УДК 004.032.26:004.932

ПРИМЕНЕНИЕ РЕКУРРЕНТНЫХ НЕЙРОННЫХ СЕТЕЙ В ЗАДАЧЕ РЕКОНСТРУКЦИИ ТРЕХМЕРНЫХ ОБЪЕКТОВ ПО ИЗОБРАЖЕНИЯМ

Редько А.В., Рыбкин С.В.

Калужский филиал ФГБОУ ВО «Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет)», Калуга, e-mail: a.redko4@gmail.com

В данной статье была рассмотрена модель для решения задачи реконструкции трехмерного объекта по изображениям с различных ракурсов. Разработанная модель представляет собой нейронную сеть с многослойной архитектурой, в основе которой лежит рекуррентная сеть типа LSTM (Long Short Term Memory). Представлена архитектура многослойной нейронной сети, используемой для решения задачи реконструкции. Подробны описаны основные компоненты модели: сверточная нейронная сеть, используемая для кодирования 2D изображений в вектор более низкой размерности, рекуррентная LSTM нейронная сеть, декодирующая нейронная сеть на основе сверточной архитектуры, преобразующая скрытое состояние LSTM сети в трехмерную воксельную модель. Приведено описание реализации рассматриваемой модели с использованием языка Python и библиотеки Theano. Отмечены особенности процесса реконструкции при использовании нейронных сетей LSTM типа. Выполнен анализ качества реконструкции трехмерных объектов, выполненной при помощи предложенной модели. Приведено описание метрик, которые использовались для оценки качества реконструкции. Рассмотрена зависимость качества реконструкции от различных конфигураций LSTM-сети. Указаны основные преимущества и недостатки предложенной модели реконструкции трехмерных моделей по изображениям с различных ракурсов.

Ключевые слова: реконструкция трехмерных сцен, 3D модель, сверточные нейронные сети, рекуррентные нейронные сети, LSTM-сети

3D OBJECT RECONSTRUCTION USING RECURRENT NEURAL NETWORK

Redko A.V., Rybkin S.V.

Kaluga branch of the Federal State Budgetary Education Institution of Higher Education «Moscow State Technical University named after N.E. Bauman (National Research University)», Kaluga, e-mail: a.redko4@gmail.com

In this article, we describe a model designed to solve the Multiview 3D object reconstruction problem. The proposed model is a neural network that has a multilayer architecture and it is based on a LSTM network. We describe the architecture of the multilayer neural network used to solve the reconstruction problem. All main components of the network are showed in details: a convolutional neural network for encoding 2D images to a low-dimension feature vector, a recurrent LSTM network and a deconvolutional neural network that is used to decode hidden state of the LSTM network in order to obtain a 3D object. The specifics of the reconstructions process using LSTM neural network is considered. We made the analysis of the 3D reconstruction quality. Two metrics to estimate the reconstruction quality are described. We consider the quality of the reconstruction depending on different configuration of the LSTM network. The main advantages and disadvantages of the model are showed.

Keywords: 3D reconstruction, 3D model convolutional neural network, recurrent neural network, LSTM network

Получение трехмерной формы объекта по его изображениям с различных ракурсов является перспективным способом быстрого и качественного построения 3D моделей, которой можно найти множество различных областей применения [1–4].

Реконструкция трехмерных объектов по изображениям – это процесс получения 3D-модели объекта по набору его изображений, выполненных с различных ракурсов. В данной статье предлагается модель решения задачи реконструкции. Рассматриваемая модель представляет собой многослойную нейронную сеть, в основе которой лежит рекуррентная LSTM сеть.

Архитектура нейронной сети для реконструкции трехмерных сцен

Для решения задачи реконструкции трехмерных сцен используется нейронная сеть. Входными данными для нейронной сети является набор изображений трехмерного объекта с различных ракурсов. Выходом нейронной сети будет являться трехмерная модель.

Нейронная сеть состоит из трех основных компонентов:

- сверточная нейронная сеть для 2D изображений (Кодирующая нейронная сеть) [5, 6];
- рекуррентная Long Short Term Memory (LSTM) нейронная сеть [7–9];
- декодирующая нейронная сеть на основе сверточной архитектуры [5].

Принимая на вход одно или несколько изображений объекта с произвольных ракурсов, сверточная нейронная сеть сначала кодирует каждое входное изображение x в вектор низкой размерности $\tau(x)$. Затем данный кодированный вектор передается

на полносвязный слой. Выход полносвязного слоя соединен со входом рекуррентной LSTM нейронной сети, которая обновляет внутреннее состояние ячеек. На последнем этапе декодирующая нейронная сеть декодирует скрытые состояния блоков LSTM и генерирует трехмерную воксельную реконструкцию. Общая схема модели приведена на рис. 1.

Кодирующая сверточная нейронная сеть

Для преобразования изображения в вектор атрибутов используется многослойная сверточная нейронная сеть, структура которой изображена на рис. 2.

На вход подается RGB изображение размером 127x127 пикселей. Нейронная сеть состоит из 6 стандартных сверточных слоев. После каждого слоя свертки идет субдискретизирующий слой. В качестве функции активации для сверточных слоев используется линейный выпрямитель с утечкой (Leaky Rectified Linear Unit), который задается следующим образом:

$$LeakyReLU(x) = \begin{cases} \alpha x, x < 0 \\ x, x \geq 0 \end{cases},$$

где α – малая константа.

Выход кодирующей сети передается на вход полносвязного слоя, который сжимает выходное значение до вектора атрибутов τ размерностью 1024.

Рекуррентная LSTM нейронная сеть

Центральной частью реконструкции является рекуррентная нейронная сеть, которая позволяет модели «запомнить» ту часть объекта, которую она уже «видела», и обновлять свою память, когда появляется новые детали объекта.

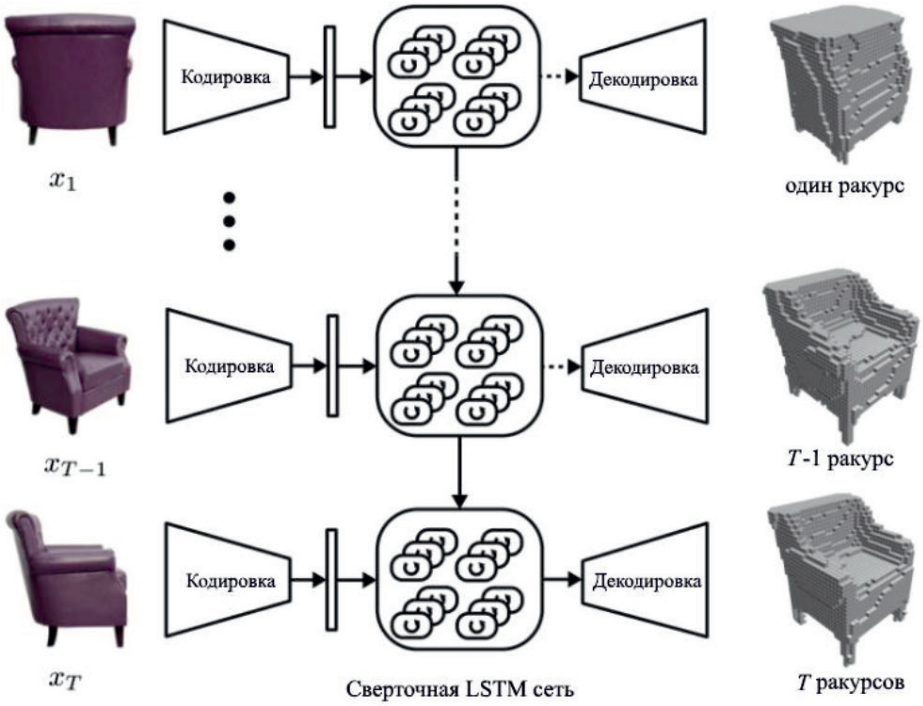

Рис. 1. Схема используемой модели реконструкции

Рис. 2. Кодирующая сверточная нейронная сеть

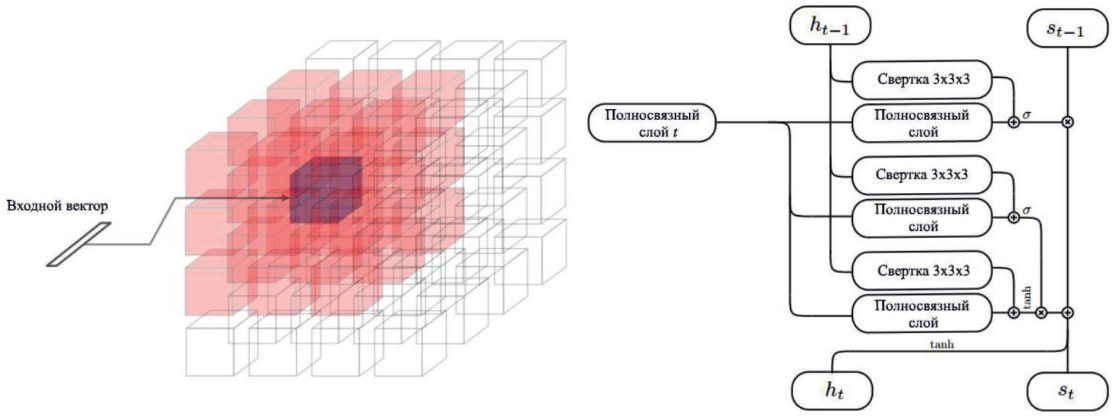

Рис. 3. Структура реконструирующей рекуррентной нейронной сети

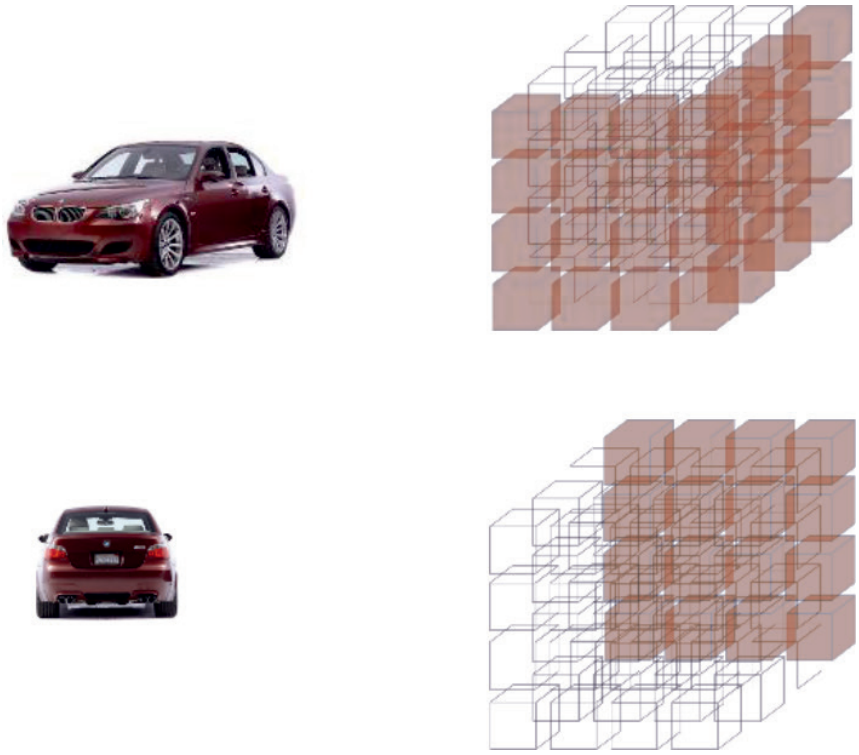

Рис. 4. Обновление состояния блоков, отвечающих только за определенный ракурс объекта

Модули LSTM организованы в трехмерную сетку размером 4x4x4. Использование классической архитектуры LSTM для выходного пространства такой большой размерности достаточно сложно. Поэтому используемая нейронная сеть отличается от стандартной LSTM сети применением сверточных слоев (рис. 3). Каждый из модулей сети отвечает за реконструкцию конкретной части конечного объекта.

Каждый модуль имеет независимое скрытое состояние $h_t(i, j, k)$ и запоминаемое состояние c_t. Для сокращения количества параметров обучения из классического LSTM-блока [9] был убран выходной фильтр. Вместо этого выходное значение сразу передается к следующему блоку без изменений.

$$f_t = \sigma(W_f \tau(x_t) + U_f h_{t-1} + b_f,$$

$$i_t = \sigma_g\left(W_i \tau(x_t) + U_i h_{t-1} + b_i\right),$$

$$c_t = f_t c_{t-1} + i_t \sigma_c\left(W_c \tau(x_t) + U_c h_{t-1} + b_c\right),$$

$$h_t = \sigma_h(c_t).$$

Интуитивно такая конфигурация позволяет блоку LSTM-сети обрабатывать несоответствие между конкретным предсказываемым участком реконструкции и моделью реального объекта, так что каждый блок учится восстанавливать одну область пространства воксела вместо того, чтобы реконструировать все пространство (рис. 4).

Кроме того, в классической LSTM сети все модули соединены друг с другом последовательно [9]. Это значит, что на текущее состояние h_t оказывают влияние все h_{t-1} элементы. В используемой архитектуре влияние на скрытое состояние оказывают только элементы, находящиеся по соседству с текущим. Количество соседей определяется размерностью используемого ядра свертки.

Декодирующая нейронная сеть

На последнем этапе используется декодирующая нейронная сеть, задачей которой является преобразование скрытого состояния (размерностью 4x4x4) к более высокому выходному разрешению (32x32x32). Количество слоёв свертки определяется требуемым в результате разрешением трехмерной модели.

Структура декодирующей нейронной сети изображена на рис. 5.

В используемой модели применяется 5 сверточных слоев. После последнего сверточного слоя результат, представляющий собой вектор $\nu \in R^{4\times4\times4\times2}$, преобразуется к трехмерной матрице вероятности занятости вокселем каждой ячейки (i, j, k). Данное преобразование выполняется с помощью функции softmax, применяемой к каждому вокселю:

$$softmax(v_i) = \frac{e^{v_i}}{\sum_j e^{v_j}}.$$

При использовании данной функции суммирование ведется по всем нейронам слоя сети. Такой выбор обеспечивает сумму выходов слоя, равную единице при любых значениях входа x и позволяет минимизировать сумму перекрестных энтропий.

Функция потерь нейронной сети

Функция потерь для модели нейронной сети определяется как сумма перекрестных энтропий всех вокселей:

$$L(\mathrm{X}, y) = \sum_{i,j,k} y_{(i,j,k)} \log\left(p_{(i,j,k)}\right) + \\ + \left(1 - y_{(i,j,k)}\right) \log\left(1 - p_{i,j,k)}\right),$$

где X – множество входных изображений, $y_{(i,j,k)}$ – соответствующее истинное значение занятости вокселя, $p_{(i,j,k)}$ – предсказываемая вероятность занятости вокселя.

Реализация

Для реализации модели нейронной сети использовалась библиотека Theano [10]. Для обучения сети использовались 3D модели, на основе которых генерировались входные изображения и выходные сетки вокселей. Все ракурсы выбирались случайно. В качестве входных использовались изображения размером 127x127, размер выходной сетки вокселей – 32x32x32

Исходные 3D модели и их изображения были взяты из базы данных ShapeNet. Набор данных был разделен на два подмножества: 80% данных использовалось для обучения, 20% для тестирования сети. Для слоя LeakyReLU использовался коэффициент α, равный 0,1. Для обучения использовался метод стохастического градиентного спуска [5].

Оценка качества реконструкции

На рис. 6 приведены результаты реконструкции для некоторых моделей.

Для оценки качества реконструкции использовались две метрики. В качестве основной использовалась метрика Intersection-over-Union, равная отношению количества вокселей, получаемых при пересечении результата реконструкции с исходной моделью, к количеству вокселей, получаемых при объединении результата с исходной моделью.

$$IoU = \frac{\sum_{i,j,k}\left[I\left(p_{(i,j,k)} > t\right) I\left(y_{(i,j,k)}\right)\right]}{\sum_{i,j,k}\left[I\left(p_{(i,j,k)} > t\right) + I\left(y_{(i,j,k)}\right)\right]},$$

где I – индикаторная функция, t – порог вероятности, при котором считается, что воксель должен быть закрашен, $p_{(i,j,k)}$ – вероятность закраски вокселя (результат реконструкции), $y_{(i,j,k)}$ – закраска вокселя в исходной модели.

Чем больше значение функции IoU, тем более точная модель была получена в результате реконструкции.

В качестве второй метрики использовалась сумма перекрестных энтропий вокселей:

$$L(\mathrm{X}, y) = \sum_{i,j,k} y_{(i,j,k)} \log\left(p_{(i,j,k)}\right) + \\ + \left(1 - y_{(i,j,k)}\right) \log\left(1 - p_{i,j,k)}\right).$$

Чем меньше значение данной метрики, тем выше качество реконструкции.

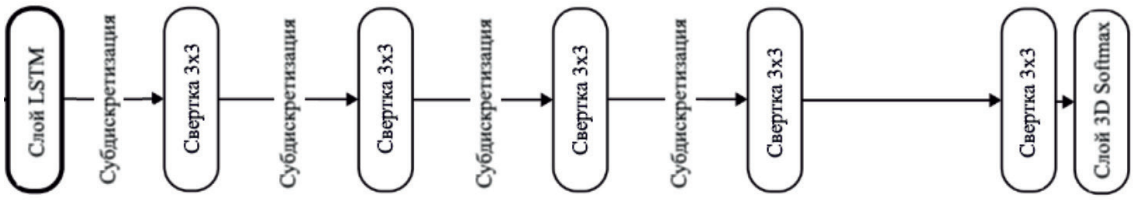

Рис. 5. Декодирующая нейронная сеть

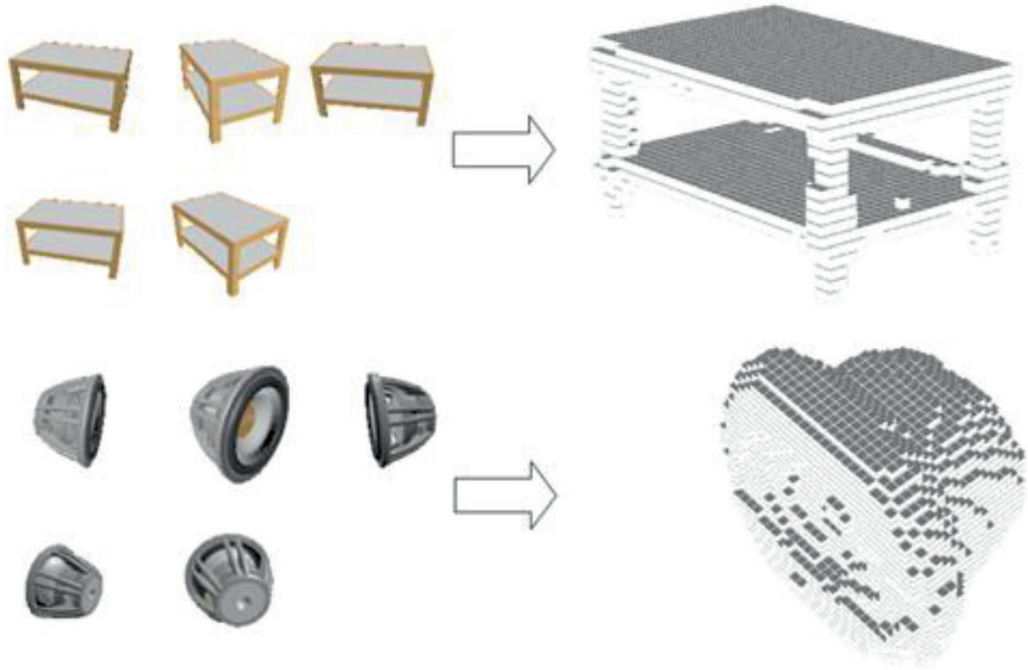

Рис. 6. Результаты реконструкции объектов

Исходная трехмерная модель преобразовывалась к воксельному представлению размером 32x32x32. Затем данная модель сравнивалась с моделью, полученной в ходе реконструкции.

Было проведено сравнение качества реконструкции в зависимости от используемого размера ядра свертки.

Средние результаты для LSTM сети и ядра свертки размером 1x1x1 и 3x3x3 приведены в таблице.

Результаты оценки качества реконструкции объектов LSTM сетью в зависимости от размера ядра свертки

Ядро свертки	IoU	Сумма перекрестных энтропий
1x1x1	0,499	0,116
3x3x3	0,539	0,106

Результаты показывают, что для ядра свертки размером 3x3x3 качество реконструкции выше. Использование ядра свертки размером 3x3x3 означает, что соседние рекуррентные блоки соединены между собой, тогда как при использовании ядра свертки размером 1x1x1 значения соседних рекуррентных блоков не учитываются.

Таким образом, была рассмотрена модель реконструкции трехмерных сцен по изображениям с различных ракурсов на основе нейронных сетей LSTM типа. Основным преимуществом использования такой сети является ее способность эффективно обрабатывать самоперекрывающиеся объекты. Если в очередном входном изображении видны детали объекта, которые ранее были перекрыты из-за ракурса съемки, то такая сеть обновит состояния LSTM для ранее невидимых деталей объекта, но при этом состояния других частей объекта остаются прежними. Это позволяет получить

достаточно качественную трехмерную модель объекта по небольшому количеству входных изображений, тогда как классические алгоритмы реконструкции [11–13] требуют не менее 10 различных ракурсов для получения качественной реконструкции. Однако, в отличие от классических алгоритмов, при увеличении количества входных изображений увеличение качества реконструкции происходит незначительно.

Список литературы

1. Редько А.В., Рыбкин С.В. Обзор методов реконструкции трехмерной сцены по изображениям // Электронный журнал: наука, техника и образование. – 2016. – № 4(9). – С. 86–91. URL: http://nto-journal.ru/uploads/articles/4a7f792360758d98e672a06eb07d88be.pdf (дата обращения: 30.03.2018)/

2. Кузнецов Г.С., Белов Ю.С. Обзор метода 3D распознавания лиц без преобразования лицевой поверхности // Электронный журнал: наука, техника и образование. – 2016. – № 2 (6). – С. 104–110. URL: http://nto-journal.ru/uploads/articles/c55b621e3573609edbc25b8820cddcb0.pdf (дата обращения: 30.03.2018).

3. Белов Ю.С., Гришанов К.М. Морфологические операции для уменьшения шума на изображении // Электронный журнал: наука, техника и образование. – 2016. – № 2(6). – С. 90–95. URL: http://nto-journal.ru/uploads/articles/1a285e863136e8b1a05511d28ee530c1.pdf (дата обращения: 30.03.2018).

4. Глебов С.А., Редько А.В. Применение методов фильтрации в задаче построения поверхности по облаку точек // Электронный журнал: наука, техника и образование. – 2017. – № СВ2(13). – С. 92–97. URL: http://nto-journal.ru/uploads/articles/8fb0d4bd92c1faf8d2b852c91d528e1f.pdf (дата обращения: 30.03.2018).

5. Sutskever I., Vinyals O., Le Q.V. Sequence to sequence learning with neural networks // Advances in neural information processing systems. – 2014. – P. 3104–3112.

6. Гришанов К.М., Рыбкин С.В. Тестирование сверточной нейронной сети в задачах машинного зрения // Электронный журнал: наука, техника и образование. – 2017. – № 2(12). – С. 186–193. URL: http://nto-journal.ru/uploads/articles/c578cad060b02dc5d46322258d9ad57a.pdf (дата обращения: 31.03.2018).

7. Sundermeyer M., Schluter R., Ney H. Lstm neural networks for language modeling // In: INTERSPEECH. – 2012. – P. 194–197.

8. Liu F., Shen C., Lin G. Deep convolutional neural fields for depth estimation from a single image. // IEEE Conf. Computer Vision and Pattern Recognition, 2015. https://arxiv.org/abs/1411.6387 (accessed 30.03.2018).

9. Olah C. Understanding LSTM Networks [Электронный ресурс]. URL: http://colah.github.io/posts/2015-08-Understanding-LSTMs/ (accessed 30.03.2018).

10. Theano [Электронный ресурс]. – URL: http://deeplearning.net/software/theano/ (accessed 30.03.2018).

11. Barnes C., Shechtman E., Finkelstein A., Goldman D. Patchmatch: A randomized correspondence algorithm for structural image editing // ACM Transactions on Graphics-TOG. – 2009. – no. 28(3). URL: http://gfx.cs.princeton.edu/pubs/Barnes_2009_PAR/patchmatch.pdf (accessed 30.03.2018).

12. Seitz M., Curless B., Diebel J., Scharstein D., Szeliski R. A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms // Proceedings of the IEEE conference on computer vision and pattern recognition. – 2006. – T. 1. – P. 519–526.

13. Белов Ю.С., Молчанов А.Н., Редько А.В. Использование алгоритмов определения ключевых точек в задаче реконструкции трехмерных сцен // Электронный журнал: наука, техника и образование. – 2017. – № СВ2(13). – С. 59–66. URL: http://nto-journal.ru/uploads/articles/fe26f5f18cd1b9e9c0d32299da69ef5a.pdf (дата обращения: 30.03.2018).