

Определение геометрии помещения по импульсной характеристике с применением свёрточных нейронных сетей

Д. А. Полещенко, К. Э. Мельникова, Д.Е. Назаров

*Старооскольский технологический институт им. А.А. Угарова (филиал)
Национального исследовательского технологического университета «МИСИС», Россия,
г. Старый Оскол*

Аннотация: В данной статье исследовался способ определения геометрии помещений с помощью эхолокации, а именно - звуколокации. В работе использовались свёрточные нейронные сети, которые обучались на большом наборе данных, состоящем из 48000 импульсных характеристик и ряда параметров, для геометрии помещения, им соответствующих. Обученная свёрточная нейронная сеть тестировалась на записанной импульсной характеристике реального помещения и показала точность в оценке по различным параметрам размера помещения от 92,2% до 98,7%.

Ключевые слова: свёрточные нейронные сети, геометрия помещения, эхолокация, импульсная характеристика, робототехника, распознавание, бесконтактные методы измерения объектов, звуколокация, предсказание геометрии, виртуальная реальность.

Одним из самых перспективных направлений в науке на сегодняшний день является оптимизация существующих или открытие новых методов оцифровки окружающей нас среды для накопления большого объёма данных с последующим применением этой информации для ускорения научно-технического прогресса.

Научить какой-либо объект определять геометрию помещения, а после идентифицировать себя в нём, это задача, которую ставят перед собой многие разработчики роботизированных устройств [1].

Однако использование фото- и видеоаппаратуры накладывает на выполнение требуемой от устройства задачи много условий, таких, как:

- наличие освещения в определённом диапазоне;
 - дополнительная защита объектива от механических воздействий;
 - энергоёмкий источник питания;
 - большие вычислительные мощности для обработки фото- и видеоинформации в реальном времени.
-

Также определение границ помещения относительно какого-либо объекта - это задача, которую решают специалисты в области дополненной реальности. Здесь отчасти присутствуют те же проблемы, что и в сфере робототехники.

Чен Минг и Джеймс А. Симмонс использовали рекуррентные и свёрточные нейронные сети для определения количества основных отражающих поверхностей и расстояния между ними, составляющих структуру объекта, для определения его геометрии. Их метод основан на измерении и анализе сигнала эхолокации летучих мышей с использованием специально разработанного высокочастотного (20-80 кГц) микрофона звукового поля и пространственной аудиотехники [2].

Сакари Терво и Тимо Тоссавайнен в своем исследовании проводили оценку геометрии помещения по пространственным импульсным откликам помещения. Их алгоритм не требовал априорной информации о форме помещения, количестве стен и порядке отражений, он итеративно выводил набор плоскостей, объясняющих измеренные положения и ковариации источников и изображений [3].

Существующие методы определения геометрии замкнутого пространства с помощью эхолокации подразумевают наличие в помещении большого количества дополнительного оборудования (источников и приёмников звука). Описанный нами метод не подразумевает наличие каких-либо априорных знаний об окружающем нас пространстве. В работе для создания и съёма реальных импульсных характеристик использовался один источник и один приемник звука [4].

Для предсказания геометрии помещения использовалась свёрточная нейронная сеть, которая обучалась на искусственно созданных импульсных характеристиках.

Генерация импульсных характеристик производилась в пакете программ MatLab. За основу был взят метод, описанный в работе [5].

В функцию, которая будет генерировать многоканальные импульсные отклики помещения, передавались следующие параметры:

- скорость звука при 20 °С (340 м/с);
- частота дискретизации (16 000 Гц);
- время реверберации (0.4 с);
- тип микрофона по диаграмме направленности;
- порядок отражения;
- ориентация микрофона в пространстве;
- значение True/False для фильтра высоких частот.

В процессе работы алгоритма по имитации импульсных характеристик было сгенерировано 3000 помещений со случайными размерами в следующих пределах: длина от 6 до 10 метров, ширина от 5 до 8 метров, высота от 4 до 6 метров. В каждой комнате было промоделировано 16 случайных относительно друг друга и в пространстве положений для источника и приёмника звукового сигнала. Полученные 48 000 импульсных характеристик, длительностью 0,256 секунды, сохранились в цифровом аудиоформате WAV [6].

В качестве примера приведена характеристика, которая была получена для комнаты со следующими параметрами: длина – 6.0196 м., ширина – 7.2332 м., высота – 5.0205 м. На рис. 1 приведен пример полученной импульсной характеристики.

Для неё самая большая амплитуда у первой регистрации возмущения приёмником звука – это сигнал, который пришёл на прямую от источника звука к приёмнику и был самым громким. Далее наблюдается несколько всплесков примерно одинаковой амплитуды – это первично отражённый звук от стен, пола и потолка. Далее вплоть до конца измерения мы наблюдаем

процесс реверберации – постепенного затухания звука после отключения источника [7].

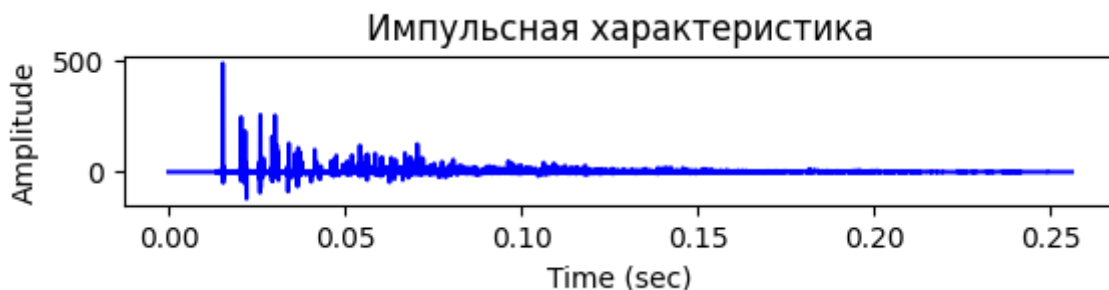


Рис. 1. – Импульсная характеристика

Полученные импульсные характеристики подавались на свёрточную нейронную сеть со следующей структурой:

Входной слой для модели с ожидаемой формой данных. Затем следует 6 слоев свертки - Conv1D с последующей нормализацией (BatchNormalization) и функцией активацией 'relu'. Эти слои применяются последовательно друг за другом. Количество выходных фильтров в слоях Conv1D: 10,20,40,80,160,160, ядро свёртки и шаг свёртки всегда имеют значение - 4. Далее идут “сглаживающий” слой Flatten, Dropout слой с 50% исключением нейронов и последний скрытый слой Dense с 40 нейронами. На выходе стоит слой Dense с функцией активации 'sigmoid' [8,9].

Сеть обучалась со следующими параметрами обучения: batch_size=150, epochs=30, обучающая и тестовая выборка составили 38400 и 9600 импульсных характеристик, соответственно.

В ходе исследования было обучено 5 нейронных сетей одинаковой структуры, которые после обучения обеспечили следующую точность на тестовой выборке:

1. 0.00235 м
2. 0.02273 м
3. 0.04349 м
4. 0.00509 м

5. 0.01466 м

Самая удачная из моделей может предсказывать размеры помещений с точностью вплоть до 2 мм.

После получения положительных результатов на искусственно созданных импульсных характеристиках была выполнена проверка работоспособности описанного метода в реальных условиях.

Для проведения эксперимента было выбрано помещение с размерами: длина – 14.44 м., ширина – 5.78 м., высота – 2.77 м. Комната имеет стандартное для офисного помещения наполнение: несколько столов, два десятка стульев и несколько стеллажей. В качестве приёмника звука использовался микрофон FIFINE K669, в качестве источника звука для получения импульсной характеристики выступал щелчок пальцев.

Для этого эксперимента был сгенерирован датасет со следующими параметрами:

- размеры комнат: длина от 10 до 20 метров, ширина от 1 до 10 метров, высота от 3 до 4 метров;
- тип микрофона по диаграмме направленности – “кардиоидный”;
- время реверберации – 1.2 секунды.

Остальные параметры были оставлены по умолчанию. Структура нейронной сети была описана выше. В результате была получена модель с точностью предсказаний на тестовой выборке до 1.2 см.

Тренды для сгенерированной и реальной импульсной характеристики, представлены на рис. 2.

С целью тестирования полученной модели на реальных данных на нейронную сеть была подана реальная импульсная характеристика и были получены следующие результаты в формате - длина, ширина, высота в метрах: [15.8496 9.9983 4.9987].

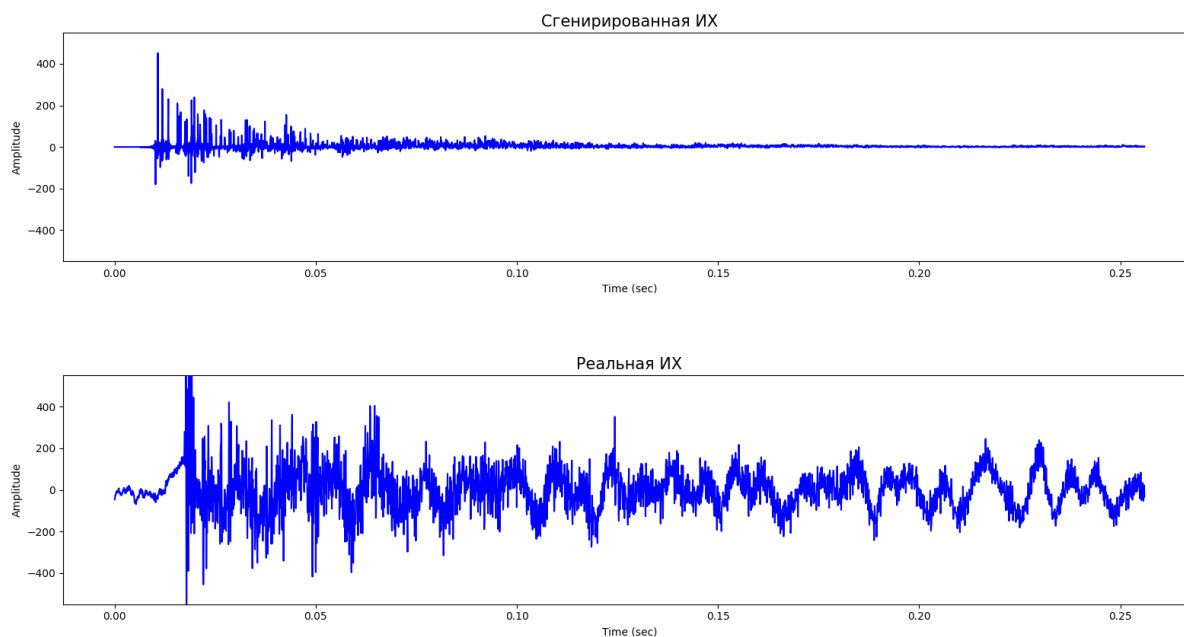


Рис. 2. – Сравнение импульсных характеристик

Обученная модель абсолютно не справилась с поставленной задачей. Предсказывая значения ширины и высоты комнаты, она выдавала предельно возможные значения по этим параметрам, которые при этом были максимально далеки от реальных значений.

Основной причиной подобного результата стало то, что на реальной импульсной характеристике присутствуют посторонние шумы, которые не позволили получить на модели точные результаты. Попытка удалить шумы программами для очистки звука привела к искажению структуры импульсной характеристики и было принято решение наложить шум на все сгенерированные характеристики [10,11].

Для этого был записан белый (стационарный) шум в том же помещении, где снимались реальные импульсные характеристики, а после случайная часть из звуковой дорожки этого шума накладывалась на нашу характеристику (формула (1)).

$$F(t) = f(t) + g(t), \quad (1)$$

где $f(t)$ – смоделированная импульсная характеристика; $g(t)$ – равномерная шумовая составляющая комнаты; $F(t)$ – зашумленная импульсная характеристика.

Реализацию такого наложения можно увидеть на рис. 3.

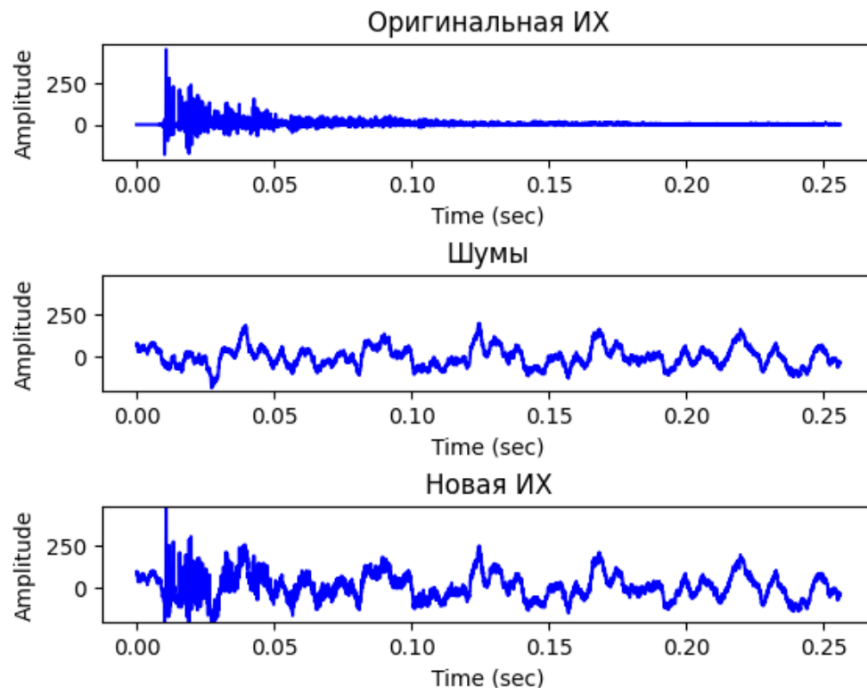


Рис. 3. – Наложение шума на импульсную характеристику

После обучения на зашумленных импульсных характеристиках на полученную нейронную сеть была вновь подана реальная импульсная характеристика и были получены следующие результаты в формате - длина, ширина, высота в метрах: [14.6292 5.3534 2.9869].

Относительное отклонение по длине составило – 1,31 %, по ширине 7,38%, по высоте – 7,83 %.

Перспективы дальнейшего исследования данной темы мы видим в повышении точности модели для предсказания размера помещения. Для этого планируется исследовать более широкий спектр структур и типов нейронных сетей и улучшить алгоритмы наложения шумов на сгенерированные импульсные характеристики.

Литература

1. Чачхиани Т.И., Серова М.Г. Мобильный робот в нестационарной среде //Инженерный вестник Дона. 2017, №4. URL: ivdon.ru/ru/magazine/archive/n4y2017/4546.
 2. Ming Chen, Simmons James A. Target Geometry Estimation Using Deep Neural Networks in Sonar Sensing. 2022. URL: deepai.org/publication/target-geometry-estimation-using-deep-neural-networks-in-sonar-sensing.
 3. Tervo Sakari, Tossavainen Timo. 3D room geometry estimation from measured impulse responses. 2012. URL: research.aalto.fi/en/publications/3d-room-geometry-estimation-from-measured-impulse-responses.
 4. Rajapaksha T., Qiu X., Cheng E., Burnett I. Geometrical room geometry estimation from room impulse responses. 2016, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 331–335.
 5. Allen J. B., Berkley D. A. Image method for efficiently simulating small-room acoustics. 1979, J. Acoust. Soc. Am., vol. 65, no. 4, pp. 943–950.
 6. Erlangen I. A. L. RIR generator. 2014. URL: audiolabs-erlangen.de/fau/professor/habets/software/rir-generat.
 7. Jager I., Heusdens R., Gaubitch N. D. Room geometry estimation from acoustic echoes using graph-based echo labeling. 2016, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1– 5.
 8. Пучков Е.В. Сравнительный анализ алгоритмов обучения искусственной нейронной сети // Инженерный вестник Дона. 2013, №4. URL: ivdon.ru/ru/magazine/archive/n4y2017/4546.
 9. Krizhevsky A., Sutskever I., Hinton G. E. Imagenet classification with deep convolutional neural networks. 2012, Proceedings of the 25th International Conference on Neural Information Processing Systems. URL: image-net.org/static_files/files/supervision.
-

10. Simonyan K., Zisserman A. Very deep convolutional networks for large-scale image recognition. 2014, CoRR, vol. abs/1409.1556. URL: arxiv.org/abs/1409.1556.

11. Abdel-Hamid O., Mohamed A., Jiang H., Deng L., Penn G., Yu D. Convolutional neural networks for speech recognition. 2014, IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 22, no. 10, pp. 1533–1545.

References

1. Chachkhiani T.I., Serova M.G. Inzhenernyj vestnik Dona. 2017, №4 URL: [magazine/archive/n4y2017/4546](https://magazine.archive/n4y2017/4546).

2. Ming Chen, Simmons James A. Target Geometry Estimation Using Deep Neural Networks in Sonar Sensing. 2022. URL: deepai.org/publication/target-geometry-estimation-using-deep-neural-networks-in-sonar-sensing.

3. Tervo Sakari, Tossavainen Timo. 3D room geometry estimation from measured impulse responses. 2012. URL: research.aalto.fi/en/publications/3d-room-geometry-estimation-from-measured-impulse-responses.

4. Rajapaksha T., Qiu X., Cheng E., Burnett I. Geometrical room geometry estimation from room impulse responses. 2016, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 331–335.

5. Allen J. B., Berkley D. A. Image method for efficiently simulating small-room acoustics. 1979, J. Acoust. Soc. Am., vol. 65, no. 4, pp. 943–950.

6. Erlangen I. A. L. RIR generator. 2014. URL: audiolabs-erlangen.de/fau/professor/habets/software/rir-generat.

7. Jager I., Heusdens R., Gaubitch N. D. Room geometry estimation from acoustic echoes using graph-based echo labeling. 2016, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1– 5.



8. Puchkov E.V. Inzhenernyj vestnik Dona, 2013, №4. URL: ivdon.ru/ru/magazine/archive/n4y2017/4546.

9. Krizhevsky A., Sutskever I., Hinton G. E. Imagenet classification with deep convolutional neural networks. 2012, Proceedings of the 25th International Conference on Neural Information Processing Systems. URL: imagenet.org/static_files/files/supervision.

10. Simonyan K., Zisserman A. Very deep convolutional networks for large-scale image recognition. 2014, CoRR, vol. abs/1409.1556. URL: arxiv.org/abs/1409.1556.

11. Abdel-Hamid O., Mohamed A., Jiang H., Deng L., Penn G., Yu D. Convolutional neural networks for speech recognition. 2014, IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 22, no. 10, pp. 1533–1545.