什么可以移动？ 在3D激光数据中寻找汽车，行人和自行车手

摘要 - 本文解决了从城市场景的三维激光扫描中分割出来的问题。特别是，我们希望检测自动驾驶应用中感兴趣的类别的实例 - 汽车，行人和自行车 - 在显着的背景混乱中。我们的目标是提供一个端到端的流水线布局，当这些流水线由三维数据的原始流提供时，会产生不同的点组，这些点可以输送到下游分类器进行分类。我们假设，对于本文所考虑的特定类，解决分类任务（即，首先将数据分离成前景和背景）优于直接解决多类问题的方法。这是通过收集城市街景的习惯和第三方数据集来确定的。虽然我们的系统对部署的特定聚类算法是不知道的，但是我们探索了使用欧几里德最小生成树来进行端到端分割流水线，并设计了基于RANSAC的边缘选择标准。

介绍

在本文中，我们提出了一个端到端的系统，可以从三维激光数据的原始数据流中检测汽车，行人和骑行者的实例，从而实现自动驾驶应用。自主驾驶已经成为机器人研究的一个重要应用领域。这一领域的出版物聚集在一起[1]，[2]，[3]。 DARPA Grand [4]和Urban Challenges [5]的成功，以及Google为促进数据收集目的而推动自动驾驶的努力[6]，都提高了人们对自动驾驶汽车能够在复杂的现实环境中运行的期望。我们的社区渴望创造自动驾驶汽车，这进一步突出了环境理解的重要性，并将重点放在了这方面。大量的研究工作正在被用于检测和分类与驾驶真实道路环境有关的物体，这些物体都使用视觉和激光数据。特别感兴趣的是潜在的动态对象 - 即可能移动的对象 - 因为它们的存在和潜在的状态变化将影响动作和轨迹的规划。这里介绍的工作也属于这一类。特别是，我们仅限于从通常部署在自主车辆上的传感器获得的3D激光数据流中检测汽车，行人和骑车人，基于每帧的形状信息。采取基于形状的方法，因为潜在移动的物体可能实际上并不移动。

然而，获得这样的分割被广泛认为是一个难题[7]，[8]，[9]，因为场景中的对象数量通常是未知的，只有一小部分数据包含相关的类信息。 这激发了这里提出的工作。

这项工作的目的是将原始数据流的显著的子集分组成与感兴趣对象相对应的连续和完整的实体，而不必事先知道存在的对象的数量和位置。由于我们已经了解了目标类别（而且它们的集合相对较小），所以我们采用了监督方法。我们研究基于图的技术在这个问题上的应用，并且确定，对于这个工作中考虑的特定类，解决一个二元分类任务（即将数据分离成前景和背景）首先要比解决多类问题直。我们认为这是本文的主要贡献。此外，为了提供布局的端到端流水线，我们演示了使用特定的基于图的聚类算法作为我们分割方法的后端（参见图1的端到端输出的典型示例系统）。在下一节对相关工作进行调查之后，我们将在第三节介绍在本文中使用的基于图的聚类算法。第IV节详细介绍了从原始3D激光点流提取前景数据的一些方案。我们在第五部分评估这些方案，并在第六部分结束。

相关工作：

现有的三维激光数据中的物体检测和识别工作可粗略地分为三类

。第一种情况通常假设代表整个对象的点云已经被分割出来，因此主要集中在分类上。例子包括Teichman等人[7]，[10]，他们将完整的分割对象轨迹划分为汽车，行人，骑自行车者或背景之一。赖等。 [8]结合形状和外观的优点与Kinect风格的传感器分类室内对象使用稀疏距离度量学习与集团套索regulariser。在这种情况下，分割任务由对象所处的受控环境来促进。Endres et al。 [9]另一方面采取无监督的方法来发现潜在的狄利克雷分配（LDA）呈现的段中的对象类别。在3D激光数据中从经常大量的背景杂波中分离出期望的物体是这种系统的关键前兆。现有的作品包括Douillard et al。 [11]，其中假定存在一个地平面，而物体段是以非监督的方式从非地面数据中导出的。 Klasing等人[12]基于单个激光点之间的欧几里得距离进行聚类，隐含地假设物体没有被场景混乱连接。

第二类方法将场景直接标记到属于对象类（可能具有背景类）的区域中，但不区分单独的对象实例。 Anguelov等。 例如，采用基于马尔科夫随机场（MRF）的监督方法，使用在各个数据点处计算的局部特征来产生全局一致的标签。 Triebel等人 [14]采用基于在特征空间和欧几里得空间中构造的条件随机场（CRF）的方法来获得通常对应于重复模式的对象类别的场景分割。

第三类方法侧重于数据的有针对性的分割。在这里感兴趣的类是已知的，并且设计了一个分割方案，以适应这个特定的情况。一个例子是Spinello等人的工作。 [15]，其中集中明确检测行人。我们在这里介绍的方法填充上面的第二和第三类之间的空间。虽然我们也事先了解感兴趣的类别，但是我们的工作旨在满足一系列类别（即汽车，行人和骑自行车者）的需求，从而牺牲了相对较窄的细分问题的好处。为了达到这个目的，我们使用与[14]中使用的相同的预分割算法和相同的描述符。然而，与[14]相反，我们的监督方法产生的对象群只对应于感兴趣的对象类别。另外，我们系统的输出区分对象实例，而不是将场景分割成一般属于对象类的区域。

最后，我们提到一个相关的工作组，可以检测即时动态物体，即在检测时正在移动的物体（参见Katz等[16]或Yang和Wang [17]）。 与这些作品相反，这里所讨论的问题包括检测和分类可能移动但在记录数据时可能不移动的实体。

基于图形的聚类

在单个场景中，通常存在未知数量的感兴趣对象。 这些对象的成功分类要求能够区分数据流内的单独对象实例，即使在clutterfree条件下（例如在去除背景之后，参见第IV节）。 在本节中，我们将这个问题作为一个聚类任务来制定。 无监督的数据聚类已经成为数十年来的一个活跃的研究领域，并且存在许多避免缺少先验信息的方法，例如存在的聚类数量。 例如，变分贝叶斯方法[18]提供了一个有吸引力的机制，但常常受到融合问题的困扰。 Jenssen等人 [19]使用信息理论测量模型选择从各种可能性中确定最佳数量的聚类。

另一种流行的方法是使用从数据构建的欧几里德最小生成树（EMST）进行基于图的聚类[20]。 基于EMST的技术早在20世纪70年代就已经在文献中出现[21]，并且在预计群集边界不规则时经常使用。 给定一个有限集合P⊂Rd基于EMST的算法首先计算整个图G =（V，E）上的最小生成树，其中V = {i：pi∈P}，E = {{i，j} ∈P，pj∈P，i 6 = j}，边的权重由成对的欧式距离给出。 在整个树中收集的边缘统计信息用于确定在哪里断开链接。 例如，文献[22]表明，通过删除EMST中K-1个最长边，可以得到一个聚类，这个聚类最大化了K点组中所有可能的不相交分区的空间中的最小簇间距离。 当K未知时，就像我们的情况一样，启发式用来确定删除哪条边。

例如，Zahn [21]使用MST中边缘权值的本地统计量来定义不一致性度量，并删除违反任何一个一致性标准的边。 Grygorash等人 [23]提出了一系列边缘去除操作，使得边缘权重的标准偏差最小化。 当达到（局部）最小值时，找到最佳的簇数。 我们的方法还利用EMST。 具体而言，我们观察到，由于生成树的形成过程，相同对象实例的连接点的边缘倾向于具有与传感器的样本宽度相对应的相似长度（直到传感器噪声）。 连接单个对象实例的边缘往往比较长。

我们通过使用RANSAC范式[24]来利用这个观察来估计边权重中的异常值。 生成树在任何地方都被破坏 生成树在任何地方都被破坏发现异常值。为了说明，考虑图2，其中在包含汽车，两名行人和一名骑车者（来自真实数据集的所有示例）的合成场景上执行聚类。 在后面的章节中，我们将这种聚类算法称为EMST-RANSAC算法。 它可以用来将点云分割成多个实体，而不需要事先知道场景中包含的对象的数量。 然而，由于该算法基于欧几里德距离对数据进行聚类，在算法可以被应用之前，必须去除属于背景 杂波的不需要的点（即，除了感兴趣的对象类别以外的任何其他点）。 在下面的章节中，我们将这种背景杂波的去除作为一个有监督的分类任务。

IV。前景提取

考虑三维激光扫描仪产生的一组点P⊂R3。为了应用上一节中描述的EMST-RANSAC算法，我们需要将P分解成前景数据集Pf⊆P - 即属于感兴趣对象类的属性及其补码，背景数据集Pb = P-Pf的。

我们采用自下而上的方法，首先对点云进行预处理，以一组点云图像的形式进行过分割。虽然我们并不要求分割是完美的，但每个细分市场都有必要跨越一个单一的兴趣类别。然后为每个补丁提取功能。这个表示被用于每个补丁的分类，作为其Pf的成员资格。

1. 预处理

与其他作品一样，我们基于点云中的点法线估计执行现成的预分割步骤，以获得一组超级体素作为我们的实体分割方法的原子输入（参见第IVB节） 。为了获得对分区算法的输入的可靠的正常估计，我们遵循普通计算的常用方法，其针对每个基准pi找到搜索半径r内的最近邻居的局部集合，然后假定局部平面性，对其执行PCA。对应于最小特征值的特征向量被认为是估计的法线方向，因此基本上对最邻近的区域执行最小二乘平面特征。

有关pda的原理：

PCA的思想是将n维特征映射到k维上（k<n），这k维是全新的正交特征。这k维特征称为主成分，是重新构造出来的k维特征，而不是简单地从n维特征中去除其余n-k维特征。PCA将n个特征降维到k个，可以用来进行数据压缩，例如100维的向量最后可以用10维来表示，那么压缩率为90%。同样图像处理领域的KL变换使用PCA做图像压缩，人脸检测和匹配。

<http://blog.csdn.net/zhongkelee/article/details/44064401>

最大方差和最小乘法：

这种方法已经在经验上显示出在鲁棒性和计算开销之间的权衡方面表现最好[25]。 超体素分割所需的边集由N个最近邻居链接给出。 那就是E = {{i，j}：pi∈P，pj∈NN}，（1）其中NNi表示点pi的N个最近邻点的集合，排除点本身。

1. 片段分割

为了获得最初的片段分割，我们遵循Triebel等人提出的方法。 对Felzenszwalb和Huttenlocher [26]介绍的流行分割算法进行了改进，以对P中的点的正常估计进行操作。该算法对无向图G =（V，E）进行操作，边权重表示不相似性度量在相邻点之间。 从每个顶点i∈V开始，作为一个单独的片段，算法按照非递减权重的顺序遍历边，在没有边界证据时合并相邻片段。 考虑如（1）中的顶点集合V = {i：pi∈P}和边集合E. 不相似度量由

w（i，j）= 1- | ni·nj |定义 ，（2）

其中ni表示在点pi估计的正常值。 光滑表面之间的交叉点将因此产生分割边界。

1. 特征提取

对于每个补丁来说，一个固定的特征向量是通过连接五组共同的不变描述符来构造的。 描述符包括在质心和垂直Z轴上计算的50维自旋图像[27]，使用成对欧几里得距离作为形状函数的32维形状分布[28]，使用点积之间的幅度的32维形状分布 在点对的法线作为形状函数，三维形状因子[29]，沿着PCA方向的边界框的三维。 这些产生了120维的特征向量。

1. 补丁分类

对于前景和背景的分离，可以采用任何一种分类框架。 前景课程简单地构成了汽车，步行和自行车课程的联合。 我们提出了两种在贴片分类阶段合并这三个前景类的方案，以产生一个干净的场景前景 - 背景分割，从而使得EMST-RANSAC算法是适用的。

F / B二进制：在这个方案中，三个前景类汇集成一个类，二进制分类器被训练来分离它们。 F / B N级：在这里，分别用于汽车，行人，自行车和背景类别的N个一元二分类分类器。 分类完成后，三个前景类的输出被合并成一个单一的集合。

实验结果

我们使用公开可用的数据集以及使用我们自己的自主车辆收集的数据来评估我们的分割方法。 特别是，我们利用[7]向公众发布的斯坦福赛道集（STC）数据集。 STC包含大量被标注的感兴趣对象（汽车，行人和自行车），并且还有一个优点，即使用同样的传感器集中在汽车上。但是，这个数据库原本是用于轨道分类的，因此只包含可追踪对象的实例。 场景混乱尤其不足（见表1）。 因此，为此，我们使用配备了Velodyne HDL-64E SE2激光测距仪的Bowler Wildcat研究平台（图3）收集的数据来增强STC。

图

显示F / B N级方案结果的示例场景框架。 （a）地面真相现场标签。 （b）F / B N级检测到的物体。 （c）场景中的地面真实物体。 （d）由F / B N级产生的对象群。 在（a）和（b）中，红色，蓝色，绿色和灰色区域分别对应于属于汽车，行人，自行车和背景类别的点。 在（c）和（d）中，不同的颜色表示不同的对象实例，随机选择颜色。 这个数据最好用颜色来看。

A

为了分类，我们采用支持向量机（SVM）分类器与非线性径向基函数（RBF）内核。参数是用交叉验证来训练的。对于方案F / B二进制，单个二进制SVM分类器被训练用于前景和背景类别。对于F / B N级和N级方案，。四个独立的支持向量机分类器在每个车，步行者，自行车和背景类中进行一对一的配置训练

N级方案的背景对所有分类器的性能与为背景类评估的二进制前景/背景分类器的性能相同。这表明，车，步行者，自行车班级的分离分类在这些类别之间引入了重大的混淆，通过将它们整理成单一的前景类别来弥补。关于这一点的进一步的证据可以在图5所示的N级方案的混淆矩阵中找到。这意味着前景和背景之间最大的混淆是由背景数据被错误地分类为汽车造成的。另一方面，个别的前景课也存在重大的混乱。这些结果表明，为了在三维激光数据中将车辆，行人和骑行者从背景中分离出来，这里采用的主要形状特征不是足够的。这为直观的概念提供了进一步的支持，即过度分割的补丁没有足够的形状信息来正确分类。另一方面，将任务作为一个二元分类问题来解决，因为前景和背景类在形状特征的表征中显得更加适合分离。

B.整体系统

评估在本节中，我们评估整个系统的性能，从原始数据流作为输入，执行预处理，补丁分类和EMST-RANSAC集群。然而，为了证明基于前景/背景的方案对N级方案的效率，获得的聚类必须被分类到三个前景对象类之一中。为此，我们使用与在F / B二进制方案返回的整个实体集群（而不是代表对象部分的补丁）相同的第IV-C节中列出的相同特征集来训练一个N类SVM分类器。对于F / B N级方案，由于它在中间阶段使用了一个N级分类器，因此我们保留了来自分块分类的预测分数，并通 过构成激光点之间的多数表决来确定实体分类的类别。票数是通过预测的确定来加权的。对于N级方案，通过独立运行EMST-RANSAC算法得到检测结果，实体集群类别是不言而喻的。但是，我们强调分类不是本文的重点。存在更复杂的分类方法（例如，参见[7]）。

为了进行评估，我们将在360°旋转范围内返回的数据点整理成单个场景（帧）进行处理。 EMST-RANSAC算法只涉及一个参数：内点支持宽度w来评估假设[24]。 该参数被训练用于从STC数据集中提取的200个帧，这些帧与用于产生用于片分类的训练数据的那些帧不相交。 对于F / B二进制和F / B N类方案，由于EMST-RANSAC算法仅在属于前景类的补丁上应用一次，因此确定了单个值w。 对于N级方案，将EMST-RANSAC应用于每个对象类，得到三元素向量w。 我们为N级方案独立训练了三个支撑宽度。

为了评估系统的性能，我们从繁忙的城市手动标记了100个随机选择的帧场面采取在当地镇中心。 这些数据完全独立于任何培训阶段使用的数据。 图6显示了F / B N级方案的样本框架的一组定性结果.F / B二进制方案的相似结果如图1所示。对于定量分析，我们采用导出的评估指标 从视觉社区中流行的对象检测挑战中使用的那些PASCAL可视对象类挑战[30]。 特别是，如果一个检测与一个超过50％的地面实况标注重叠，则标记为正确。 重叠度量计算如下