

УДК 004.932

Р.А. Чугунов, А.Д. Кульневич, С.В. Аксенов

Методика построения карт глубины стереоизображения с помощью капсулой нейронной сети

Дана оценка работе капсулой нейронной сети в задаче построения карты глубины по стереоизображению. Выполнены обзор существующих решений и теоретический обзор для решения задачи разработки программы. Описана структура нейронной сети и процесс ее обучения.

Ключевые слова: стереоизображение, карта глубины, машинное обучение.

doi: 10.21293/1818-0442-2019-22-1-83-86

Обработка стереоизображений с целью получения карты глубины изображения является важной задачей для исследования в области компьютерного зрения. Карта глубины изображения содержит информацию о расстоянии до объектов на данном изображении. Основными применениями могут быть как 3D-телевидение (3DTV) и телевидение со свободной точкой зрения (Free Viewpoint Television) [1–4], так и создание 3D-сенсоров. Такие сенсоры, позволяющие определять расстояние до объектов и строить трехмерную модель окружающего пространства, могут помочь при решении многих задач робототехники, дополненной и виртуальной реальности [5–7]. В частности, в робототехнике задача SLAM (одновременное построение карты и локализация робота). В [8] карты глубины используются для навигации антропоморфного робота. На рис. 1 представлен пример карты глубины.

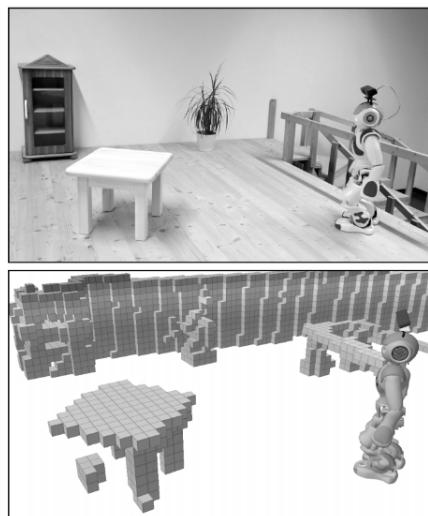


Рис. 1. Навигация антропоморфного робота с помощью карты глубины

В данной работе предлагается использовать капсулую нейронную сеть для решения задачи построения карты глубины по двум RGB-изображением стереопары. В отличие от традиционных методов стереореконструкции предлагаемая методика основана на использовании капсулой нейронной сети и напрямую не использует никаких знаний о геометрии стереозрения.

Целью данной работы является использование капсулой нейронных сетей как средства построение карт глубины для систем стереозрения, что является хорошим компромиссом между точностью карт и скоростью их построения. Также использование методов глубокого обучения может помочь строить карты глубины для изображений со сложным освещением, текстурой и пересечениями объектов.

Обзор литературы

Методы стереореконструкции широко исследовались с самого появления компьютерного зрения. Главной проблемой, которую данные методы пытаются решить, является проблема совмещения: имея два изображения одной и той же сцены, требуется совместить пиксели первого изображения с пикселями второго.

Существует большое число алгоритмов для решения данной проблемы, они обычно принадлежат одной из двух категорий: локальные методы и глобальные методы [9]. Как правило, эти методы сводят задачу совмещения пикселей исходных изображений к совмещению пикселей ректифицированных изображений [10, 11]. По паре соответствующих точек можно выполнить триангуляцию и определить координаты их прообраза в трехмерном пространстве. Зная трехмерные координаты прообраза, глубина вычисляется как расстояние до плоскости камеры.

Локальные методы часто используются, когда требуется работа в реальном времени и производительность ценится выше точности. Такие методы часто включают в себя минимизацию фотометрической ошибки, которая рассчитывается по окружающим пикселям изображений. Для расчета ошибки могут использоваться различные функции, например: сумма модулей разностей (SAD), взаимная корреляция, L2-норма и т.д. Эти методы, как правило, имеют большой процент ошибок в сценах со сложным освещением, зачастую полученные карты глубины зашумлены.

Глобальные методы используются в случае, когда точность построенной карты глубины важнее вычислительной эффективности алгоритма. Поэтому такие методы, как правило, используется в тех приложениях, в которых не требуется работа в реальном времени. Эти алгоритмы добавляют условие пространственной гладкости к задаче минимизации фотометрической ошибки.

Капсулльные нейронные сети уже применялись в задаче сегментации изображений [12]. В данной работе применена нейронная сеть такой же структуры, однако в применении к задаче построения карты глубины. В такой сети используются сверточные и капсулльные слои. Данные алгоритмы относятся к классу локальных методов. В [13, 14] сверточная сеть использовалась для расчета карты глубины по одному изображению, в [15] сеть такого типа использовалась для определения соответствующих друг другу частей левого и правого изображений, однако построение карты глубины основывалось на традиционных методах стереореконструкции. В [16] полностью связная нейронная сеть обучалась на разреженных векторах признаков для построения карты глубины по одному изображению. Применение капсулльных сетей в прикладных задачах пока не отражено широко в научных публикациях.

Проверка алгоритма

Выборка для разработки алгоритма создавалась процедурно. Генерация синтетической выборки изображений для обучения осуществлялась с помощью программы для создания трехмерной компьютерной графики, а также языка программирования Python. Всего было получено 5619 изображений трехмерных фигур на произвольном расстоянии и под произвольным углом к стереокамере.

Сверточные сети часто применяются в задачах сегментации, где входное изображение преобразуется в маску, каждый пиксель которого может иметь одно из дискретных значений – индекс класса, к которому относится данный объект изображения. Од-

нако если не преобразовывать выходное изображение в бинарную маску, то сверточные или капсулльные сети можно также использовать для построения карты глубины. В данном случае на вход нейронной сети поступают левое и правое изображения, объединенные по вертикали в единое изображение размера 720×640 пикселей (части стереоизображения по отдельности имеют размер 360×320 пикселей). На рис. 2 представлен пример стереоизображения.

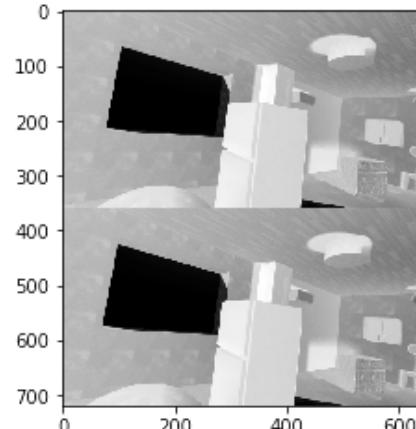


Рис. 2. Пример стереоизображения

Используется нейронная сеть, построенная на основе сети SegCaps.

Данная структура капсулльной нейронной сети основана на сверточной нейронной сети U-Net [17] для сегментации изображений. На рис. 3 представлена структура нейронной сети SegCaps.

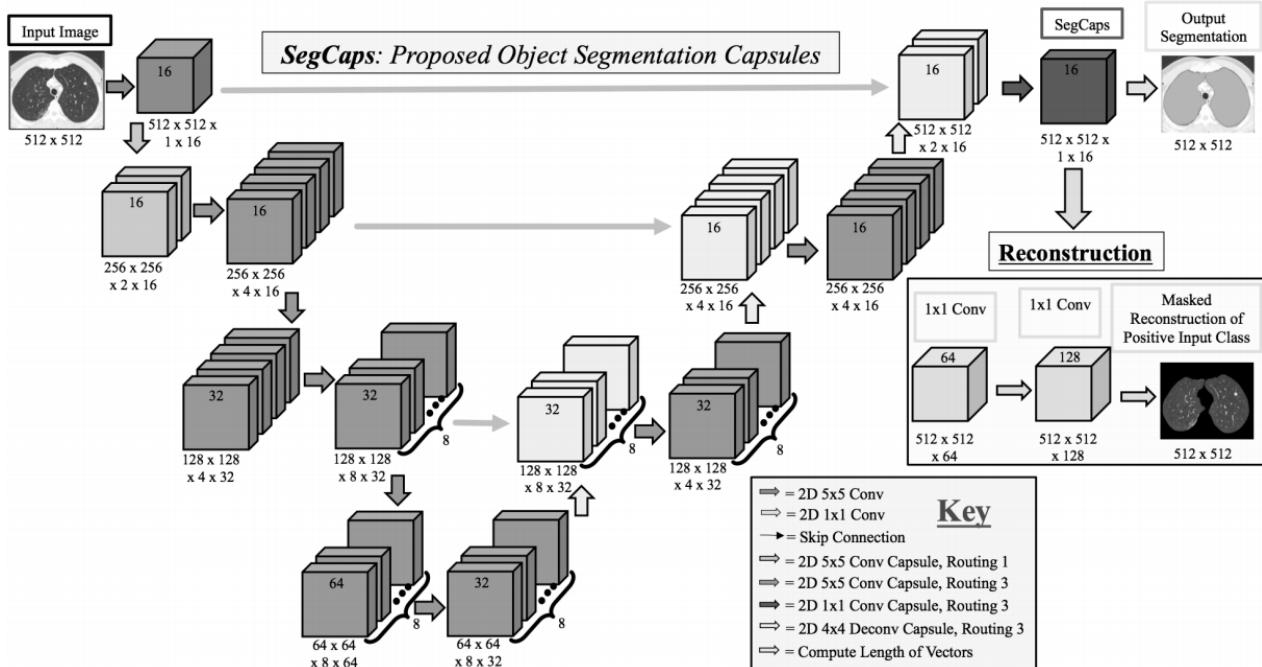


Рис. 3. Структура нейронной сети SegCaps

Процесс обучения нейронной сети состоит из следующих шагов:

1. Считываются сгенерированные стереоизображения и карты глубины.

2. Значение каждого пикселя делятся на 255, значения карт глубины делятся на 10, чтобы быть в пределах от 0 до 1.

3. Созданная выборка изображений случайным образом (распределение вероятностей – равномерное) разбивается на тестовую, валидационную и обучающую в пропорции 10/10/80%. Обучающая выборка используется для обучения нейронной сети, валидационная используется для оптимизации параметров. Тестовая выборка требуется для окончательной несмешенной оценки работы алгоритма.

4. Создается нейронная сеть приведенной выше структуры.

5. Обучающая выборка подается на вход модели.

6. Выполняется обучение модели, в качестве оптимизируемой функции: среднеквадратическая ошибка между истинными картами глубины и прогнозируемыми ответами (mean squared error) на обучающей выборке. Модель обучается до тех пор, пока за последние 5 эпох есть уменьшение минимальной среднеквадратической ошибки на валидационной выборке (проверка выполняется каждую эпоху обучения).

Результаты

По итогам обучения нейронной сети была оценена средняя относительная ошибка расчета глубины на тестовой выборке. На рис. 4 представлен пример построенной нейросетью карты глубины.

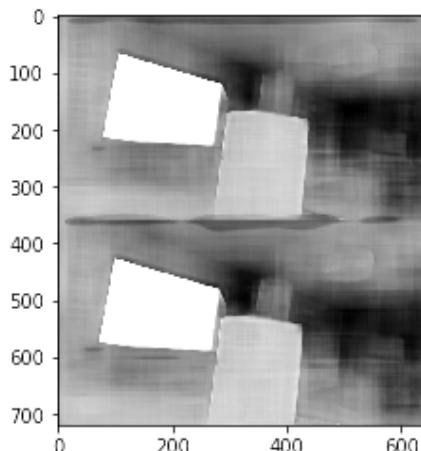


Рис. 4. Пример карты глубины, построенной нейросетью на тестовой выборке

Было проведено сравнение по качеству и скорости работы метода на основе нейронной сети с классической сверточной сетью U-Net, обученной на той же выборке. Результаты сравнения представлены в таблице.

Результаты сравнения

Метод	Классический U-Net	Капсульная нейронная сеть
Средняя абсолютная ошибка, м	0,43	0,25
Время построения одного изображения, с	0.02	0,0233
Частота обработки, кадров/с	50	43

Капсульная сеть позволяет точнее строить карты глубины, чем полностью сверточная нейронная сеть.

Заключение

Показана эффективность работы капсульных нейронных сетей в задаче построения карты глубины по стереоизображению. Методика не требует использования других способов обработки изображений, в частности, алгоритмов получения разностного изображения и фильтрации шумов. Данная методика позволяет заметно снизить ошибку определения глубины стереоизображения.

Литература

- Zhang K. Cross-based local stereo matching using orthogonal integral images / K. Zhang, J. Lu, G. Lafruit // Circuits and Systems for Video Technology // IEEE Transactions on. – 2009. – Vol. 19(7). – P. 1073–1079.
- Graham T. 3d image sequence acquisition for tv & film production / T. Graham, O. Grau // In 3DPVT. – IEEE. – 2002. – 320 p.
- Flack J. Low-bandwidth stereoscopic image encoding and transmission / J. Flack, P.V. Harman, S. Fox // In Electronic Imaging 2003. – International Society for Optics and Photonics. – 2003. – P. 206–214.
- Fehn C. Depth-image-based rendering (dibr), compression, and transmission for a new approach on 3d-tv // In Electronic Imaging 2004. – International Society for Optics and Photonics. – 2004. – P. 93–104.
- Kim H. 3d scene reconstruction from multiple spherical stereo pairs / H. Kim, A. Hilton // International journal of computer vision. – 2013. – Vol. 104(1). – P. 94–116.
- Dynamic 3d scene reconstruction in outdoor environments / H. Kim, M. Sarim, T. Takai, J.Y. Guillemaut, A. Hilton // In Proc. IEEE Symp. on 3D Data Processing and Visualization. – 2010. <https://core.ac.uk/download/pdf/30340214.pdf>
- Schonbein M. Omnidirectional 3d reconstruction in augmented manhattan worlds / M. Schonbein, A. Geiger // In Intelligent Robots and Systems (IROS 2014). – 2014. – IEEE / RSJ International Conference. – 2014. – P. 716–723.
- Maier D. Real-Time Navigation in 3D Environments Based on Depth Camera Data / D. Maier, A. Hornung, M. Bennewitz // 12th IEEE-RAS International Conference on Humanoid Robots (Humanoids 2012). – 2012. – P. 692–697.
- Mohan D. A Review on Depth Estimation for Computer Vision Applications / D. Mohan, Dr. A. R. Ram // International Journal of Engineering and Innovative Technology (IJEIT). – 2015. – Vol. 4. – Is. 11. – <https://pdfs.semanticscholar.org/f75b/2030713b36addde9b9d551ee003c20bac3ca.pdf>
- Fusiello A. A compact algorithm for rectification of stereo pairs / A. Fusiello, E. Trucco, A. Verri // Machine Vision and Applications. – 2000. – Vol. 12(1). – P. 16–22.
- Szeliski R. Computer vision: algorithms and applications // Springer Science & Business Media. – 2010. – http://szeliski.org/Book/drafts/SzeliskiBook_20100903_draft.pdf
- LaLonde R. Capsules for Object Segmentation / R. LaLonde U. Bagci // CorR. – 2018. – <https://openreview.net/pdf?id=B1L4Gb3sf>
- Eigen D. Depth map prediction from a single image using a multi-scale deep network / D. Eigen, C. Puhrsch, R. Fergus // In Advances in Neural Information Processing Systems. – 2014. – P. 2366–2374.
- Liu F. Deep convolutional neural fields for depth estimation from a single image / F. Liu, C. Shen, G. Lin // 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2015. – P. 5162–5170.
- Zbontar J. Stereo Matching by Training a Convolutional Neural Network to Compare Image Patches / J. Zbontar, Y. LeCun // Journal of Machine Learning Research. – 2016. – Vol. 17, No. 65. – P. 1–32.

16. Depth estimation from monocular infrared images based on svm model / L. Xi, S. Sun, L. Li, F. Zou // *Laser & Infrared*. – 2012. – Vol. 11:025. – P. 111–115.
17. Ronneberger O. U-Net: Convolutional Networks for Biomedical Image Segmentation / O. Ronneberger, P. Fischer, T. Brox // *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Springer, LNCS, 2015. – Vol. 9351. – P. 234–241.
-
- Чугунов Роман Анварович**
Магистрант отд. информационных технологий (ОИТ)
Инженерной школы информационных технологий
и робототехники (ИШИТР)
Национального исследовательского
Томского политехнического университета (НИ ТПУ)
Ленина пр-т., 30, г. Томск, Россия, 634050
Тел.: +7-913-106-49-63
Эл. почта: manofnovel@gmail.com
- Кульневич Алексей Дмитриевич**
Магистрант отд. ОИТ ИШИТР НИ ТПУ
Ленина пр-т., 30, г. Томск, Россия, 634050
Тел.: +7-952-892-68-90
Эл. почта: kulnevich94@mail.ru
- Аксенов Сергей Владимирович**
Канд. техн. наук, доцент отд. ОИТ ИШИТР НИ ТПУ
Ленина пр-т., 30, г. Томск, Россия, 634050
Тел.: +7-913-887-47-90
Эл. почта: axyonov@tpu.ru

Chugunov R.A., Kulnevich A.D., Aksenov S.V.
Depth estimation based on convolutional neural networks

In this paper, the SegCaps neural networks were tested on the task of depth map prediction using stereo images. Overview of existing solutions was carried out to develop the algorithm. The article describes the structure of the neural network and the network training process.

Keywords: stereo image, depth map, machine learning.
doi: 10.21293/1818-0442-2019-22-1-83-86

References

1. K. Zhang, J. Lu, G. Lafruit. Cross-based local stereo matching using orthogonal integral images. *Circuits and Systems for Video Technology. IEEE Transactions on Circuits and Systems for Video Technology* on, 2009, vol. 19(7), pp. 1073–1079.
2. T. Graham, O. Grau. 3d image sequence acquisition for tv & film production. In 3DPVT. *IEEE*, 2002, 320 p.
- 3 J. Flack, P.V. Harman, S. Fox. Low-bandwidth stereoscopic image encoding and transmission. In Electronic Imaging 2003. *International Society for Optics and Photonics*, 2003, pp. 206–214.
4. Fehn C. Depth-image-based rendering (dibr), compression, and transmission for a new approach on 3d-tv. In Electronic Imaging 2004. *International Society for Optics and Photonics*, 2004, pp. 93–104.
5. H. Kim, A. Hilton . 3d scene reconstruction from multiple spherical stereo pairs. *International journal of computer vision*, 2013, vol. 104(1), pp. 94–116.
6. H. Kim, M. Sarim, T. Takai, J.Y. Guillemaut, A. Hilton. Dynamic 3d scene reconstruction in outdoor environments. In Proc. *IEEE Symp. on 3D Data Processing and Visualization*, 2010. <https://core.ac.uk/download/pdf/30340214.pdf>
7. M. Schonbein, A. Geiger. Omnidirectional 3d reconstruction in augmented manhattan worlds. In *Intelligent Robots and Systems (IROS 2014)*, 2014, IEEE/RSJ International Conference, 2014, pp. 716–723.
8. D. Maier, A. Hornung, M. Bennewitz. Real-Time Navigation in 3D Environments Based on Depth Camera Data. *12th IEEE-RAS International Conference on Humanoid Robots (Humanoids 2012)*, 2012, pp. 692–697.
9. D. Mohan, Dr. A. R. Ram. A Review on Depth Estimation for Computer Vision Applications. *International Journal of Engineering and Innovative Technology (IJEIT)*, 2015, Vol. 4, Is. 11. <https://pdfs.semanticscholar.org/f75b/2030713b36addde9b9d551ee003c20bac3ca.pdf>
10. A. Fusello, E. Trucco, A. Verri. A compact algorithm for rectification of stereo pairs. *Machine vision and applications*, 2000, vol. 12(1), pp. 16–22.
11. Szeliski R. Computer vision: algorithms and applications. *Springer Science & Business Media*, 2010. http://szeliski.org/Book/drafts/SzeliskiBook_20100903_draft.pdf
12. R. LaLonde U. Bagci. Capsules for Object Segmentation. CoRR, 2018. <https://openreview.net/pdf?id=B1L4Gb3sf>
13. D. Eigen, C. Puhrsch, R. Fergus. Depth map prediction from a single image using a multi-scale deep network. In *Advances in Neural Information Processing Systems*, 2014, pp. 2366–2374.
14. F. Liu, C. Shen, G. Lin. Deep convolutional neural fields for depth estimation from a single image. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 5162–5170.
15. J. Zbontar, Y. LeCun. Stereo Matching by Training a Convolutional Neural Network to Compare Image Patches. *Journal of Machine Learning Research*, 2016, vol. 17, no. 65, pp. 1–32.
16. L. Xi, S. Sun, L. Li, F. Zou. Depth estimation from monocular infrared images based on svm model. *Laser & Infrared*, 2012, vol. 11:025, pp. 111–115.
17. O. Ronneberger, P. Fischer, T. Brox. Ronneberger O. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Springer, LNCS, 2015, vol. 9351, pp. 234–241.

Roman A. Chugunov

Master student, Department of Information Technologies (DIT), Engineering School of Information Technology and Robotics (ESITR), National Research Tomsk Polytechnic University (NR TPU) 30, Lenina pr., Tomsk, Russia, 634050 Phone: +7-913-106-49-63 Email: manofnovel@gmail.com

Aleksey D. Kulnevich

Master student DIT, ESITR, NR TPU 30, Lenina pr., Tomsk, Russia, 634050 Phone: +7-952-892-68-90 Email: a.d.kulnevich@gmail.ru

Sergey V. Aksenov

Candidate of Engineering, Assistant Professor, DIT, ESITR, NR TPU 30, Lenina pr., Tomsk, Russia, 634050 Phone: +7-913-887-47-90 Email: axyonov@tpu.ru