

АНАЛИЗ МЕТОДОВ ОПРЕДЕЛЕНИЯ РЕАЛЬНЫХ ГЕОМЕТРИЧЕСКИХ РАЗМЕРОВ ОБЪЕКТОВ И РАССТОЯНИЙ С ПРИМЕНЕНИЕМ КОМПЬЮТЕРНОГО ЗРЕНИЯ

Д. А. ГОРДЕВИЧ, д-р техн. наук, доц. Р. П. БОГУШ
*(Полоцкий государственный университет
имени Евфросинии Полоцкой, Беларусь)*

Аннотация. Рассматриваются основные существующие подходы, которые могут быть использованы для определения реальных размеров объектов и расстояний по видеоданным. Представлены классические методы на основе мерного объекта, стереозрения, RGB-D, а также с использованием специально обученных для данных задач нейронных сетей. Приведены особенности их применения и недостатки.

Ключевые слова: компьютерное зрение, измерение размеров объектов на изображении, нейронные сети.

Системы компьютерного зрения все шире используются для решения различных прикладных проблем. При этом ключевым шагом в такой системе является обнаружение объекта и определение его параметров, в том числе во многих случаях требуется определить реальные геометрические размеры объекта на основе видеоданных. Можно выделить следующие основные подходы для такой задачи: применение мерных объектов [1], с использованием стереозрения [2], при помощи камеры с датчиком глубины [3], с применением искусственных нейронных сетей [4].

Реальные геометрические размеры объекта интереса по его изображению методом мерных объектов могут быть получены на основе соотношения известных реальных размеров мерных объектов с их размерами на полученном изображении. Такой подход реализован в библиотеке OpenCV [5]. Однако погрешность в данном примере достигает 3 мм, что для небольших объектов является существенной величиной. Низкая точность обусловлена отсутствием калибровки камеры. Без калибровки камеры фотографии могут быть подвержены искажениям из-за влияния характеристик оптической системы. Методы определения размеров объектов при помощи мерных объектов является неудобным в эксплуатации и ограничивает функциональные возможности.

Подход получения информации о дальности и размере объекта, не использующих мерный объект, состоит в использовании стереозрения, предполагающий регистрацию нескольких изображений объектов под различными ракурсами. Для каждой точки на одном изображении выполняется поиск парной ей точки

на другом изображении. По паре соответствующих точек можно выполнить триангуляцию и определить координаты их прообраза в трехмерном пространстве. На основе трехмерных координат прообраза вычисляется глубина как расстояние до плоскости камеры. В результате получается карта глубины, изображение которой в каждом пикселе содержит информацию о расстоянии поверхностей объекта до камеры.

Методика повышения точности измерения расстояний и размеров объекта, представленная в [6], основана на принципах фотограмметрии и корреляционной обработке цифровых изображений стереопары. Получение информации о дальности заключается в регистрации нескольких изображений объектов под различными ракурсами. При сопоставлении этих изображений можно реконструировать трехмерную структуру сцены. На основе координат исследуемых объектов могут быть определены линейные размеры исследуемых объектов. Для нахождения разности координат объекта (смещения) в плоскостях первого и второго изображений стереопары используется соответствующая методика корреляционной обработки изображений.

Сдвиг между изображениями определяется по положению максимального значения функции схожести, как правило двумерной нормированной корреляционной функции. Однако могут быть использованы другие функции схожести [7].

Точность такой методики такова, что, для камеры Canon PowerShot A550 с матрицей 7,1 мегапикселя и фокусом объектива 5,8 мм, на расстояниях до 50 м достигается погрешность измерения лучше 1 см. На расстояниях порядка 1 км относительная погрешность измерения составляет 0,6 % для базы 50 см и 0,3 % – для базы 1 м.

Метод с использованием стереозрения, представленный в [8], реализован с использованием языка программирования C++ и библиотеки OpenCV. Он позволяет проводить измерения объекта в режиме реального времени. Результаты показывают, что измерение расстояния до объекта определяются с погрешностью от 10 до 25 см, а измерение размера объекта с погрешностью не более 3 см. Такая точность связана с уменьшенным разрешением изображения, чтобы процесс мог быть выполнен в режиме реального времени. Более высокое разрешение камеры позволило бы производить более точные измерения, но тогда процесс не сможет быть выполнен в режиме реального времени, так как частота кадров стала бы слишком низкой. Так же частота кадров снижается, если в поле зрения камер будет попадать больше одного объекта, а из-за использования вычитания фона для обнаружения объекта алгоритм требует постоянной среды освещения, что может вызвать трудности в удобстве использования.

Ещё одним подходом для определения размеров объектов на изображении является использование камеры с датчиком глубины. С RGB-Depth (RGB-D) сенсоров можно получить цветное изображение (RGB изображение) и соответствующую ему

карту глубины. Применяется алгоритм сегментации изображения, чтобы выделить объекты на изображении. Для каждого выделенного объекта вычисляется его геометрические характеристики и используя информацию о глубине изображения, определяются размеры объектов в трехмерном пространстве.

В исследовании с использованием данного подхода [6] для определения высоты объектов координаты каждого пикселя изображения пересчитали в координаты реального измерения. Для сегментации изображения использовали два метода: метод кластеризации на основе глубины и метод обнаружения объектов по RGB изображению при помощи сверточной нейронной сети совместно с кластеризацией на основе глубины. Первый метод позволяет получить четкие границы объекта. Пиксели изображения глубины группируют в три кластера: фон, передний план и пиксель без информации о глубине. Кластеризация на основе глубины позволяет сегментировать объекты в тех случаях, когда сегментация на основе RGB затруднена. Погрешность такого метода составила от 3 до 6 сантиметров. Но этот метод применим только в случаях, когда на изображении присутствует только один объект. Использование сегментации только на основе глубины недостаточно, когда в сцене присутствует несколько объектов, поэтому второй метод, использует готовые предварительно обученные сверточные нейронные сети для обнаружения и получения ограничивающих рамок вокруг интересующих объектов. При этом используется только информация RGB для получения ограничивающей рамки. После чего кластеризация на основе глубины позволяет получить четкие границы объекта уже внутри ограничивающей рамки. Погрешность такого метода составила от 6 до 10 сантиметров. В ходе исследования было обнаружено, что информация об облаке точек может отсутствовать в случае черных или металлических объектов. Кроме того, существует также пространственный сдвиг между цветной камерой и датчиком глубины, который необходимо учитывать.

В статье [10] представлена методика, использующая Microsoft Kinect для получения RGB-D изображения. Кадры для измерения размеров книги, монитора и шкафа были выбраны на основе изображений на которых целевой объект был четко виден с минимальным размытием изображения и дрожанием камеры, но расположен в случайных местах, некоторые из которых были ближе к центру, в то время как другие были ближе к краям. Методика, описанная в данной статье, вычисляет размеры с точностью 90–95%. Такая точность обусловлена неточностями вычисления глубины самой камеры и неспособности обнаружения глубины в местах, где имеются поглощающие, глянцевые и отражающие поверхности, а также острые края объектов.

Метод представленный в [11] использует сверточную нейронную сеть (СНС) YOLOv3 для идентификации людей и оценки расстояния с помощью монокулярной камеры. В качестве набора данных для обучения нейронной сети были использованы изображения людей в разных позах, сегментированные изображения и карты

глубины. Для обучения использованы наборы данных с максимальным расстоянием до человека 30 и 50 метров. В результате тестирования модель, обученная на наборе данных до 30 метров, имеет погрешность от 0,89 до 3,21 метра. Модель, обученная на наборе данных до 50 метров, характеризуется погрешностью от 1,4 до 3 метров. В сравнении с данным методом, другие методы такие как DisNet [12] и LASSO [13] показывают погрешность большую на 50–60% соответственно. Однако, недостатком данного метода является то, что нейронная сеть YOLOv3 была обучена на людях с одинаковыми характеристиками, такими как возраст, телосложение и одежда, что может повлиять на результат использования на других объектах интереса. Для решения данной проблемы требуется обучение на значительно больше базе данных объектов реального мира.

В методе, представленном в [14], для обнаружения объектов используется YOLOv5, а для оценки глубины – DepthNet. Таким образом две модели обучали двумя наборами данных: состоящим из синхронизированных RGB изображений объектов и их карты глубин и набор данных взятый из COCO, содержащий изображения объектов 80 различных классов. Для оценки результатов авторы самостоятельно подготовили набор данных, состоящий из 100 изображений. При этом расстояние от камеры до объекта измерялось оператором. Погрешность в тестах составила от 0.09 до 0.69, причем, с увеличением расстояния погрешность также увеличивалась.

ЛИТЕРАТУРА

1. Идентификация размеров объекта с использованием эталонного объекта. [Электронный ресурс]. – Режим доступа: <https://www.вестник-науки.пф/article/3993>. – Дата доступа: 18.02.2024.
2. Основы стереозрения. [Электронный ресурс]. – Режим доступа: <https://habr.com/ru/articles/130300/>. – Дата доступа: 18.02.2024
3. Камеры глубины. Опыт сканирования помещений. [Электронный ресурс]. – Режим доступа: <https://medium.com/phygitalism/rgb-d-sensors-room-scan-c0c70fbc1946>. – Дата доступа: 18.02.2024.
4. Искусственные нейронные сети и их приложения. [Электронный ресурс]. – Режим доступа: https://kpfu.ru/staff_files/F1493580427/NejronGafGal.pdf. – Дата доступа: 18.02.2024
5. Измерение размера объектов на изображении с помощью OpenCV [Электронный ресурс]. – Режим доступа: <https://pyimagesearch.com/2016/03/28/measuring-size-of-objects-in-an-image-with-opencv/>. – Дата доступа: 19.02.2024.
6. Козлов В.Л., Кузьмичев И.Р. Вестник БГУ. Сер. 1. – 2011. – № 1. – С. 33–38.
7. Bohush, R. Image Similarity Estimation Based on Ratio and Distance Calculation between Features / R.P. Bohush, S.V. Ablameyko, E.R. Adamovskiy and D. Savca // Pattern Recognition and Image Analysis. – 2020. – Vol. 30, № 2. – P. 147–159.
8. Измерение расстояния и размера объекта с помощью системы стереовидения. [Электронный ресурс]. – Режим доступа: https://www.researchgate.net/publication/274773924_Object_Distance_and_Size_Measurement_Using_Stereo_Vision_System. – Дата доступа: 01.03.2024.
9. Object Localization and Size Estimation from RGB-D Images. [Электронный ресурс]. – Режим доступа: <https://arxiv.org/pdf/1808.00641.pdf>. – Дата доступа: 03.03.2024.

10. Calculating real world object dimensions from Kinect RGB-D image using dynamic resolution. [Электронный ресурс]. – Режим доступа: https://www.researchgate.net/publication/282954352_Calculating_real_world_object_dimensions_from_Kinect_RGB-D_image_using_dynamic_resolution. – Дата доступа: 05.03.2024.
11. Human detection and distance estimation with monocular camera using YOLOv3 neural network. [Электронный ресурс]. – Режим доступа: <https://ims.ut.ee/www-public2/at/2019/msc/atprog-courses-magistrit55-loti.05.036-asif-sattar-text-20190603.pdf>. – Дата доступа: 13.03.2024.
12. DisNet: A novel method for distance estimation from monocular camera. [Электронный ресурс]. – Режим доступа: <https://project.inria.fr/ppniv18/files/2018/10/paper22.pdf>. – Дата доступа: 13.03.2024.
13. Real-time hand gesture feature extraction using depth data. [Электронный ресурс]. – Режим доступа: <https://core.ac.uk/download/pdf/96919675.pdf>. – Дата доступа: 13.03.2024.
14. Absolute distance prediction based on deep learning object detection and monocular depth estimation models. [Электронный ресурс]. – Режим доступа: https://www.researchgate.net/publication/355872639_Absolute_distance_prediction_based_on_deep_learning_object_detection_and_monocular_depth_estimation_models. – Дата доступа: 14.03.2024.