基于深度卷积神经网络的双目立体视觉匹配算法

肖进胜[1,2*]，田红[1]，邹文涛[1]，童乐[1]，雷俊锋[1]

[1]武汉大学电子信息学院，湖北 武汉 430072；

[2]地球空间信息技术协同创新中心，湖北 武汉 430079

摘要 对于基于块进行立体匹配的深度学习方法而言，网络结构的设计对匹配代价的计算至关重要，同时，卷积神经网络(CNN)在图像处理时的耗时问题也亟待解决。提出一种基于“缩小型”网络的 CNN 立体匹配方法。利用 CNN 训练左右图像块的相似性，计算出立体匹配的匹配代价。其中，CNN 特征提取阶段，通过对每个层增加相应的批归一化层，可以使训练使用更大的学习率，加快网络训练收敛速度。另外，网络设计中全连接层采用“逐层缩小”的形式，结合上述网络优化和损失函数改善，在保证精度的同时提高了运行速度。使用 KITTI 数据集对算法进行验证，实验结果证明，相比目前国内外先进方法，本文算法在精度方面有一定优势，相比部分方法，速度有较大提升。

关键词 机器视觉；立体匹配；匹配代价；相似性学习；卷积神经网络

中图分类号 TP391.41 **文献标识码** A **doi**：10.3788/AOS201838.0815017

Stereo Matching Based on Convolutional Neural Network

Xiao Jinsheng[1,2*], Tian Hong[1], Zou Wentao[1], Tong Le[1], Lei Junfeng[1]

[1] *Electronic Information School, Wuhan University, Wuhan, Hubei* 430072, *China*;

[2] *Collaborative Innovation Center of Geospatial Technology, Wuhan, Hubei* 430079, *China*

Abstract For the stereo matching method of deep learning based on patches, the network structure is vital for the calculation of the matching cost, and the time-consuming of convolutional neural network (CNN) in the image processing field also needs to be solved. We propose a stereo matching method of CNN based on a “shrink network”. The CNN method is utilized to train the similarity of the left and right image patches, and the matching cost of the stereo matching is obtained by the similarity. At the feature extraction stage, by adding batch normalization layers to each layer, the gradient dispersion in the backward propagation can be improved effectively. Besides, the full-connection layer adopts a "layer-by-layer reduction" form with other network optimizations to increase the speed while ensuring the accuracy. We utilize the KITTI datasets to test the algorithm. Experimental results demonstrate that the proposed method increases the accuracy and speed fairly compared to some other methods.

Key words machine vision; stereo matching; matching cost; similarity learning; convolutional neural network

OCIS codes 150.0155; 100.6640, 200.4260

1 引 言

在过去的数十年间，立体视觉被广泛运用于三维重建、机器人、智能驾驶等领域。从立体图像获取的信息中得到稠密精确的深度图一直是国内外研究者们竞相追求但又是十分困难的事情。其中，立体匹配是立体视觉最重要也是最难的一环。到目前为止，传统的立体匹配方法大致可分为三类：全局匹配，局部匹配，以及二者相结合形成的半全局匹配。全局匹配方法[1]试图求解的是全局内的最优解，计算十分缓慢，且未必能找到全局最优解。全局方法不需要代价聚合这一步骤，所以匹配代价的计算方式和最优视差的选择策略对全局方法影响很大。局部匹配算法主要通过比较待匹配点一定范围内的局部特性进行匹配，因此，十分依赖于匹配窗口的合理性，而且对弱纹理区域和遮挡区域的处理效果不好。基于局部区域[2-3]获取信息的匹配方法有多种，如基于自适应窗口[4]的方法将像素预测和局部平滑组合成能量函数。

收稿日期：2018-03-27；**修回日期**：2018-05-07；**录用日期**：2018-05-11

基金项目：国家自然科学基金(61471272)、湖北省自然科学基金(2016CFB499)

* **E-mail**：xiaojs@whu.edu.cn

半全局块匹配[5]是全方法和局部方法的一种结合，执行逐像素匹配代价计算，在一维平滑约束的最佳路径搜索中利用动态规划算法实现最优路径的搜索。这些方法均使用手动设计的代价函数，而且在大多数情况下仅学习到了数据特征之间的线性组合。

近年来，利用卷积神经网络(CNN)[6]进行立体匹配的方法越来越多。Žbontar 等[7]充分利用了基于块的图像匹配在计算机视觉领域的广泛应用[8]，使用 CNN 找到各块之间准确的对应关系，然后根据不同块的对应关系进行不同程度的相似性打分。相似性的负值被定义为立体匹配的代价，用于后续的代价聚合和视差计算。当然，由于存在太多影响图像外观的因素，譬如镜头畸变、照明变化、遮挡等，决定两块是否匹配的问题仍非常具有挑战性。正如前面讨论过的局部匹配，手动设计特征描述符可能无法以最佳的方式考虑到所有决定块外观的因素。受此启发，在 Žbontar 等[7]提出的基于卷积神经网络的匹配代价(MC-CNN)的基础上，改进其网络结构，并在学习到块相似性的基础上，将相似性的负值定义为代价进行代价聚合，分别对半全局匹配、交叉代价聚合，以及斜平面平滑[9]等后处理方法运用网络学习到的匹配相似性，计算得到更加稠密的视差图。

本文的主要贡献：1）摒弃传统固定尺寸的全连接层，使用逐次“缩小”的全连接层，结合后续网络优化，在提高运行速度的同时保证了输出精度；2）在网络训练的后续阶段设计损失函数，使正负样本的相似性逼近一个更加松弛的约束，提高了输出网络的匹配效果；3）在特征提取阶段，在每个卷积层后引入相应的批归一化层[10]，在使用较大学习率的情况下加速网络训练的收敛。

分别用 KITTI2012 和 KITTI2015 数据集进行验证。实验证明，相比于传统的全局匹配方法[1,11]或局部匹配方法[5,9]，本文算法在立体匹配获取稠密视差图方面有更好的表现。同时，本文算法与 MC-CNN[7]和 Luo 等[12]深度学习方法的结果也有一定的可比性。

2　相关工作

除了之前大量的全局[1]或局部[4,13]的立体匹配算法以外，深度学习用于立体匹配的大致可以分为两大类：基于块相似性比较的方法和端到端从左右图像对学习视差图的方法。Luo 等[12]提出了一个基于块的匹配网络，能够在不到 1 s 的时间内计算产生非常准确的视差结果。该方法利用一个简单的内积计算暹罗结构产生的两个网络层，并将网络认定为多分类问题，而不是单纯将匹配结果归为“相似”或“不相似”。相比于 MC-CNN[7]，在损失一定主观效果的情况下，该方法的运行速度大幅提高。Zagoruyko 等[14]同样利用块匹配原理设计了多个适用于立体视觉的神经网络结构，并且研究了这些不同架构的权衡和优势。

对于端到端的学习方法，最近，Kendall 等[15]提出一种基于三维(3D)卷积端到端学习视差的方法。该方法利用图像的几何特性进行深度特征提取，以此构造匹配代价体。其中，3D 卷积用来理解语义并改善视差估计。该方法能够端到端地实现亚像素精度的视差学习，无需额外的后处理或正则化。另外一种经典的端到端网络是由 Güney 等[16]提出的 Displet 结构：使用基于稀疏视差估计和图像语义分割的反向图形技术对特定对象类的视差进行建议来规范化视差范围。该方法优化了具有反射性和无纹理平面的处理效果，解决了立体匹配的歧义问题。2016 年，Mayer 等[17]提出了 DispNet 网络结构，将光流估计的应用扩展到视差估计，基于改进的 FlowNet 结构[18]提出了一个实时估计视差的神经网络。FlowNet 结构是一个“沙漏型”网络，由“收缩”和“放大”两部分组成。其中，“收缩”部分主要由卷积层组成，用于深度地提取两个图片的一些特征。“放大”部分可以恢复“收缩”部分导致的分辨率降低，主要由反卷积组成。DispNet 网络充分利用图像的帧间信息，可以实现实时预测光流。DispNet 还有一个贡献：由于双目视觉数据集的匮乏，因此仿真了几千张的双目图像及其基准视差图像，在该数据集上训练的网络在实际数据以及 KITTI 数据集上也有较好的验证结果。受到 FlowNet 等方法的启发，Pang 等[19]提出了输入信息更加丰富的级联残差学习(CRL)网络。CRL 由 DisFullNet 和 DisResNet 两个部分组成，其中，DisResNet 的输入既包括 DisFullNet 网络的输出视差图，也包括视差图对右图进行 warp 操作生成的左图，以及原始左图和生成左图的差值图。这样丰富的输入可以保证在优化初始视差图的前提下，给后续网络提供误差先验，使网络的训练变得更加简单。

3　基于 MC-CNN 方法的改进

3.1　基于 MC-CNN 方法的立体匹配

众所周知，立体匹配的过程即在匹配图中找到与基准图相对应像素点的过程。然而，众多研究学

者将深度学习用于立体匹配不是直接在左右原图中找匹配点，而是通过将图像划分为多个块，比较各块的相似性试图找到相对应的像素点。如图 1 所示，假设(a)、(b)分别是来自左右图像的块，需要计算左图上的某点 $p=(x,y)$和右图上的某点 $q=(x-d,y)$的匹配代价，其中，d 为所有在考虑范围的视差，即场景中同一点在左右图像中的像素坐标差。基于图像块匹配的思想，该算法可解释为以 p 点为中心的块与以 q 点为中心的块相似度，相似度越小，匹配代价越大。

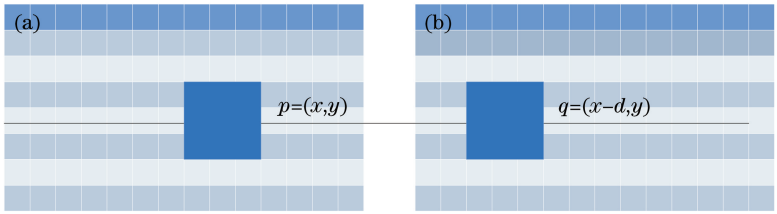

图 1　基于块的匹配。(a)左图中的块；(b)右图对应的块

Fig. 1　Matching based on patch. (a) Patch in the left image; (b) patch in the right image

原始 MC-CNN 方法的流程图如图 2 左边所示，结合上述相似性比较的分析可总结到，基于 MC-CNN 方法的双目立体视差计算的具体步骤可归纳如下。

1) 分别使用 4 层卷积神经网络层提取左右图像块的不同特征信息。

2) 联结左右特征信息，利用 4 层全连接层网络对特征进行分类判断。其中，损失函数使用交叉熵代价函数：

$$t\ \log(s)+(1-t)\log(1-s), \tag{1}$$

式中 s 为相似性比较网络的输出，t 为样本标记，当输入为正样本时 $t=1$，输入为负样本时 $t=0$。

3) 对相似/不相似的判断结果以相似性打分的形式输出，用于后续的立体匹配。

4) 利用相似性打分的反比例构造代价函数，图像块越相似代价，反之，图像块相似性越小，代价越大。

5) 对代价函数分别进行交叉代价聚合、半全局匹配等后处理，计算得到最终视差图。

虽然 MC-CNN 方法相比大部分传统立体匹配算法在精度上有一定的提高，但是算法的网络结构决定此方法耗时较长。另外，损失函数的设计大幅影响网络对立体匹配的效果。通过分析 MC-CNN 的网络架构和算法各模块的运行时间，从网络结构和损失函数方面对现有 MC-CNN 方法进行改进。

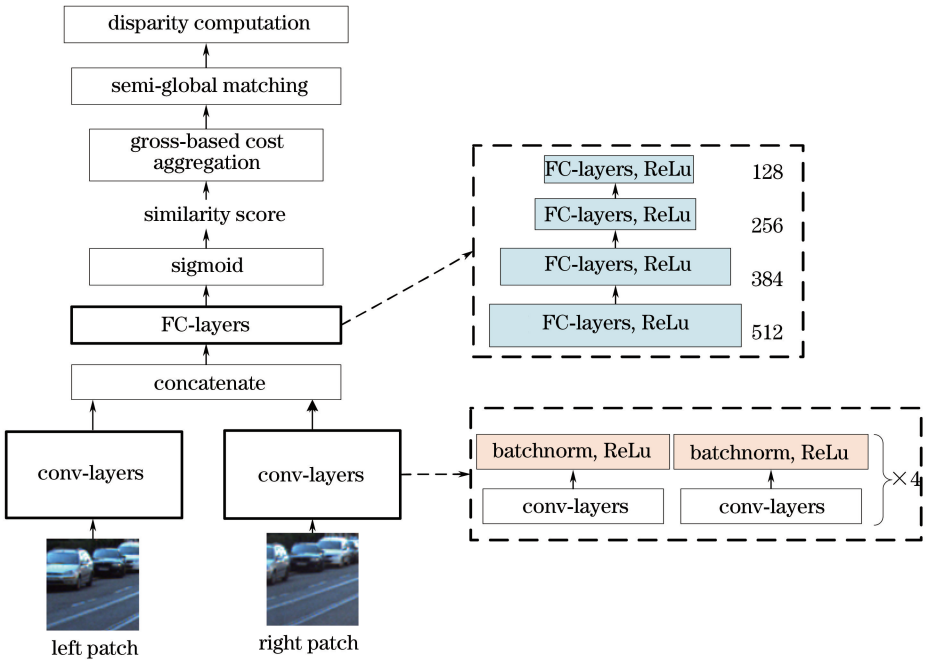

图 2　MC-CNN 改进前后的网络结构图

Fig. 2　Structure of MC-CNN net before and after modification

3.2　网络结构上的改变

如图 2 右部分所示，基于 MC-CNN 网络，提出一个改进的网络结构。框图中虚线框内的部分为本文在网络结构上的主要改进。首先，在每个卷积层后面增加批归一化层，加速网络训练的收敛速度。同时，批归一化使训练能够使用更高的学习率而无需太多的初始化操作。其次，原始 MC-CNN 网路结构中使用 4 层或者 5 层尺寸相同的全连接层来实现对块相似性的判定，本文将 4 层全连接层设计成“缩小”形式。其中，“缩小型”全连接层的神经元尺寸分别为 512、384、256、128。结合其他网络优化操作，在保证计算精度不下降的同时，每个左右图像对的测试时间降低了约 20%。

由图 2 可知，本文的网络设计流程：1）经过 4 次卷积层＋批归一化＋线性整流函数（ReLu）激活层的迭代；2）将左右特征经过一个联结层进行连接；3）进行 4 次全连接层＋ReLu 激活层迭代，其中，全连接层的尺寸逐次缩小；4）经过 Sigmoid 激活函数对所选块的相似性进行打分。

3.3　损失函数的改进

在建立数据集的正负样本时，选用 KITTI2012 和 KITTI2015 立体数据集来构建二分类数据集。在真实视差已知的每个图像像素点处，分别提取一个正样本和负样本，以保证数据集包含相同数量的正负样本。以图 1 为例，为了丰富正负样本的数据，通过将右图像块的中间像素点设置为以下形式得到正样本：

$$q=(x-d+o_{\mathrm{pos}},y),\tag{2}$$

式中 o_{pos} 是 $[-V_{\mathrm{pos}},V_{\mathrm{pos}}]$ 范围内的一个随机数，每次迭代重新随机生成一个数。根据大量实验证实，V_{pos} 被设置为 4 时匹配效果最佳。同理，将负样本设计为

$$q=(x-d+o_{\mathrm{neg}},y),\tag{3}$$

式中 o_{neg} 是 $[-Z_{\mathrm{low}},Z_{\mathrm{high}}]$ 范围内的一个随机数，与 o_{pos} 的生成同理，Z_{low} 和 Z_{high} 分别被设置为 4 和 18。这里，o_{pos} 没有被设置为 0 与后续将使用到立体方法有关，即当“优良”匹配和近似“优良”匹配的匹配代价均被设置为很小时，交叉代价聚合的表现更好。

但是，基于以上数据集的标记会出现一个矛盾。在网络训练的后阶段，由于算法对正负样本分类设计的基准视差分别是 [1,0]，也即对每一个正样本而言，网络希望它越来越逼近相似性为 1，而负样本的相似性越来越逼近 0。如前所述，本算法设计的正负样本并不是所有图像块均满足左右完全匹配或不匹配（因 o_{pos}、o_{neg} 的存在），所以训练网络接近收敛时，应该给予网络一定的“放松”。可以通过设计交叉熵损失函数来实现：

$$D_{\mathrm{loss}}=\begin{cases}t\ln(\Delta+s)+(1-t)\ln(1-s), & t=1,1-3\Delta\leqslant s\leqslant 1-\Delta\\ t\ln(s)+(1-t)\ln(\Delta+1-s), & t=0,\Delta\leqslant s\leqslant 3\Delta\\ t\ln(s)+(1-t)\ln(1-s), & \text{others}\end{cases},\tag{4}$$

式中：s 表示网络的输出；正样本时 $t=1$，负样本时 $t=0$。本文中设置 Δ 为 0.05。改进前后的损失函数曲线如图 3 所示。算法分别在正样本相似性接近 1、负样本相似性接近 0 时放宽容限。

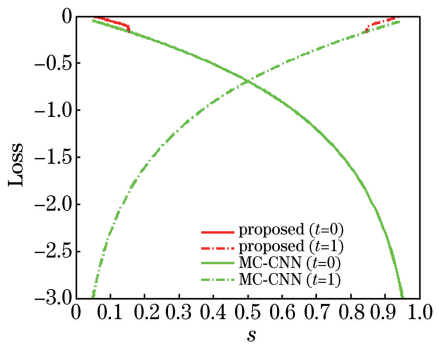

图 3　改进前后损失函数曲线图

Fig. 3　Loss function curve before and after modification

4　实　　验

分别在 KITTI2012 和 KITTI2015 双目图像数据集上对本文方法进行效果验证。KITTI2012 训练集包含了 195 对左右图像，KITTI2015 有 200 对图像。借鉴 MC-CNN 等多数深度学习方法，分别将两个数据集中的 40 对图像用于验证网络，其余图像用于训练网络。另外，当使用 KITTI2012（KITTI2015）数据集训练网络来验证 KITTI2015（KITTI2012）数据集时，所有的数据均被用作训练对象。所有的立体图像均分别通过减去平均值并除以其像素强度值的标准偏差来进行预处理。另外，使用颜色信息并不能改善视差图的质量，因此，本实验将所有彩色图像转换为灰度图像。

4.1　训　　练

采用小批量梯度下降的方式进行训练，经过一定数量的训练-验证重复实验，本文将训练时的批尺寸设置为 150、动量设为 0.9。在神经网络训练中，学习率是影响训练速度和训练精度的重要因素之一。若学习率太小，收敛性易得到保证，但收敛速度太慢；学习率太大，学习速度快，但可能导致振荡或发散。尝试经过不同参数值，本文网络训练时学习率定为 0.02，为了适应不同阶段的权值修正幅度，迭代后期逐渐减小学习率，当训练到第 18 个 epoch（1 个 epoch 就是将所有样本全部通过网络训练一次）时，损失函数接近收敛。实验比较发现，将块尺寸设为 9×9 时测试结果最好，所以后续实验结果均为基于 9×9 的块进行的实验。本文实验平台为 NVIDIA Titan X，环境为 Torch7。

4.2　实验对比

MC-CNN 网络改变前后的损失值（基于 KITTI2015 数据集）对比曲线如图 4 所示。可以看出，本文算法的初始损失值相比原始 MC-CNN 有大幅度下降。虽然 MC-CNN 在前几次 epoch 中损失值下降迅速，但本文网络无论是在迭代前期还是在后期，损失值均低于 MC-CNN 方法。

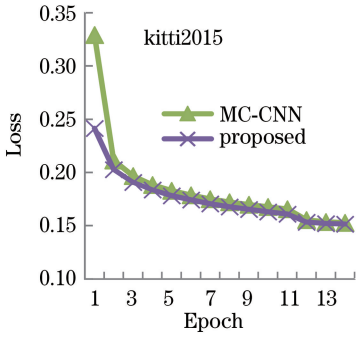

图 4　网络改变前后损失收敛曲线对比图

Fig. 4　Comparison of loss convergence curves before and after network modification

表1给出了本文算法在网络优化的前提下，改进损失函数前后的客观误差指标对比（基于 KITTI2015 数据集）。该表中指标为视差值与基准视差差距大于 3 pixel 的像素点比例。从左至右分别是训练 KITTI2012 验证 KITTI2012、训练 KITTI2015 验证 KITTI2015、训练 KITTI2012 验证 KITTI2015、训练 KITTI2015 验证 KITTI2012 的误差对比。从表中可以看出，在网络结构相同的情况下，改进损失函数后算法获得的视差图误差指标有所下降，证明了损失函数修改的有效性。表 2 给出了改进后损失函数中不同 Δ 值得到的最终视差误差对比，从表中可以看出，当 Δ 为 0.05 时，实验结果误差最小。

表 1　改进损失函数前后误差对比

Table 1　Error comparison for the improvement of loss function　%

Method	KITTI2012	KITTI2015	KITTI2012 on KITTI2015	KITTI2015 on KITTI2012
Original loss	2.63	3.28	4.03	3.92
Proposed loss	**2.61**	**3.25**	**4.02**	**3.89**

表 2　不同 Δ 值下视差误差对比

Table 2　Error comparison between different Δ values　%

Δ	Error
0.02	3.261
0.04	3.266
0.05	3.252
0.06	3.284
0.08	3.267

由于本文算法是基于原始 MC-CNN-slow 版本的改进，将本文算法单独与 MC-CNN-slow 版本的客观误差指标结果相比较，如表 3 所示。值得一提的是，原始 MC-CNN-slow 版本每对图像的运行速度为 35 s，本文算法运行时间为 28 s。结合表 3 可以看出，本文算法相比 MC-CNN-slow 版本而言，不仅在速度上约有 20%的提升，而且在主观和客观视差结果指标上没有下降，甚至有一定程度的改善。

表 3 本文算法与 MC-CNN-slow 版本的误差对比

Table 3　Error comparison between MC-CNN-slow and proposed method　%

Training set	KITTI2012		KITTI2015	
	MC-CNN-slow	Proposed	MC-CNN-slow	Proposed
KITTI2012	2.63	2.61	4.01	4.02
KITTI2015	4.32	3.89	3.27	3.25

除了与 MC-CNN 方法进行比较，本文还将改进网络的结果与双目视觉领域内经典算法（高效大规模立体匹配（Elas）[1]、半全局立体匹配（SGM）[5]、斜平面平滑（SPSS）[9]、快速 CNN 匹配[12]，以及 MC-

CNN[7]的 fast 和 slow 版本)的客观误差指标进行实验结果对比,如表 4、5 所示。表中指标为视差值与基准视差差距大于 m(m=2,3,4,5)个像素的像素点比例。从表中可以看出,本文算法的误差值在总体上均小于对比算法,具有较好的匹配效果。

表 4　各算法的视差结果误差对比(KITTI2012)

Table 4　Error comparison of disparity with different algorithms (KITTI2012)　%

Algorithm	>2 pixel	>3 pixel	>4 pixel	>5 pixel
Elas	22.72	21.07	20.23	19.66
SGM	6.28	4.98	4.14	3.57
SPSS	4.86	3.79	3.17	2.76
Fast CNN	4.98	3.07	2.39	2.03
MC-CNN-fast	4.88	3.03	2.30	1.93
MC-CNN-slow	**4.28**	.63	2.02	1.72
Proposed	4.36	**2.61**	**2.00**	**1.70**

表 5　各算法的视差结果误差对比(KITTI2015)

Table 5　Error comparison of disparity with different algorithms (KITTI2015)　%

Algorithm	>2 pixel	>3 pixel	>4 pixel	>5 pixel
Elas	24.09	19.21	17.59	16.82
SGM	10.03	6.93	5.47	4.48
SPSS	7.15	4.58	3.46	2.93
Fast CNN	6.78	4.38	2.56	2.03
MC-CNN-fast	7.53	4.01	2.84	2.33
MC-CNN-slow	**6.38**	3.27	2.37	1.97
Proposed	6.56	**3.25**	**2.33**	**1.92**

本文方法得到的视差主观结果如图 5、6 所示。其中,误差图是指计算所得的视差图与基准视差各

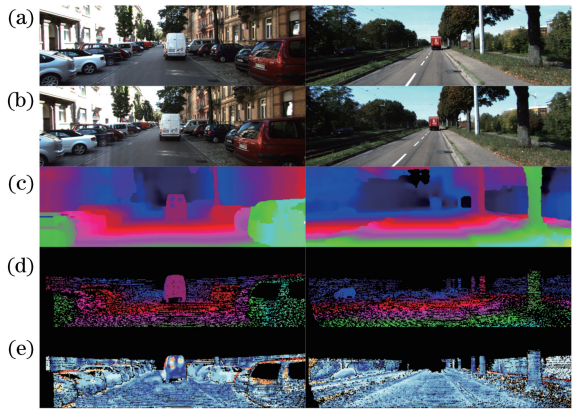

图 5　本文算法进行立体匹配的视差结果(Ⅰ)。(a)原输入左图;(b)原输入右图;(c)视差图;(d)基准视差;(e)误差图

Fig. 5　Disparity map of stereo matching obtained by proposed method (Ⅰ). (a) Original left input image; (b) original right input image; (c) disparity map; (d) ground truth; (e) error graph

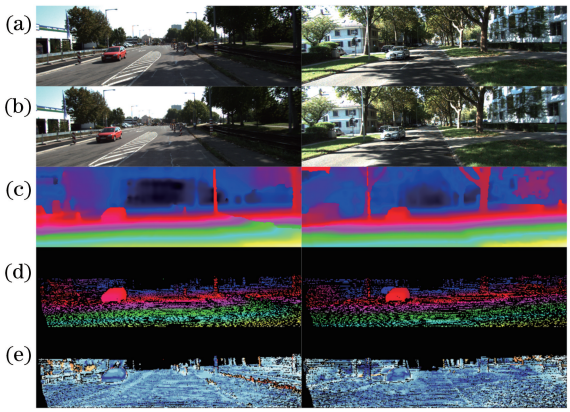

图 6　本文算法进行立体匹配的视差结果(Ⅱ)。(a)原输入左图;(b)原输入右图;(c)视差图;(d)基准视差;(e)误差图

Fig. 6　Disparity map of stereo matching obtained by proposed method (Ⅱ). (a) Original left input image; (b) original right input image; (c) disparity map; (d) ground truth; (e) error graph

个像素点的差距,以图像像素的形式展现出来。从视差图可以看出,在 4 种道路场景下,本文算法均能得到光滑稠密的视差图,特别是在目标物边缘区域,较为明显地保留了原目标的边缘信息,例如图 5 和图 6 左图中的车辆边缘、图 5 和图 6 右图中车辆、树干和电线杆等,匹配效果较好。

5　结　论

提出一种基于卷积神经网络“缩小型”结构的立体匹配方法。在 KITTI2012 和 KITTI2015 数据集上进行实验,相比经典传统算法和部分深度学习方法,本文方法立体匹配的结果在精度上有一定优势,特别是相比原始 MC-CNN 方法有较大的速度提升。但是,目前边界处理在卷积神经网络,以及其他双目视觉方法上都有一定的局限性,后期工作可能会重点关注经神经网络匹配后的视差图像,以及利用分割信息等方法对视差的后处理进行优化。

参　考　文　献

[1] Geiger A, Roser M, Urtasun R. Efficient large-scale stereo matching [C]. Tenth Asian Conference on Computer Vision, 2010: 25-38.

[2] Liu J, Zhang J X, Dai Y, *et al*. Dense stereo matching based on cross-scale guided image filtering [J]. Acta Optica Sinica, 2018, 38(1): 0115004.
刘杰, 张建勋, 代煜, 等. 基于跨尺度引导图像滤波的稠密立体匹配[J]. 光学学报, 2018, 38(1): 0115004.

[3] Ma N, Men Y B, Men C G, *et al*. A small baseline stereo matching method based on extended phase correlation[J]. Acta Electronica Sinica, 2017, 45(8): 1827-1835.
马宁, 门宇博, 门朝光, 等. 基于扩展相位相关的小基高比立体匹配方法[J]. 电子学报, 2017, 45(8): 1827-1835.

[4] Xu Y, Zhao Y, Ji M. Local stereo matching with adaptive shape support window based cost aggregation[J]. Applied Optics, 2014, 53(29): 6885-6892.

[5] Hirschmuller H. Accurate and efficient stereo processing by semi-global matching and mutual information[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2005: 807-814.

[6] Zhou F Y, Jin L P, Dong J. Review of convolutional neural network[J]. Chinese Journal of Computers, 2017, 40(6): 1229-1251.
周飞燕, 金林鹏, 董军. 卷积神经网络研究综述[J]. 计算机学报, 2017, 40(6): 1229-1251.

[7] Žbontar J, LeCun Y. Stereo matching by training a convolutional neural network to compare image patches[J]. Journal of Machine Learning Research, 2016, 17(65): 1-32.

[8] Xiao J S, Liu E Y, Zhu L, *et al*. Improved image super-resolution algorithm based on convolutional neural network[J]. Acta Optica Sinica, 2017, 37(3): 0318011.
肖进胜, 刘恩雨, 朱力, 等. 改进的基于卷积神经网络的图像超分辨率算法[J]. 光学学报, 2017, 37(3): 0318011.

[9] Yamaguchi K, McAllester D, Urtasun R. Efficient joint segmentation, occlusion labeling, stereo and flow estimation[C]. European Conference on Computer Vision, 2014: 756-771.

[10] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]// International Conference on Machine Learning, 2015: 448-456.

[11] Chu J, Gong W, Miao J, *et al*. A tree structure dynamic programming stereo matching algorithm based on linear filtering[J]. Acta Automatica Sinica, 2015, 41(11): 1941-1950.
储珺, 龚文, 缪君, 等. 基于线性滤波的树结构动态规划立体匹配算法[J]. 自动化学报, 2015, 41(11): 1941-1950.

[12] Luo W, Schwing A G, Urtasun R. Efficient deep learning for stereo matching[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 5695-5703.

[13] Sinha S N, Scharstein D, Szeliski R. Efficient high-resolution stereo matching using local plane sweeps[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 1582-1589.

[14] Zagoruyko S, Komodakis N. Learning to compare image patches via convolutional neural networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 4353-4361.

[15] Kendall A, Martirosyan H, Dasgupta S, *et al*. End-to-end learning of geometry and context for deep stereo regression[C]// IEEE International Conference on Computer Vision (ICCV), 2017: 66-75.

[16] Güney F, Geiger A. Displets: Resolving stereo ambiguities using object knowledge[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 4165-4175.

[17] Mayer N, Ilg E, Häusser P, *et al*. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 4040-4048.

[18] Dosovitskiy A, Fischer P, Ilg E, *et al*. FlowNet: Learning optical flow with convolutional networks[C]// Proceedings of the IEEE International Conference on Computer Vision, 2015: 2758-2766.

[19] Pang J, Sun W, Ren J S, *et al*. Cascade residual learning: A two-stage convolutional neural network for stereo matching[C]// International Conference on Computer Vision-Workshop on Geometry Meets Deep Learning (ICCVW 2017), 2017: 887-895.