

Оценивание координат объектов с использованием сверточных сетей и карт глубины

Н.А. Андриянов
Финансовый университет при Правительстве Российской Федерации
Москва, Россия
naandriyanov@fa.ru

Аннотация—В работе рассмотрена задача оценивания координат объектов, обнаруживаемых по видеопоследовательности изображений с камеры Intel Real Sense. За счет использования сверточной нейронной сети YOLOv5 разработана система обнаружения объектов, для которой реализован алгоритм оценивания их координат в трехмерном пространстве на базе совместной обработки информации с карт глубины и оценок пиксельных координат объектов на изображении. Проведенный сравнительный анализ с методом инструментальных измерений показывал низкую погрешность предложенного алгоритма.

Ключевые слова— компьютерное зрение, искусственный интеллект, стереозрение, обнаружение объектов, оценивание координат.

1. ВВЕДЕНИЕ

Сегодня системы искусственного интеллекта все глубже и шире проникают в различные отрасли экономики. Особый интерес в ряде прикладных задач представляют системы компьютерного зрения, эффективность работы которых значительно повысилась с появлением глубоких сверточных нейронных сетей [1]. Сегодня существует множество популярных архитектур, которые используются в задачах обнаружения и распознавания объектов на изображениях, например, сети архитектуры YOLO [2]. Однако в ряде задач оценки положения объекта в плоскости оптического изображения недостаточно. Например, в робототехнике системы компьютерного зрения должны оценивать положение объектов в трехмерном пространстве [3]. Например, при автоматизации сбора урожая [4], необходимо точно знать, в какое положение направить манипулятор для захвата.

В данной работе рассматривается комплексный алгоритм, который позволяет оценить местоположение в трехмерном пространстве обнаруживаемого на видео объекта по результатам оценки его координат на кадре изображения с помощью сети YOLOv5 и построенной с помощью дальномера камеры Intel Real Sense карты глубины.

2. ДАННЫЕ С КАМЕРЫ И ИХ ОБРАБОТКА

Задача определения дальности до объектов сегодня не составляет большого труда и во многом является решенной. В используемой в работе камере Intel Real Sense Depth Camera D415 (рис. 1) применяется алгоритм измерения стереоглубины. Расстояние рассчитывается по различиям двух принимаемых изображений инфракрасных камер, направленных в одну сторону.



Рис. 1. Вид камеры Intel Real Sense Depth Camera D415

На рис. 1 слева направо: ИК-камера принимающая, ИК-излучатель, ИК-камера принимающая, камера оптического диапазона (непосредственно видеочкамера).

Преимуществом использования такой стереокамеры является то, что карта глубины и цветовое изображение не требуют дополнительного совмещения. Таким образом, мы получаем карту глубины, в которой для каждого пикселя на изображении рассчитано расстояние от камеры и цветовую картинку в каналах яркости RGB. Однако если обратить внимание на рис. 1, то можно заметить, что для корректного позиционирования относительно центра камеры следует учитывать смещение видеочкамеры от центра вдоль горизонтальной оси.

На рис. 2 показан интерфейс программы Real Sense Viewer при построении цветового изображения, а на рис. 3 – соответствующая цветному изображению карта глубины.

Следует отметить, что отдельно взятые кадры, поступающие с камеры, можно рассматривать как статичные изображения. Для этого будет использоваться библиотек OpenCV на Python. Также подключим сверточную нейронную сеть YOLOv5 с уже предобученными весами. В таком формате данная нейронная сеть позволяет производить обнаружение объектов, принадлежащих к стандартным классам датасета COCO [5].

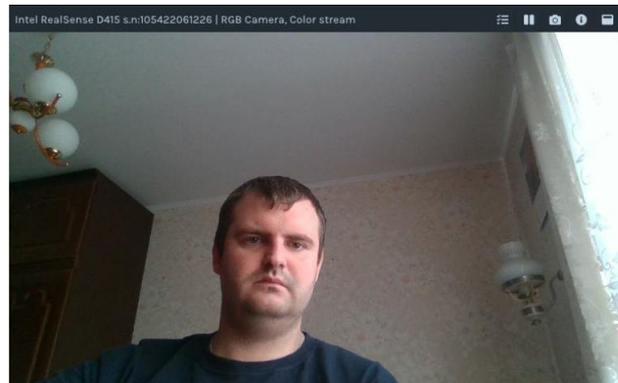


Рис. 2. Изображение в канале RGB

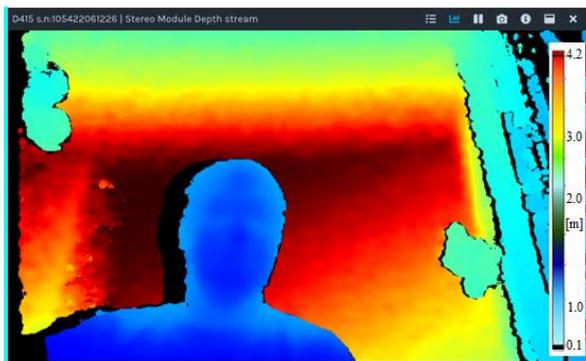


Рис. 3. Карта глубины в Intel RealSense Viewer

На рис. 4 представлен пример совместной обработки информации с цветного и глубинного изображений с использованием YOLOv5. По данным с RGB камеры определяется объект, который обводится прямоугольником, а расстояние до объекта определяется, как значение карты глубины в его центре.

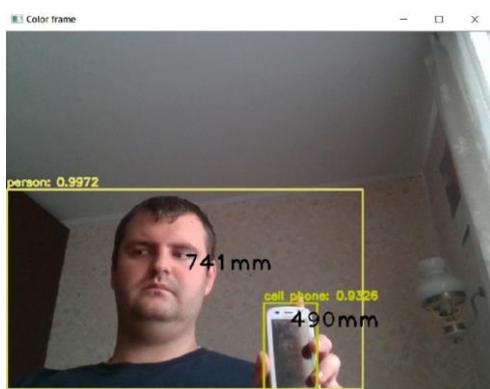


Рис. 4. Определение расстояния до объекта

Анализ рис. 4 показывает, что система компьютерного зрения обнаружила 2 объекта: человека и мобильный телефон на расстоянии 741 мм и 490 мм соответственно. При этом расстояние определяется как расстояние до центральной точки ограничивающего прямоугольника. Зная координаты ограничивающих прямоугольников в пикселях (принимая левую верхнюю точку за начало координат), получим выражение для пересчета относительных координат объектов [6]:

$$\begin{aligned} X[mm] &= d_{x_0,y_0} \frac{(C_x - x_0[px])}{f_x}, \\ Y[mm] &= d_{x_0,y_0} \frac{(C_y - y_0[px])}{f_y}, \end{aligned} \quad (1)$$

где сдвиги относительно камеры по осям X и Y рассчитаны в миллиметрах, d_{x_0,y_0} – значение карты глубины в центре ограничивающего объект прямоугольника, C_x и C_y – координаты центра изображения в пикселях, x_0 и y_0 – координаты центра ограничивающего прямоугольника в пикселях.

Для камеры Intel Real Sense Depth Camera D415 смещение видеофиксатора относительно центра составляет 35 мм, поэтому сделаем поправку:

$$X'[mm] = X[mm] - 35. \quad (2)$$

Понятно, что подставляя $Y[mm]$ из выражения (1), а $X[mm]$ из выражения (2), легко найти $Z[mm]$ по формуле евклидова расстояния, зная значение d_{x_0,y_0} .

Проверка алгоритма выполнялась для 5 произвольных объектов. Сравнение осуществлялось с измерениями с помощью угольников и линеек. В таблице 1 представлены результаты исследования.

Таблица 1. ОЦЕНКА ПОГРЕШНОСТИ ИЗМЕРЕНИЙ

№ объекта	Измерения		
	Пара рассчитанных координат (x,y)	Пара измеренных координат (x,y)	Корень из среднеквадратичной ошибки
1	(186, 25)	(190, 24)	(7.33, 5.27)
2	(320, 85)	(308, 82)	
3	(-76, -64)	(-82, -57)	
4	(-163, 96)	(-155, 100)	
5	(95, 252)	(92, 244)	

По результатам таблицы 1 можно сделать вывод о том, что погрешности в среднем не превышают 1 см.

3. ЗАКЛЮЧЕНИЕ

В работе предложено выполнять оценку координат системами компьютерного зрения с помощью комплексирования карт глубины и оценок на базе глубокого обучения. Исследования предложенного алгоритма показали, что оценки измерений имеют низкую погрешность: в среднем 5-7 мм.

БЛАГОДАРНОСТИ

Исследование выполнено при поддержке РФФИ, Грант №19-29-09048 мк.

ЛИТЕРАТУРА

- [1] Krizhevsky, A. ImageNet Classification with Deep Convolutional Neural Networks / A. Krizhevsky, I. Sutskever, G.E. Hinton // NIPS'12 Proc 25th Int Conf Neural Inf Process Syst. – 2012. – Vol. 1. – P. 1097-1105.
- [2] Nepal, U. Comparing YOLOv3, YOLOv4 and YOLOv5 for Autonomous Landing Spot Detection in Faulty UAVs / U. Nepal, H. Eslamiat // Sensors. – 2022. – Vol. 22. – P. 464. DOI: 10.3390/s22020464.
- [3] Švec, M. Calculation of object position in various reference frames with a robotic simulator / M. Švec, I. Farkas // Proceedings of 36th Annual Conference of the Cognitive Science Society. – 2014. – P. 1-4.
- [4] Andriyanov, N. Intelligent System for Estimation of the Spatial Position of Apples Based on YOLOv3 and Real Sense Depth Camera D415 // N. Andriyanov, I. Khasanshin, D. Utkin, T. Gataullin, S. Ignar, V. Shumaev, V. Soloviev // Symmetry. – 2022. – Vol. 14. – P. 148. DOI: 10.3390/sym14010148.
- [5] COCO [Electronic resource]. – Access mode: <https://cocodataset.org/> (08.02.2022).
- [6] Laganier, R. Robust object pose estimation from feature-based stereo / R. Laganier, S. Gilbert, G. Roth // IEEE Trans. Instrum. Meas. – 2006. – Vol. 55. – P. 1270-1280.