторам, которые обычно вызывают затруднения. Кроме того, обучаемая модель предоставляет возможности для до обучения на новых данных, что делает её особенно ценным инструментом в условиях постоянно меняющейся среды и требований.

Практическая значимость разработанной системы заключается в её применении в различных сферах – от медицины и промышленного контроля до систем безопасности и анализа мультимедийного контента. Высокая скорость обработки и высокая точность сопоставления обеспечивают возможность её использования в реальном времени, что является критичным для многих приложений.

Таким образом, обучаемая система для сопоставления изображений не только расширяет существующие возможности обработки визуальной информации, но и закладывает основу для дальнейших исследований и внедрений в области интеллектуального анализа данных. Перспективы развития связаны с улучшением архитектуры нейронных сетей, оптимизацией алгоритмов обучения и интеграцией дополнительных источников данных, что позволит повысить как эффективность, так и универсальность таких систем в будущем.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Balntas V., Lenc K., Vedaldi A., Mikolajczyk K.* Hpatches: A Benchmark and Evaluation of Handcrafted and Learned Local Descriptors // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – Los Alamitos: IEEE Computer Society, 2017. – P. 5173-5183.
2. *Chen H., Luo Z., Zhou L. [et al.].* Aspanformer: Detector-free image matching with adaptive span transformer // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – Los Alamitos: IEEE Computer Society, 2022. – P. 5277-5287.
3. *Schonberger J.L., Frahm J.-M.* Structure from-motion revisited // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – Los Alamitos: IEEE Computer Society, 2016. – P. 4104-4113.
4. *Sun J., Shen Z., Wang Y., Bao H., Zhou X.* Loftr: Detector-free local feature matching with transformers // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – Los Alamitos: IEEE Computer Society, 2021. – P. 8918-8927.
5. *Truong P., Danelljan M., Van Gool L., Timofte R.* Learning accurate dense correspondences and when to trust them // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – Los Alamitos: IEEE Computer Society, 2021. – P. 5710-5720.
6. *DeTone D., Malisiewicz T., Rabinovich A. Superpoint:* Self-supervised interest point detection and description // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – Los Alamitos: IEEE Computer Society, 2018. – P. 224-236.
7. *Karpur A., Perrotta G., Martin-Brualla R. [et al.].* Lfm-3d: Learnable feature matching across wide baselines using 3d signals // Proceedings of the International Conference on 3D Vision (3DV). – Los Alamitos: IEEE, 2024. – P. 1234-1245.
8. *Зуев В.М.* Сравнение обнаружения объектов средствами искусственного интеллекта в сравнении с классическими методами // Проблемы искусственного интеллекта. – 2024. – № 3 (34). – С. 30-35.
9. *Dai A., Chang A.X., Savva M. [et al.].* Scannet: Richly-annotated 3D reconstructions of indoor scenes // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – Los Alamitos: IEEE Computer Society, 2017. – P. 5828-5837.
10. *Silveira T.L., Jung C.R.* Dense 3D scene reconstruction from multiple spherical images for 3-DOF+ VR applications // IEEE Conference on Virtual Reality and 3D User Interfaces (VR). – Los Alamitos: IEEE, 2019. – P. 9-18.
11. *Tang L., Jia M., Wang Q. [et al.].* Emergent correspondence from image diffusion // arXiv. 2023. arXiv:2306.03881.
12. *Torii A., Havlena M., Pajdla T.* From Google Street View to 3D city models // Proceedings of the International Conference on Computer Vision. – Los Alamitos: IEEE Computer Society, 2009. – P. 2188-2195.
13. *Tian Y., Fan B., Wu F.* L2-Net: Deep learning of discriminative patch descriptor in Euclidean space // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – Los Alamitos: IEEE Computer Society, 2017. – P. 661-669.
14. *Носков В.П., Курьянов А.Н.* Использование комплексированных дескрипторов в решении SLAM-задачи // Известия ЮФУ. Технические науки. – 2022. – № 1 (225). – С. 268-278.
15. *Морев К.И., Боженюк А.В.* Сопоставление изображений по особым точкам различных категорий // Известия ЮФУ. Технические науки. – 2020. – № 3 (213). – С. 192-201.
16. *Truong P., Danelljan M., Van Gool L., Timofte R.* Learning accurate dense correspondences and when to trust them // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – Los Alamitos: IEEE Computer Society, 2021. – P. 5710-5720.
17. *Tyszkiewicz M., Maninis K.-K., Popov S., Ferrari V.* RayTran: 3D pose estimation and shape reconstruction of multiple objects from videos with ray-traced transformers // Proceedings of the European Conference on Computer Vision. – Cham: Springer, 2022. – P. 456-472.
18. *Бондаренко В.А., Каплинский Г.Э., Павлова В.А., Тупиков В.А.* Метод поиска и сопоставления ключевых особенностей изображений для распознавания образов и сопровождения объектов // Известия ЮФУ. Технические науки. – 2019. – № 1 (203). – С. 281-293.

19. *Vaswani A., Shazeer N.M., Parmar N. [et al.].* Attention is all you need // Advances in Neural Information Processing Systems. – Red Hook: Curran Associates, 2017. – P. 5998-6008.
20. *Verdie Y., Yi K.M., Fua P., Lepetit V.* TILDE: A temporally invariant learned detector // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – Los Alamitos: IEEE Computer Society, 2015. – P. 5263-5271.
21. *Ma J., Jiang X., Fan A. [et al.].* Image matching from handcrafted to deep features: A survey // International Journal of Computer Vision. – 2020. – Vol. 129, No. 1. – P. 23-79.
22. *Ono Y., Trulls E., Fua P.V., Yi K.M.* LF-Net: Learning local features from images // Advances in Neural Information Processing Systems. – Red Hook: Curran Associates, 2018. – P. 6234-6244.
23. *Radenovic F., Iscen A., Tolias G., Avrithis Y., Chum O.* Revisiting Oxford and Paris: Large-scale image retrieval benchmarking // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – Los Alamitos: IEEE Computer Society, 2018. – P. 5706-5715.
24. *Deng J., Dong W., Socher R. [et al.].* ImageNet: A large-scale hierarchical image database // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – Los Alamitos: IEEE Computer Society, 2009. – P. 248-255.
25. *Atanassov K.* Remark on a property of the intuitionistic fuzzy interpretation triangle // Notes on Intuitionistic Fuzzy Sets. – 2002. – No. 8 (1). – P. 34-36.
26. *Roessle B., Nießner M.* End2end multi-view feature matching with differentiable pose optimization // Proceedings of the International Conference on Computer Vision. – Paris: IEEE, 2023. – P. 477-487.
27. *Li Z., Snavely N.* MegaDepth: Learning single-view depth prediction from internet photos // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. – Los Alamitos: IEEE, 2018. – P. 2041-2050.
28. *Устенко В.Ю., Бондаренко В.И.* Разработка программного комплекса аннотирования данных для задач компьютерного зрения: объектно-ориентированный подход на основе winforms // Проблемы искусственного интеллекта. – 2024. – № 4 (35). – С. 151-163.
29. *Xue F., Budvytis I., Cipolla R.* SFD2: Semantic-guided feature detection and description // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – Los Alamitos: IEEE Computer Society, 2023. – P. 5206-5216.

REFERENCES

1. *Balntas V., Lenc K., Vedaldi A., Mikolajczyk K.* Hpatches: A Benchmark and Evaluation of Handcrafted and Learned Local Descriptors, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society, 2017, pp. 5173-5183.
2. *Chen H., Luo Z., Zhou L. [et al.].* Aspanformer: Detector-free image matching with adaptive span transformer, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society, 2022, pp. 5277-5287.
3. *Schonberger J.L., Frahm J.-M.* Structure from-motion revisited, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* Los Alamitos: IEEE Computer Society, 2016, pp. 4104-4113.
4. *Sun J., Shen Z., Wang Y., Bao H., Zhou X.* Loftr: Detector-free local feature matching with transformers, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society, 2021, pp. 8918-8927.
5. *Truong P., Danelljan M., Van Gool L., Timofte R.* Learning accurate dense correspondences and when to trust them, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society, 2021, pp. 5710-5720.
6. *DeTone D., Malisiewicz T., Rabinovich A. Superpoint:* Self-supervised interest point detection and description, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society, 2018, pp. 224-236.
7. *Karpur A., Perrotta G., Martin-Brualla R. [et al.].* Lfm-3d: Learnable feature matching across wide baselines using 3d signals, *Proceedings of the International Conference on 3D Vision (3DV)*. Los Alamitos: IEEE, 2024, pp. 1234-1245.
8. *Zuev V.M.* Sravnenie obnaruzheniya ob"ektov sredstvami iskusstvennogo intellekta v sravnenii s klassicheskimi metodami [Comparison of object detection using artificial intelligence methods versus classical methods], *Problemy iskusstvennogo intellekta* [Problems of Artificial Intelligence], 2024, No 3 (34), pp. 30-35.
9. *Dai A., Chang A.X., Savva M. [et al.].* Scannet: Richly-annotated 3D reconstructions of indoor scenes, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society, 2017, pp. 5828-5837.
10. *Silveira T.L., Jung C.R.* Dense 3D scene reconstruction from multiple spherical images for 3-DOF+ VR applications, *IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. Los Alamitos: IEEE, 2019, pp. 9-18.

11. *Tang L., Jia M., Wang Q. [et al.]*. Emergent correspondence from image diffusion, *arXiv*. 2023. arXiv:2306.03881.
12. *Torii A., Havlena M., Pajdla T.* From Google Street View to 3D city models, *Proceedings of the International Conference on Computer Vision*. Los Alamitos: IEEE Computer Society, 2009, pp. 2188-2195.
13. *Tian Y., Fan B., Wu F.* L2-Net: Deep learning of discriminative patch descriptor in Euclidean space, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society, 2017, pp. 661-669.
14. *Noskov V.P., Kur'yanov A.N.* Ispol'zovanie kompleksirovannykh deskriptorov v reshenii SLAM-zadachi [Using complex descriptors for SLAM task solution], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2022, No. 1 (225), pp. 268-278.
15. *Morev K.I., Bozhenyuk A.V.* Sopostavlenie izobrazheniy po osobym tochkam razlichnykh kategoriy [Image matching by feature points of various categories], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2020, No. 3 (213), pp. 192-201.
16. *Truong P., Danelljan M., Van Gool L., Timofte R.* Learning accurate dense correspondences and when to trust them, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society, 2021, pp. 5710-5720.
17. *Tyszkiewicz M., Maninis K.-K., Popov S., Ferrari V.* RayTran: 3D pose estimation and shape reconstruction of multiple objects from videos with ray-traced transformers, *Proceedings of the European Conference on Computer Vision*. Cham: Springer, 2022, pp. 456-472.
18. *Bondarenko V.A., Kaplinskiy G.E., Pavlova V.A., Tupikov V.A.* Metod poiska i sopostavleniya klyuchevykh osobennostey izobrazheniy dlya raspoznavaniya obrazov i soprovozhdeniya ob"ektov [Method of search and matching of key image features for pattern recognition and object tracking], *Izvestiya YuFU. Tekhnicheskie nauki* [Izvestiya SFedU. Engineering Sciences], 2019, No. 1 (203), pp. 281-293.
19. *Vaswani A., Shazeer N.M., Parmar N. [et al.]*. Attention is all you need, *Advances in Neural Information Processing Systems*. Red Hook: Curran Associates, 2017, pp. 5998-6008.
20. *Verdie Y., Yi K.M., Fua P., Lepetit V.* TILDE: A temporally invariant learned detector, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society, 2015, pp. 5263-5271.
21. *Ma J., Jiang X., Fan A. [et al.]*. Image matching from handcrafted to deep features: A survey, *International Journal of Computer Vision*, 2020, Vol. 129, No. 1, pp. 23-79.
22. *Ono Y., Trulls E., Fua P.V., Yi K.M.* LF-Net: Learning local features from images, *Advances in Neural Information Processing Systems*. Red Hook: Curran Associates, 2018, pp. 6234-6244.
23. *Radenovic F., Iscen A., Tolias G., Avrithis Y., Chum O.* Revisiting Oxford and Paris: Large-scale image retrieval benchmarking, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society, 2018, pp. 5706-5715.
24. *Deng J., Dong W., Socher R. [et al.]*. ImageNet: A large-scale hierarchical image database, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society, 2009, pp. 248-255.
25. *Atanassov K.* Remark on a property of the intuitionistic fuzzy interpretation triangle, *Notes on Intuitionistic Fuzzy Sets*, 2002, No. 8 (1), pp. 34-36.
26. *Roessle B., Nießner M.* End2end multi-view feature matching with differentiable pose optimization, *Proceedings of the International Conference on Computer Vision*. Paris: IEEE, 2023, pp. 477-487.
27. *Li Z., Snavely N.* MegaDepth: Learning single-view depth prediction from internet photos, *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE, 2018, pp. 2041-2050.
28. *Ustenko V.Yu., Bondarenko V.I.* Razrabotka programmnogo kompleksa annotirovaniya dannykh dlya zadach komp'yuternogo zreniya: ob"ektno-orientirovannyy podkhod na osnove winforms [Development of data annotation software complex for computer vision tasks: object-oriented approach based on WinForms], *Problemy iskusstvennogo intellekta* [Problems of Artificial Intelligence], 2024, No. 4 (35), pp. 151-163.
29. *Xue F., Budvytis I., Cipolla R.* SFD2: Semantic-guided feature detection and description, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society, 2023, pp. 5206-5216.

Морев Кирилл Иванович – Акционерное общество Научно-конструкторское бюро вычислительных систем; e-mail: morev-ki@ya.ru; г. Таганрог, Россия; программист.

Morev Kirill Ivanovich – Joint Stock Company "Scientific Design Bureau of Computing Systems"; e-mail: morev-ki@ya.ru; Taganrog, Russia; programmer.