****

**毕业论文**

**题 目：基于端到端深度学习的立体匹配研究**

**学 号 M201671791**

**姓 名 王润**

**专 业 信息与通信工程**

**指 导 教 师 杨铀**

**院（系、所） 电子信息与通信学院**

**华中科技大学研究生院制**

**填表注意事项**

一、本表适用于攻读硕士学位研究生选题报告、学术报告，攻读博士学位研究生文献综述、选题报告、论文中期进展报告、学术报告等。

二、以上各报告内容及要求由相关院（系、所）做具体要求。

三、以上各报告均须存入研究生个人学籍档案。

四、本表填写要求文句通顺、内容明确、字迹工整。

# 1 研究背景及意义

随着科学技术的发展，人们对智能化生活的需求日益增长。各个国家政府也将人工智能提升到了国家战略的层面，相继提出了各自的人工智能发展规划。人类获取的信息中有70%-80%来源于视觉[1]，而计算机视觉的目标就是使机器、计算机像人的视觉系统一样“看”世界。显然，计算机视觉是实现智能化生活的一个重要切入口。人眼天然地具有感知周围三维环境的能力，即感知距离的能力。而如何获取深度一直都是计算机视觉中的研究热点。根据在成像过程中是否主动发出能量，目前计算机视觉领域获取深度的方式可以分为两类，即主动深度传感与被动测距传感。

主动深度传感器主要包括飞行时间（TOF）、结构光和激光雷达，如微软公司推出的Kinect传感器和用于无人驾驶汽车上的LiDAR传感器。TOF相机通过发射连续的近红外脉冲，根据接收到的脉冲相位差推算出光脉冲的传输延迟，从而计算出传感器到场景的深度。结构光传感器将结构光投射至场景，由于结构光的模式图案会因为物体的形状发生形变，通过形变程度利用三角原理就可以计算得到场景中各点的深度信息[2]。但上述主动深度传感器通常存在成像分辨率低、易受噪声干扰和成像距离有限等缺点[3,4]。另一方面，虽有传感器价格在不断降低，但相对高昂的价格也限制了主动深度传感器的普及度。被动测距传感中最常用的方法就是双目立体视觉。类似人类视觉系统的左右眼，双目立体视觉技术使用两台相机在不同位置，对同一场景进行拍摄。从而物理世界的同一目标点在不同相机平面上成像，通过寻找两幅图像上的匹配点形成视差，最后利用三角测量原理将视差转化为深度。

相对于主动深度传感器，双目立体视觉有着低成本、灵活、易扩展和易实现的优势。因而双目立体视觉在诸多领域被广泛研究和使用，如三维场景重建、自动驾驶、机器人导航、工业检测和虚拟现实等，存在巨大的商业价值和军事价值。

如图1所示，双目立体视觉系统通常包含以下流程：离线相机标定、图像采集、双目校正、立体匹配和深度计算。上述五个步骤中难度最大以及最重要的环节就是立体匹配，立体匹配的好坏直接决定了最终的深度图质量[5]。立体匹配

图1. 双目立体视觉系统

技术主要是在校正后的左右视点图像对上寻找匹配点，然后根据匹配点在左右视点上的水平坐标差计算得到视差。寻找匹配点对的一个潜在假设就是同一表面在左右视点成像是相似的，而在实际成像过程中，立体匹配往往会演化成病态问题。造成立体匹配困难的原因是多方面，一方面存在如光照变化等成像噪声，另一方面，左右视点图像存在遮挡区域、弱纹理区域、纹理重复区域、透视畸变区域和视差不连续区域。虽然立体匹配算法经过多年的发展，已经取得了相当不错的进展，但是上述问题依然无法得到很好解决。近来，由于深度学习的突破性进展，深度神经网络在图像识别[6]、目标检测[7]和语义分割[8,9]等任务上都取得了目前最好的效果。基于深度神经网络的立体匹配算法也激发了研究者们广泛的研究兴趣，并展现出了十足的潜力，为后续的立体匹配算法研究也提供了另一条思路。但是，现有算法对于立体匹配中存在的病态问题依然无法能够很好地解决，立体匹配算法的精度和效率都有很大的提升空间。

综上所述，本文研究基于双目视觉的立体匹配算法具有重要的学术价值和实用价值。

# 2 国内外研究现状

立体匹配作为双目视觉中最关键的环节，到目前，来自全球的研究者们已经做出了大量的工作。2002年，Scharstein 和 Szeliski[9]对立体匹配算法进行了全面的综述调研，并将立体匹配归纳为了四个步骤：匹配代价计算、代价聚合、视差最优化计算和视差精细化。同时，这个综述报告也介绍了第一个Middlebury数据集和相应的评价指标，此数据集包含通过结构光获得的真实深度图。2012年，Geiger[10]等人将LiDAR传感器安装在汽车顶部来进行移动拍摄，发布了一个更大的数据集KITTI 2012。 Menze[11]在2015年继续扩展了这个数据集，发布了KITTI 2015。这些带有Ground Truth的数据集使得利用统计学习算法来进行立体匹配成为可能。尤其是深度学习近年来取得的突破性进展，深度神经网络凭借其强大的特征表达能力，在图像识别、目标检测和语义分割等任务上达到了目前最好的效果。相继地，研究者们利用深度神经网络来处理立体匹配问题，目前的结果也展现了此思路的潜力。同论文[12]一样，在本文中，我们根据是否利用了深度学习将现有的立体匹配算法分为两大类，即基于传统方法的立体匹配和基于深度学习的立体匹配。

## 2.1 基于传统方法的立体匹配

基于传统立体匹配算法可以按照在图像中参与计算的区域大小分为基于局部区域优化的立体匹配算法，基于全局优化的立体匹配算法和基于半全局优化的立体匹配算法。

**2.1.1 基于局部优化的立体匹配**

基于局部区域优化的立体匹配算法是通过提取图像的角点、边缘、平面等局部信息估计视差，优点是速度快，在特征区分明显的情况下效果好，缺点是视差图稀疏，鲁棒性差。基于光度测量学不变性原理区域的立体匹配算法以领域像素的灰度作为匹配基元，用相关性作为是否匹配的判断依据，可以获得相对稠密的视差图，缺点是在弱纹理场景和遮挡区域难以得到较好的视差图。代表性的方法如计算像素绝对误差和（SAD）、平方误差和（SQD）和图像块的描述相关性(NCC)[13]。Zabih 和 Woodfill[14]将非参数变换引入到区域立体变换中，通过将领域的像素灰度值与匹配点像素的灰度值比较，进行二值编码，然后再根据汉明距离计算匹配代价，可以有效改善噪声明显、亮度不均匀的情况。Yoon[15]基于同一物体周围颜色的相似性设计了一个自适应的权重学习模型来进行立体匹配。

**2.1.2 基于全局优化的立体匹配**

基于局部的立体匹配算法由于局部窗口内的特征较为有限，不能讲窗口外的信息考虑进来，并且窗口的尺寸大小难以设定，在弱纹理区域等场景下匹配效果不好。为了解决这个问题产生了基于全局优化的立体匹配算法，本质上是把对应点的匹配问题转化为寻找能量函数全局最优解的问题，将图像的每个像素都考虑 进来构建全局能量函数。如Boykov[16]采用马尔科夫随机场对视差可能的取值进行连续性约束，进而进行全局优化。Vladimir[17]提出基于图割的立体匹配算法构建网络后通过最小割方法获取网络的最大流。Taniai[18]将窗口内的图像特征提取后用图割算法进行优化计算取得很好的实验结果。基于图割的算法计算视差虽然迭代次数较高但是能去掉条纹瑕疵。Sun[19]利用置信传播算法进行视差的全局优化，对于弱纹理和遮挡区域的处理效果大为改善。

**2.1.3 基于半全局优化的立体匹配**

由于全局优化算法是对二维平面上所有像素的优化过程，时间复杂度较高，因此产生了半全局优化算法Semi-Global Matching（SGM）[20]。其核心思想是用多个一维空间的优化结果来替代二维空间的优化结果。其优化函数与全局算法的优化函数类似，具有代价项和平滑项。代价项用来描述某处像素的匹配代价，平滑项用来描述待匹配点与其周围像素视差值的差异程度。依据目标像素预测视差值与周围像素的预测视差值的差值设计不同的惩罚系数进行优化，选取待匹配像素八联通域上的八个方向上的所有像素作为搜索区域，从而将全局优化二维问题转化为八个方向上的一维最优问题，每个一维方向的视差计算采用动态规划算法。半全局优化算法具有速度快、精度高的优点，应用十分广泛。Rothermel[21]设计了由粗到细的SGM算法，将图像先缩放到低分辨率上进行SGM计算得到较为粗糙的视差图，并将其用在高分辨率的图像上进行训练获得更为精准的视差图。

## 2.2 基于深度学习的立体匹配

基于深度学习的立体匹配算法可以分为两类，将深度学习与传统方法相结合的立体匹配算法和基于端到端的整体优化的深度学习立体匹配算法。

**2.2.1 基于传统算法与深度学习结合的立体匹配**

在立体匹配算法中一个非常重要的环节就是计算匹配代价，本质上是度量两个图像块的相似度。卷积神经网络有强悍的特征表达能力，从而可以在高层特征上进行相似度计算[22]。Zbontar 和 LeCun[23,24]利用深度神经网络来计算匹配代价，利用成对9\*9的图像块组成训练数据，通过训练深度神经网络来预测图像块之间的相似性。在他们的方法中，只是匹配代价的计算由传统方法中人工设计的准则换成了由训练的网络模型来预测，后续依然采用了传统匹配方法中的代价聚合、SGM和视差精细化等步骤，但其效果超过了传统的匹配算法。研究者在此思路上做了更深的研究。Chen[25]引入多尺度的信息，将多个尺度的特征融合起来，最后通过计算左右视点图像融合特征的欧式距离得到匹配代价。Luo[26]等人提出了一个快速的网络，将计算匹配代价转化为一个多分类的问题。Shaked和Wolf[27]提出了一个高速通路的网络来计算匹配代价，利用交叉熵损失函数对两块匹配区域的相似进行判断。Seki[28]基于物体像素周围邻域内的视差具有连续性并利用左右一致性约束，在卷积神经网络中引入了基于SGM的置信度函数训练，后面再进行代价聚合与后处理获得最终的预测结果。

**2.2.1 基于端到端深度学习的立体匹配**

近来，端到端网络被设计来直接预测视差图。Mayer[29]等人在2015年提出了一个新颖的端到端的网络FlowNet来预测图像的光流，并且发布了一个大规模的合成数据集Scene Flow。后来Guney[30]将FlowNet的网络结构用在立体匹配中，称为DispNet。算法首先对一对左右视图各自通过两个网络学习抽象特征，提出一个新的关联系数层来计算左右两个像素块的相似性，通过上采样与卷积操做将特征图恢复到原图尺寸进行预测，由于不同网络层所学习到的特征不一样，浅层网络特征图学到更多是边缘等细节特征，深层网络特征图学到更多是抽象的语义特征，把深层的特征图通过双线性插值放大到与浅层特征图一样的大小，并将它们累叠起来，目的是融合浅层与深层特征。其次将真实视差标签图缩小成不同分辨率，针对不同分辨率的特征图辅助监督完成训练。Pang[31]等人扩展DispNet提出了一个级联残差学习网络（CRL），包含两个阶段的网络，分别计算视差图和对应的残多尺度残差，然后将两个阶段的输出加起来组成最后的视差图。Kendall[32]等人提出了一个端到端的网络GC-Net，根据与预先设定最大视差，考虑到场景的几何信息形成代价容器。CRL和GC-Net都利用了多尺度的特征来进行视差估计，将底层特征和高层特征组合起来，反复利用层次信息。

在语义分割中，聚合上下文信息往往对标注目标类别十分关键。语义分割与立体匹配这两个任务有着很强的相似性，都是像素到像素的优化任务。语义分割中考虑全局上下文的思想也可以用在立体匹配中。在语义分割中通常有两种做法：编码-解码结构和金字塔池化。编码-解码结构的主要思想是通过跳跃连接整合自顶而下和自下而上的信息。Long[33]等人提出全卷积网络（FCN）聚合由粗到细的预测结果以提升分割效果。不同于聚合预测结果，Ronneberger[34]等人提出U-Net，通过聚合由粗到细的特征在医学图像上达到了很好的分割结果。其后，又相继出现了SharpMask[35]、RefineNet[36]和标签精细网络[37]，其核心思想都是类似地聚合多尺度的特征。Newell[38]堆叠多个编码-解码结构，以此来提升特征融合。在深度网络中，网络的实际感受野比理论上的感受野要小很多，基于此提出了金字塔池化[39]。Liu[39]等人提出的ParseNet表明在FCN中加入全局池化能够扩大实际的感受野，从而能够在全局图像层级上提取信息，最终提升分割效果。Zhao[40]等人提出PSPNet表明金字塔池化能够收集有效的多尺度上下文信息。

DispNet和GC-Net都利用编码-解码的结构试图去聚合上下文信息，并且取得了不错的结果。

## 3研究内容与研究方案

## 3.1 研究内容

在第二节的分析中，目前立体匹配的方法主要有两大类：基于传统方法的立体匹配和基于深度学习的立体匹配。传统方法主要是通过人工设计准则来进行像素块匹配，对于立体匹配中的病态问题往往束手无策，如遮挡、弱纹理区域等。

基于深度学习的立体匹配算法达到了目前最好的效果。深度神经网络直接根据ground truth从大量的训练数据中学习匹配表达，凭借网络模型强悍的非线性表达能力，从而达到好的学习效果。

基于深度学习的算法又细分成两类，一种是利用深度神经网络来学习计算匹配代价，从而和传统算法流程相结合；另一种是直接实现端到端的立体匹配。基于传统算法与深度学习结合的立体匹配，将算法分解成多步优化的过程，往往达不到整体最优的结果。基于端到端深度学习的立体匹配，输入左右视点图，直接输出视差图，达到全局的整体优化。目前，基于端到端深度学习的立体匹配算法展现了更好的潜力。

本研究主要包括以下三个方面：

1. 基于端到端深度学习的立体匹配。端到端的学习模型能够较好地实现全局优化，从而达到好的学习效果。
2. 基于编码-解码结构和金字塔池化相结合的上下文信息学习。在基于块匹配的立体匹配算法中，往往考虑不到上下文信息。而考虑上下文信息对于解决立体匹配中的病态问题是十分有帮助的，如弱纹理区域、纹理重复区域等。本文主要考虑将语义分割中两种结合上下文信息的方法融入到立体匹配算法中。设计了编码-解码的跳跃连接结构和金字塔池化的方法，实现层次化的上下文学习。
3. 实现双目立体视觉系统。基于双目立体视觉系统的五个步骤，搭建双目立体视觉软硬件系统，实现三维空间测距功能。

## 3.2 立体匹配算法研究框架和技术方案



图2. 算法框架图

本方案的主要出发点是在立体匹配中充分考虑上下文信息。从语义分割中解决此类方法得到的两条线索，设计编码-解码的跳跃链接网络结构，加入金字塔池化模块。在GC-Net[32]的基础上设计了端到端的全卷积网络，实现输入左右视点图像，直接输出视差图。如图2所示为框架流程图。可以分为以下几个模块：一元特征提取、金字塔池化、构建匹配代价容器、编码-解码结构3D CNN和视差回归。

**3.2.1 一元特征提取**

卷积神经网络有很好的特征提取能力。首先我们学习到一个深度特征用于计算立体匹配成本。而不是使用原始的像素强度值来计算匹配成本。这种特征提取器在光照条件复杂的区域更具鲁棒性并且能够结合局部的上下文信息。   
  在我们的模型中，通过二维卷积操作来学习深度特征。每个卷积层后面都跟随一个批量规则化层（BN）和一个修正的非线性层。为了减少计算要求，我们首先使用5\*5的卷积核，以2为步长对输入进行下采样。这一层之后，我们加了8个残差组，每个残差组是由两个3\*3的卷积层链接而成。我们将左右两幅立体图像穿过这些层，得到了一元特征，特征图的尺寸是原始图像尺寸的1/4。我们在左右两个分路中共享参数以更加高效的学习到一致性特征。

**3.2.2 金字塔池化**



图3.金字塔池化

众所周知，global pooling可以非常好的作为全局信息的提取器，在分类任务中得到广泛使用。但是只有全局池化会丢失很大部分的空间位置信息，无法利用各子区域之间的关系。我们在一元特征feature map后使用了金字塔池化模块，希望通过不同大小感受野去提取更为丰富的全局信息以及各子区域的信息。在何凯明[41]的论文中，空间金字塔池化被设计用来去除CNN中的尺寸约束。由空间金字塔池化生成的不同级别的特征图被平整后送入全卷积层用于分类，在此之后金字塔池化被用于语义分割问题。我们采用包含不同尺度和次级区域信息的分层全局池化。在PSPNet[40]中，金字塔池化模块使用自适应平均池化把特征压缩到四个尺度上，并紧跟一个1\*1的卷积层来减少特征维度，之后低维度的特征图通过双线性插值的方法进行上采样以恢复到原始图片的尺寸。不同级别的特征图都结合成最终的金字塔池化特征图。

**3.2.3 构建匹配代价容器**

我们使用一元特征计算匹配代价来构架匹配代价容器。虽然采用的方法是简单地将左右特征图级联起来，但构成的匹配代价容器使得我们能够按照保留立体匹配几何信息的方式建模。对于每一幅立体图像，我们都构建了一个四维的匹配代价容器（高度、宽度、最大视差+1（0～maxdisp）、特征图数目）。我们将左图的每一个一元特征和右图每一个视差下的特征图级联起来，并封装成一个四维的代价容器。（对于某一个特征，匹配代价容器就是一个三维的方块，第一层是视差为0时的特征图，第二层是视差为1时的特征图，以此类推共有最大视差+1层，长和宽分别是特征图的尺寸，假设一共提取了10个特征，则有10个这样的三维方块） 至关重要的是，我们在这步操作中保留了特征的维度，不像先前许多方法使用点乘方式来降低特征维度。这使得我们可以结合上下文信息并作用于一元特征上。我们发现通过级联特征图构成匹配代容器的效果要优于削减特征或者使用距离度量函数的方法。我们的直觉是通过保持一元特征，网络能够学习到一个绝对的表征（因为不是去度量距离这种间接的方式）并把它带入到匹配代价卷中。这赋予了该结构学习语义信息的能力。相反，使用距离度量方法限制了网络只能学习特征之间的相关性表征，而不能将绝对的特征表征带入到匹配代价容器中。

**3.2.4 编码-解码结构3D CNN**

为了得到更多的上下文信息，我们使用编码-解码结构，通过跳跃连接来实现由精到粗再由粗到精的过程中的信息聚合。经过4层3D卷积下采样， 然后对应的3D反卷积。最终的结果是得到了一个规则化的尺寸为H\*W\*D的匹配代价容器。

图4. 编码-解码结构3D CNN

**3.2.5 视差回归**

通过将一元特征匹配代价转为最后维度为H\*W\*D的匹配代价容器。根据此匹配代价容器，通过遍历每个视差维度，使整体的匹配代价达到最小，通过此最小化操作得到最终的视差图。然而，这个操作会造成两个问题：1、无法达到亚像素精度的视差估计。2、此最小化操作不可导，从而不能进行反向传播训练。

为了解决上述问题，我们采用GC-Net中使用软最小化方法，此方法不仅可导，而且可以通过回归方式得到平滑的视差估计。根据由softmax操作得到预测代价 来计算每一个视差值d的可能性。预测视差值d'由每一个视差值\*其对应的可能性求和得到，如下式：



其中 对应的是softmax操作。

## 3.3 双目立体视觉系统

如图5所示，由双目立体视觉模型可以得到：



场景深度与视差成反比，根据焦距f和视差d就可以求得深度，而焦距可以通过相机标定得到，视差通过立体匹配得到。根据双目立体视觉模型，我们搭建了双目立体视觉软硬件系统。

**3.2.1 双目立体视觉硬件系统**

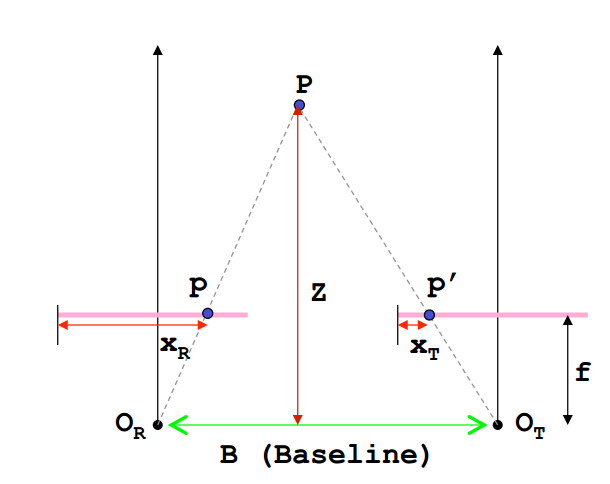


图5.双目立体视觉模型

硬件系统由两个相同的USB相机构成，组装在实验室的光学成像平台上。相机分辨率为1920\*1080，双目相机基线距离为60mm。

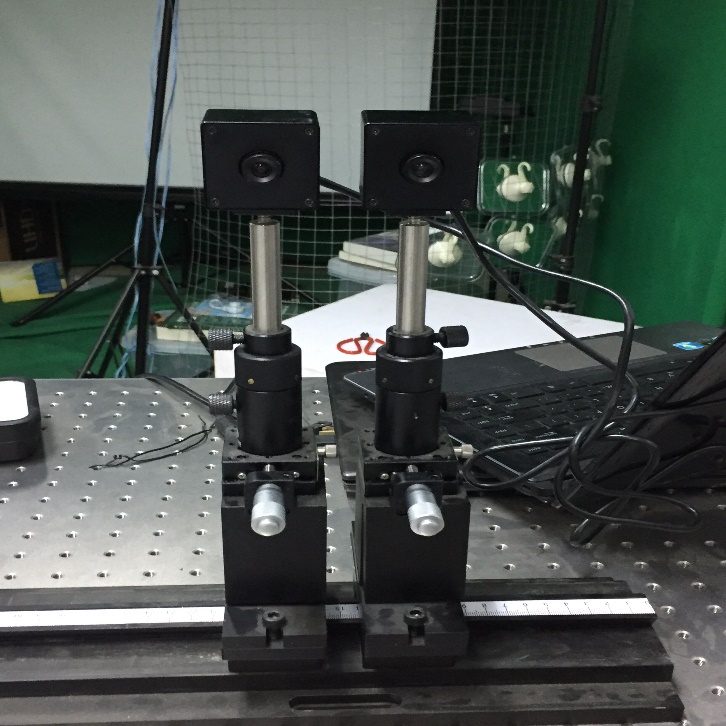


图6.双目立体相机

相机标定不是本文研究重点，我们利用成熟的张氏标定法对USB相机进行标定。得到相机内外参如下，



双目标定结果如下：

旋转矩阵：



平移矩阵：



由此得到相机的相对焦距。

**3.2.1 双目立体视觉软件系统**

如图7所示，软件系统主要包含三个模块：立体匹配模块、距离测量模块、和结果显示模块。



图7.双目立体视觉软件模块

|  |  |  |
| --- | --- | --- |
| （a） | （b） | （c） |

图8. （a）左视点图，（b）右视点图，（c）视差图

## 4 研究工具及性能参数评估

本实验在window 10 64位和Ubuntu 16.04 开发环境下，利用c++、Python和matlab语言和OpenCV、PyTorch等工具实现。

实验所用的数据库包括KITTI 2012[10]和KITTI 2015[11]。在KITTI 2012中，训练集包含194对立体图像，测试集包含195对立体图像。训练集中提供了通过LiDAR传感器得到的真实视差图。图像分辨率为1226\*370。在KITTI 2015中，训练集包含200对立体图像，测试集也包含200对立体图像。训练集中同样提供了通过LiDAR传感器得到的真实视差图。图像分辨率为1242\*375。



图9. KITTI 2015数据集部分样本图像

对于估计得到的视差与ground truth范围之差小于3个像素（<3px）就被视为匹配正确的点，反之则视为匹配错误。对于KITTI 2012和KITTI2015数据集，都以左视图为基准生成视差[10,11]。所以立体匹配算法的评价指标即为：

匹配错误率=总匹配错误的相素点数 / 总像素点数

错误率越低，算法性能越好。

# 5 研究安排与预期成果

1. 2018.11 - 2018.12，继续查找和阅读与课题相关的论文、文献等，寻找更多与课题相关的思路。实现并优化本文提出的立体匹配算法。

2. 2018.12-2019.1，进行更深入的思考，在本文提出的方案上做出进一步的改进和创新，完成双目立体视觉软件系统，编写软件界面。

3. 2019.1-2018.3，对深度网络模型做进一步的优化，考虑与马尔科夫随机场进行联合优化，进行实验验证。

4. 2019.3-2019.4，撰写毕业论文，完成文档整理，毕业答辩。

# 6 参考文献

[1] 章毓晋. 图像工程[M]. 清华大学出版社，2005.

[2] [Richardt](https://onlinelibrary.wiley.com/action/doSearch?ContribAuthorStored=Richardt%2C+Christian) C, [Stoll](https://onlinelibrary.wiley.com/action/doSearch?ContribAuthorStored=Stoll%2C+Carsten) C, [Dodgson](https://onlinelibrary.wiley.com/action/doSearch?ContribAuthorStored=Dodgson%2C+Neil+A) N A, et al. Coherent Spatiotemporal Filtering, Upsampling and Rendering of RGBZ Videos[C]. Computer Graphics Forum,2012.

[3] Zhang Z. Microsoft Kinect Sensor and Its Effect[J]. IEEE Transactions on Multimedia,2012,19(2):4-10

[4] [Khoshelham](https://www.mdpi.com/search?authors=Kourosh%20Khoshelham&orcid=) K , [Elberink](https://www.mdpi.com/search?authors=Sander%20Oude%20Elberink&orcid=) S O. Accuracy and Resolution of Kinect Depth Data for Indoor Mapping Applications[J]. Sensor,2012,12(2):1437-1454.

[5] Barnard S T and Fischler M A. Computational stereo[J]. ACM Computing Surveys, 1982, 14(4):553–572.

[5] Krizhevsky A, Sutskever I, and Hinton G E. Imagenet classification with deep convolutional neural networks[C]. In Advances in neural information processing systems, 2012, pages 1097–1105.

[6] Girshick R, Donahue J, Darrell T, and Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. In Proceedings of the IEEE conference on computer vision and pattern recognition, 2014., pages 580–587.

[7] Long J, Shelhamer E, and Darrell T. Fully convolutional networks for semantic segmentation[C]. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pages 3431–3440.

[8] [Chen](https://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22First%20Name%22:%22Liang-Chieh%22&searchWithin=%22Last%20Name%22:%22Chen%22&newsearch=true&sortType=newest)L C ,[Papandreou](https://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22First%20Name%22:%22George%22&searchWithin=%22Last%20Name%22:%22Papandreou%22&newsearch=true&sortType=newest)G, [Kokkinos](https://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22First%20Name%22:%22Iasonas%22&searchWithin=%22Last%20Name%22:%22Kokkinos%22&newsearch=true&sortType=newest), [Murphy](https://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22First%20Name%22:%22Kevin%22&searchWithin=%22Last%20Name%22:%22Murphy%22&newsearch=true&sortType=newest)K, [Yuille](https://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22First%20Name%22:%22Alan%20L.%22&searchWithin=%22Last%20Name%22:%22Yuille%22&newsearch=true&sortType=newest) A L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs[J]. [IEEE Transactions on Pattern Analysis and Machine Intelligence](https://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=34), 2018,40(4):834-848.

[9] Scharstein D and Szeliski R. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms[J]. International Journal of Computer Vision, 2002, 47(1):7–42.

[10] Geiger A, Lenz P, and Urtasun R. Are we ready for autonomous driving? the kitti vision benchmark suite[C]. In Conference on Computer Vision and Pattern Recognition , 2012.

[11] Menze M and Geiger A. Object scene flow for autonomous vehicles[C]. In Conference on Computer Vision and Pattern Recognition , 2015.

[12] 林森，殷新勇，唐延东. 双目视觉立体匹配技术研究现状和展望[J]. 2017, 17(3): 1671-1815.

[13] Hirschmüller H, Innocent P R, Garibaldi J. Real-Time Correlation-Based Stereo Vision with Reduced Border Errors[J].2002, [International Journal of Computer Vision](https://link.springer.com/journal/11263),60(2):91-110.

[14] Zabih R and Woodfill J. Non-parametric Local Transforms for Computing Visual Correspondence[C]. In Proceedings of European Conference on Computer Vision, 1994:151–158.

[15] [Yoon](https://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22Authors%22:%22Kuk-Jin%20Yoon%22&newsearch=true&sortType=newest)K J, [Kweon](https://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22Authors%22:%22In-So%20Kweon%22&newsearch=true&sortType=newest) I S. Locally adaptive support-weight approach for visual correspondence search[C]. [IEEE Computer Society Conference on Computer Vision and Pattern Recognition](https://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=9901) , 2015:648-655.

[16] [Boykov Y](https://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22First%20Name%22:%22Y.%22&searchWithin=%22Last%20Name%22:%22Boykov%22&newsearch=true&sortType=newest), [Veksler](https://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22First%20Name%22:%22O.%22&searchWithin=%22Last%20Name%22:%22Veksler%22&newsearch=true&sortType=newest)O, [Zabih](https://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22First%20Name%22:%22R.%22&searchWithin=%22Last%20Name%22:%22Zabih%22&newsearch=true&sortType=newest) R. Markov random fields with efficient approximations[C]. [IEEE Computer Society Conference on Computer Vision and Pattern Recognition](https://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=5649), 1998:648-655.

[17] Kolmogorov V, Zabih R. Multi-camera Scene Reconstruction via Graph Cuts[C]. [European Conference on Computer Vision](https://link.springer.com/conference/eccv), 2002:82-96.

[18] Taniai T, Matsushita Y, Sato Y ,Naemura T. Continuous 3D Label Stereo Matching using Local Expansion Moves[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018,40(11): 2725—2739.

[19] [Sun J](https://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22Authors%22:%22Jian%20Sun%22&newsearch=true&sortType=newest), Zheng N N, [Shum](https://ieeexplore.ieee.org/search/searchresult.jsp?searchWithin=%22Authors%22:%22Heung-Yeung%20Shum%22&newsearch=true&sortType=newest) H Y. Stereo matching using belief propagation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(7):787-800.

[20] Hirschmuller H. Stereo processing by semiglobal matching and mutual information [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008,30(2):328–341.

[21] [Rothermel](https://e.glgoo.top/citations?user=Ef8jOboAAAAJ&hl=zh-CN&oi=sra) M, [Wenzel](https://e.glgoo.top/citations?user=wQ_asB0AAAAJ&hl=zh-CN&oi=sra) K, [Fritsch](https://e.glgoo.top/citations?user=ylnCdtIAAAAJ&hl=zh-CN&oi=sra) D.  [SURE: Photogrammetric surface reconstruction from imagery](http://www.ifp.uni-stuttgart.de/publications/2012/Rothermel_etal_lc3d.pdf)[C]. Proceedings LC3D workshop,2012:8-29.

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

[23] Zbontar J and LeCun Y. Computing the stereo matching cost with a convolutional neural network[C]. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2015:1592–1599.

[24]Zbontar J and LeCun Y. Stereo matching by training a convolutional neural network to compare image patches[J]. Journal of Machine Learning Research, 2016 17(1-32):2.

[25] Luo W, Schwing A G, and Urtasun R. Efficient Deep Learning for Stereo Matching[C]. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition , 2016:5695-5703.

[26] Chen Z, Sun X, Wang L, Yu Y, and Huang C. A deep visual correspondence embedding model for stereo matching costs[C]. In Proceedings of the IEEE International Conference on Computer Vision, 2016:972–980.

[27] Shaked A and Wolf L . Improved Stereo Matching with Constant Highway Networks and Reflective Confidence Learning[C]. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2017:35-46.

[28] Seki A and Pollefeys M. Patch Based Confidence Prediction for Dense Disparity Map[C]. British Machine Vision Conference, 2016.

[29] Mayer N, Ilg E, Hausser P, Fischer P, Cremers D, Dosovitskiy A, and Brox T. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation[C]. In The IEEE Conference on Computer Vision and Pattern Recognition, 2016:4040-4048.

[30] Guney F and Geiger A. Displets: Resolving stereo ambiguities using object knowledge[C]. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015:4165–4175.

[31] Pang J, Sun W, Ren S Q, Yang C, and Yan Q. Cascade residual learning: A two-stage convolutional neural network for stereo matching[C]. In The IEEE International Conference on Computer Vision, 2017:878-886.

[32] Kendall A, Martirosyan H, Dasgupta S, Henry P, Kennedy R, et al. End-to-end learning of geometry and context for deep stereo regression[C]. In The IEEE International Conference on Computer Vision, 2017:66-75.

[33]Long J, Shelhamer E, and Darrell T. Fully convolutional networks for semantic segmentation[C]. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015:3431–3440.

[34] Ronneberger O, Fischer P, and Brox T. U-Net: Convolutional networks for biomedical image segmentation[C]. In International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015:234–241.

[35] Pinheiro P O, Lin T Y, Collobert R, and Dollar P. Learning to refine object segments[C]. In European Conference on Computer Vision, 2016:75–91.

[36] Lin G, Milan A, Shen C, and Reid I. RefineNet: Multipath refinement networks for high-resolution semantic segmentation[C]. In The IEEE Conference on Computer Vision and Pattern Recognition, 2017.

[37] Islam M A, Naha S, Rochan M, Bruce N, and Wang Y.Label refinement network for coarse-to-fine semantic segmentation. arXiv preprint arXiv:1703.00551, 2017.

[38] Newell A, Yang K, and Deng J. Stacked hourglass networks for human pose estimation[C]. In European Conference on Computer Vision, 2016:483–499.

[39] Liu W, Rabinovich A, and Berg A C. ParseNet: Looking wider to see better. arXiv preprint arXiv :1506.04579, 2015.

[40] Zhao H, Shi J, Qi X, Wang X, and Jia J. Pyramid scene parsing network. In The IEEE Conference on Computer Vision and Pattern Recognition, 2017.

[41] He K, Zhang X, Ren S, and Sun J. Spatial pyramid pooling in deep convolutional networks for visual recognition[C]. In European Conference on Computer Vision, 2014:346–361.

|  |  |  |
| --- | --- | --- |
| **研究生签字**  **指导教师签字**  **院(系、所)领导签字** |  |  |

**年 月 日**