

指导教师： 杨涛

提交时间： 2016.3.19

# CVPR2015 Paper Translation

No: 01

姓名： 王羽娟

学号： 2013302468

班号： 10011301



## 学习特定的描述符 3D 关键点检测

### 摘要

关键点检测代表大多数自动基于已建立的局部描述符之间一致的现代计算机视觉管道的第一阶段。然而，由于 3D 数据，例如点云或网状物，在细节和噪声上显示出高可变性，标准的结果还没有出现。更关键的是，现有的关于 3D 关键点检测的建议是依赖于企图达到最大限度的可重复性的几何凸函数，而不是选定区域的特殊性，后者结果可能导致所有管道性能达不到最优。为了克服这些缺点，我们把 3D 关键点检测转换为点之间的二元分类，考察这些点能不能与预定义的 3D 描述正确匹配，从而了解具体描述探测器能否完全适应不同的场景。通过针对几个公共数据集的实验，我们发现这种新的关键点检测器设计方法是一种灵活的解决方案，同时可以提供最先进的描述符匹配性能。

### 1. 介绍

当处理图像和 3D 数据时，对于可重复的和特殊的关键点的检测是现代计算机视觉的基本任务。图像的关键点检测应用于图像检索，物体检测和

识别，物体和相机跟踪，相机定标以及图像配准等等。3D 数据中的关键点常用于处理几种图形匹配任务中，例如点云配准，3D 物体识别和姿态估计，形状检索和分类。

2D 和 3D 计算机视觉处理方法包括根据人工确定的局部显著性函数的最大值定义关键点[25, 13, 24]。应用最广泛的 3D 关键点检测方法基于此类函数，可以分为固定尺度的处理方法和自适应尺度的处理方法。其中一些方法显示出良好的可重复性，以及识别可能被 3D 检测器正确匹配的区域的能力[24]。例如，在固定尺度的方法中，本质形状特征提取(ISS)[29]是一种应用广泛的又快又有效的方案；这种固定尺度探测器由 Mian 等人[15]提出，是一种比较慢的选择，点密度差异的鲁棒性好。在自适应尺度探测器中，MeshDoG[28]是一个流行的关于定义在多种由网格逼近的标量函数上的高斯差分探测器[14]的推广；自适应尺度变体在通过缩放适应性地定义点的邻域范围大小取凸函数的最大值中被提出[15]；[4]这也是为了推广 DoG 对于网格的操作。

然而，当初初始化这样的探测器时，特点和之后要计算出的描述器数量无

关。事实上，探测器通常被设计用来取重复量的最大值，而不是区域的不同点，因为这可能会导致匹配性能达不到最佳而使得到的值比正确值小。然而，辨别正确结果的能力是所有探测器/描述器进程最终的目标。在探测阶段，直接从事这个目标的重要性和用途已经显现出来。在视频追踪中，通过 Shi-Tomasi[20]最著名的角点检测，也就是将要检测的“物体”特征定义成可能得到相邻帧之间正确匹配的特征。此外，在 3D 视觉的领域里，不同的数据采集模式提供完全不一样的细节和噪声级别，这可能需要仔细调整探测器的参数，以通过不同数据集获取有用的关键点，或者也可能在应用于数据时得出一个和原始预想的情况不同的无用的探测结果(例如基于曲率的探测器在很大程度上会受到数据中的大量噪声的影响)。

在这个工作中，我们打算把 3D 关键点的检测转化为把点分为能否被预定义 3D 描述器有效编码的二元分类问题，同时也提出一个有效的算法来定义需要训练的集合。自动学习的这种分类功能有可能自适应识别那些更可能提供对特定数据集的正确结果的点，而不用通过人为设定的探测器发现的特定结构。

一部分研究者研究了机器学习技术在有关关键点检测问题的工作中的应用。就图像而言，最重要的贡献也许就是 FAST[16]了，它基于对探测拐点特性的加速分割测试：从一个训练

集中得到的测试顺序在树中被学习，以提高检测速度。这个方法攻击了几种最近的比较成功的关键点检测器的核心，比如应用于 BRISK[12] 和 ORB[17]中的方法。另一种研究方法主要从通过修剪一种规范的关键点探测器抽取出的关键点，来提高特定任务中的全部性能。在 Hartmann[9]等人提出机器学习算法来研究将要在描述符匹配阶段被放弃的点，这些点在由一个规范的关键点检测算法[3]里被提取出，这个方法叫做 DoG。通过使用随机森林算法来研究这样的“匹配性”，他们展示了这种方法可以很好地提高和加速运动结构管道特征匹配阶段。简单来说，[22]作者展示了可以通过指示 WaldBoost 分类器来保持完成的更高的可重复性，在由标准探测器提取出的点中，只有这些点被认为在给定场景中是有用的。作为一个这种方法的示例应用，他们展示了应用在城市环境中提升之后的图像匹配，分类器学会了关注人造结构以及被忽视的经受着自然改造的物体，比如植被和云彩。一个不同的方法在[26]被提出，通过一系列训练图像的重复率最高的 DoG 关键点被用于定义那些被用来训练回归量的正例，这些正例学习了一个能够凸显强烈光照变化下的相同图像点的凸函数，这些变化来自于天气，季节和每天不同时间的变化。

在 3D 计算机视觉中，用二元分类法计算关键点检测器的方法只有在处理特定数据集的时候被提出，这种数

数据集已经包括了对于兴趣点的定义，比如面部标记[5]。简单来说，作者[23]提出从训练数据而不是手动创造几何特点中学习出一个探测器，这样可以处理基准测试程序的人类使用者选择出的突出的高可变性的结构[7]。这个任务确实没有定义泛化的和可重复的关键点，但要准确找到在建立基准时人类使用者确定的需要突出的关键点。不同的是，Holzer[10]等人提出通过使用学习森林回归模型作为特征来进行二元深度比较的凸函数来加速和提高基于曲率的探测器。

在这篇论文中，我们提出研究一种分类函数作为特征点检测器，目的是辨别一些点，这些点在一个给出的描述器中编码时可以生成正确匹配。相应地，我们提出了一个方法自动生成要用来训练分类器的数据。因此，定义兴趣点时不需要数据集，正如[23,5]。尤其，被驱动的得到的数据有自动地匹配特定描述器的能力，这样生成了一个特定描述符的关键点检测器。学习一个特定描述符探测器的能力对3D视觉领域尤其重要，在这中间，与2D的杰出方法比如SIFT或SURF不同，最先进的描述器[8][19][18][11]仍然缺少一个合适的探测器。

我们也指出，和目前为止提出的2D探测器[22,9]减枝方法对比，我们直接用我们的分类器作为关键点检测器，这样避免了对于选择一个特定人工探测器作为预处理器的需要。它保

持了让更高的适应能力对不同输入数据屈服的潜力，否则可能会受到特定数据集的几何结构的适用性所反映出的特点功能受雇于选中的探测器。此外，对于这样基础的3D探测器的选择结果也有很多问题，因为目前还没有一个确定的被广泛接受的能像DoD在2D图像中一样的3D数据处理算法。

这篇论文的结构如下。第二部分提供了关于我们提出的学习特定描述符探测器的框架的综述，和详述如何创建训练集，以及我们在使用它学习一个分类函数之前，为证实它的有效性进行的实验。第三部分展示用提取出的特征训练分类器的设计，以及我们如何在试验中应用它在给定点云中提取关键点。第四部分展示了对于三种公开的可获得的数据集试验的结果，以及我们的提议与几种最先进的探测器的对比结果。第五部分就试图落实这个新奇的关键点检测器的设计过程中获得的深刻见解做出总结。

## 2.训练一个特定描述符的探测器

我们的方法主要的目的是研究探测放弃了与众不同的3D描述符的3D关键点。真实的训练集的清晰度，尤其是正例的，是至关重要的。我们开发了一系列部分重叠的3D物体的2.5D视图，这些视图关注于特定的数据集，所以定义了正例和反例。

虽然我们的框架具有普遍性，并且在大体上会被任何3D描述符适应，接

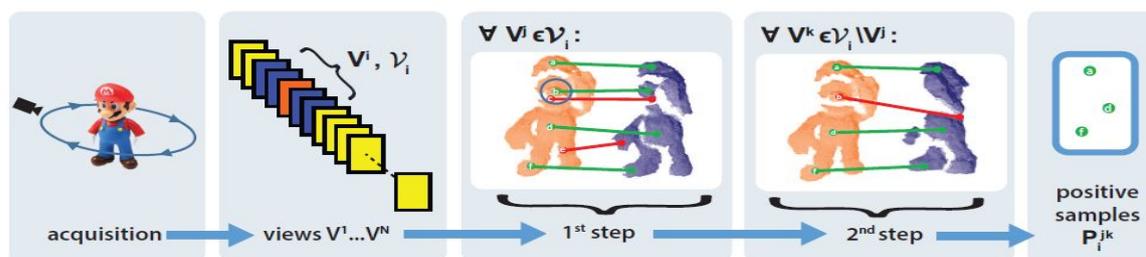


Figure 1. Overview of the definition of positive samples, from left to right: if not available, a set of 2.5D views is generated by simulating a sensor in  $N$  uniformly distributed views; for each view  $V^i$ , its overlapping views are selected; for each overlapping view, matches are established by selecting the nearest neighbor in the descriptor space for each point in  $V^i$ ; matches are removed if they are wrong matches (point e) or close to a match with more similar descriptors (point c); by using every other overlapping view  $V^k$ , the positive set is refined by removing points that do not yield correct matches (point b), because their distinctiveness is not robust enough to changes in viewpoint.

下面我们依赖 SHOT 描述符，给出它在开放源码的 PCL 中的应用以及它全部的特性[19]。此外，我们固定对描述符的支持并导出一个在一定尺度内辨别不同点的固定尺度探测器。

## 2.1 训练集的定义

从一个 3D 数据集中确定一个对象，假设  $\{V^i\}$ ,  $i=1, \dots, N$  是一组  $N$  个 2.5D 视图(如图 1)。这个对象是否需要作为一个完整的 3D 模型在数据集中提供，我们会通过执行合成渲染获得 2.5D 视图，正如例子[2,1]里做的一样。例如，我们展开[1]中采用的方法，描述视图为 3D 模型中心在二十面体中心的节点。

正如图 1 所说，为了辨别可能更有鲁棒性地被匹配的点，并用它们作为关键点正例，我们寻找物体上的在不同区域重叠部分用 SHOT 描述符计算出的结果更近似的点。

在算法上，我们是这么进行的。对于每一个视图  $V^i$ ，我们算出每一点  $p \in V^i$  的 SHOT 描述符结果  $SHOT_p^i$ 。

之后，对每一个视图  $V^i$ ，我们选择它们的子集

$$V^i = \{V^j | V^i \cap V^j \geq \tau, j=1, \dots, N, j \neq i\} \quad (1)$$

即这一系列视图依据一个阈值  $\tau$

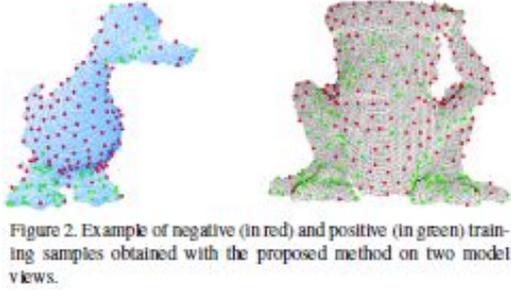
部分地与  $V^i$  重叠。对于每一个部分重叠的视图  $V^j$ ，我们进行两步选点。首先，

对于每一个描述符  $SHOT_p^i$ ，我们找到  $p \in V^j$  的点，产生最邻近的 SHOT 描述符，即  $SHOT_p^j$ 。之后我们最近根据匹配描述符之间的欧氏距离

$$d_{pq} = \| SHOT_p^i - SHOT_p^j \|_2,$$

分出了一对匹配点  $(p, q)$ 。从候选的正例清单

$C_i^j$  的第一对开始，我们检测匹配是否正确，即，两个视图中的对应点是否符合 3D 模型中的同一点。为了让学习的探测器鲁棒性更好以减小探测结果的偏移，我们考虑能否让  $P$  点和  $q$  点之间的距离在同一个坐标系之中表达出来并且完成正确的匹配，即，用已知真实的旋转和平移做视图变换，匹配



结果要小于一个阈值  $\epsilon$ 。这种检测也丢弃了不属于两个视图重合部分的点。

如果匹配是正确的，我们就把她加入这一对视图的正向采样集  $P_i^j$ 。之后我们从候选清单  $C_i^j$  中删除点  $P$ ，同时也删除它邻域内的匹配点对，来模拟空间非极大值抑制的影响，也就是通常通过显著性值来获得关键点作为极值点。对  $V^i$  中的每一个临近  $P$  点的  $P'$  点，即那些到  $P$  点的 3D 距离小于一个临界值  $\epsilon_{NMS}$  的点，我们要把这一对匹配点从匹配分类清单中移除。之后从清单的开始重复这个算法。如果匹配不正确，我们删除这一对点然后重复。

这样，在第一步的最后，我们会获得一个  $V^i$  的正向采样集  $P_i^j$ ，而  $V^j$  用来和 SHOT 描述符进行匹配。为了获取一个不被  $V^i$  和  $V^j$  中特定视点改变影响的正向集，在第二步我们要通过检测其中的点是否能在与其他我们没有用来定义  $P_i^j$  的视图  $V^i$  覆盖的区域很好地匹配，来精练正向采样集。更准

确地说，对于每一个  $V_k \in \mathcal{V}^i \setminus \{V^j\}$ ，即，每一个视图和  $V^i$  中除了已经用过的  $V^j$  之外其他的视图部分地覆盖，我们进行如下的第二步。首先，我们算出两个视图公共部分的正向点，方法是从  $P_i^j$  中对每一个正向样本，用已知真实的旋转和平移做视图变换，检测是否有  $V^k$  中的邻域在距离  $\epsilon$  内。对于每一个计算区域内的正向点，我们选择描述符空间中  $V^k$  的描述符和我们检测是否和 3D 距离测试结果正确匹配的描述符之间最近的邻域。特别地，一个点被要求有一个和  $V^k$  视图之间正确的匹配，这样可以认为它是一个正向样本。用这种方法，我们获得了一个关于每一个视图  $V^k$  精练的正向样本集  $P_i^{jk}$ 。

$$P_i = \bigcup_{\substack{j=1 \\ j \neq i}}^{|\mathcal{V}^i|} \bigcup_{\substack{k=1 \\ k \neq i, j}}^{|\mathcal{V}^i|} P_i^{jk} \quad (2)$$

最终的正向样本集和  $V^i$  之间的关系是，精练的正向集覆盖所有重叠视图，最后，正向样本集  $P$  由所有视图中的正向样本  $P^i$  集合而成。要获得反向样本，这几种替代选择是可行的：

i) 实现一个和描述的大致相同的进程，但对列表分类降序排列，这样产生最不同的描述符的点就变成了反向样本；

ii)包括反向集在内的所有不正确匹配会在滚动原始列表时被发现，直到达到预先定义的数字，这样分类器可以学习如何在产生高相似度描述符时，避免错误匹配；

iii)在反向的不在正向集中的点中随意取样，来获得对没有不同部分的点云的统一覆盖。

我们从最后一种策略开始，这种策略最容易实现，并且发现结果很不错。因此，我们把对上面提出的其他选择的实现和测试留给了未来的调查研究。为了让点散乱地分布在视图所有没有内容的部分，我们也根据取样出的新反向样本，移动所有在它周围某一个范围  $\epsilon_{neg}$  内的点。图 2 展示了两个从模型视图中提取的正向和反向样本。从图片可以看出，正向点不均衡地分布在表面上，并且不一定出现在显著位置，在几何上很好地定义了点，比如右边模型的膝盖和肘关节或尾部，这些最常规的探测器就能提取出来。

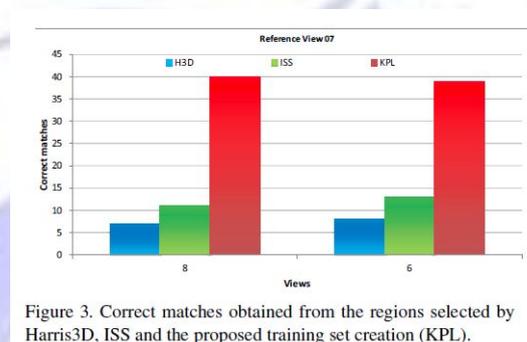


Figure 3. Correct matches obtained from the regions selected by Harris3D, ISS and the proposed training set creation (KPL).

## 2.2 训练集的确定

在定义用来训练特定分类器的特

征之前，我们研究提出的用 SHOT 描述器定义要编码的选定区域的方法的有效性。换句话说，我们希望首先发现完美的分类器能否可得到，是否我们的训练集定义的分类函数可以相对于其他基于几何特征用标准的没有描述器的探测器分类函数，随意地选择将要被 SHOT 进行匹配的区域。我们有意比较从参考视图  $V^i$  和第二步选择中的最近重叠邻域  $V^k$  匹配 SHOT 描述器得到的正确结果数目。尤其是，我们比较了 SHOT 描述器在第一次选择完成的正向集  $P_i^j$  的点上算出的匹配性能，和 ISS[29]关键点提取的结果以及 Harris3D 探测器在 PCL 中得到的结果。为了评估区域不考虑重复的唯一的描述，我们严格地把  $V^i$  中的关键点移到  $V^k$  中，检测 SHOT 空间中变换的关键点最近的邻域是否在  $\epsilon$  内。图 3 展现了一个从 Kinect 数据集[19]中的 Mario 模型的视图 7 的结果样本，它和视图 6 和 8 都有部分重叠。我们选择视图 7 和 6 中的正向点进行第一步，之后检测我们从视图 7 和 8 中得到的正确匹配有多少(最左边的直方图)，之后我们转换了视图 6 和 8 的功能。我们可以很清楚地看出，我们的点可以检测出更适合选定描述器的区域，即使它们从不同的视图学习而来。我们可以由此解决定义适当的目的是学习探测 3D 关键点放弃特殊 SHOT 探测器

的分类器的问题，通过使用提出的训练集。

### 3.分类器的设计

我们选择随机森林算法[3]作为分类器来研究我们训练集的关键点检测函数有几个原因。首先，它已经被成功地应用于计算机视觉并解决了一些问题，同时最近解决了 3D 关键点检测问题[23]。更多地，森林随机算法基于最快的分类器来测试运行时间，即使是学习复杂分类函数的时候，相反举例来说有着非线性核心程序的 SVM。在把一个分类器当探测器使用时这尤其重要，因为这已经应用于每一个输入点云。最后，一个随机森林算法很自然能够多级分类，并且因此符合我们要延伸我们的工作到支持多重大小以可能想出一个尺度合适的特定描述探测器的需求。

对特征的定义用来捕捉点云周围的几何结构，是分类器设计的重要一步。通常的特征在对图像应用随机森林算法时提出，简化为二元特征，就像随机灰度偏差[6,10]。然而，要在 3D 邻域应用随机偏差时保持循环不变性要求对局部参考系的定义，比如 SHOT[19]或 MeshHoG[28]中的例子，以对同一点的每一个二元测试正确联合。虽然，3D 描述器的从点云的稀疏子集计算的实例不同，这样一种珍贵的参考系估算需要在点云中的每一点进行运算。此外，点的偏差要处理修

改视图来展示不同点密度。

基于以上的考虑，我们选择个别的循环不变特征，它不需要定义局部参考系。由 SHOT 描述器的鲁棒性和描述性启发，我们依靠正常方位的直方图，即对每一个我们需要算出特征适量的参考点  $P$  和它相邻的范围  $r_{feat}$  内的所有点  $q$ ， $P$  和  $q$  的法向量夹角在直方图中算出并量子化，就像  $q$  所在的球的扇区。在 SHOT 中，球沿着半径分阶和点的极坐标方位角划分扇区。为了避免局部参考系的计算，我们改变对球面支撑的分割形状，只考虑沿着径向尺寸  $N_r$  细分，以及用  $N_b$  二进制计算一个直方图这样球壳就得到了。由于直方图是为球壳而计算，它们内在就具有旋转不变性，所以可以免去对局部 RF 的估算。正如[19]中的讨论，为了让计算更快鲁棒性更高，我们没有对参考点的法向量和考察点的法向量的夹角进行量子化，而只是做它们的数乘，这样相当于就是两法向量夹角的余弦。为了避免量化误差，在直方图中我们选择双线性插值法。最终，特征向量由划分它的欧几里得范数限制点密度的变化效果完成标准化。图 4 提供了一个图解的特征估算进程的概述。

给出所有数据集中的模型，我们构造训练集并在每点提取方向特征直方图，设置  $\epsilon_{neg}$  以获得一个平衡的训练

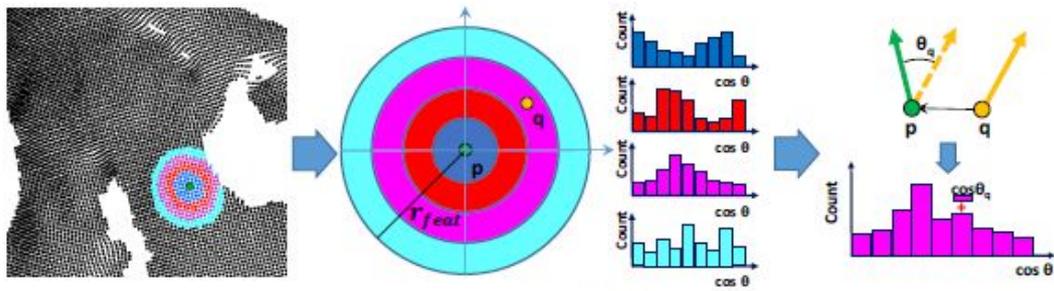


Figure 4. Overview of the feature computation. In the example,  $N_r = 4$  radial subdivisions are used (for ease of visualization, a 2D representation of the 3D spherical support is portrayed). Each subdivision is associated with a spherical volume and a histogram comprising  $N_b$  bins, which accumulates the cosine of the angle between the normal of  $p$  (central point) and that of each point  $q$  falling in the corresponding spherical volume.

集。我们之后用  $T$  树训练随机森林。当在检测不可见输入点云的关键点时，我们在每一点用分类器并且算出根据关键点分类得出的  $T_p$  树里的数量。点关联的分数是  $T_p/T$ ，如果分数大于一个最小值分  $s_{min} \geq 0.5$ ，并且是一个半径  $r_{nms}$  内局部分数的最大值，这个点会被探测为一个关键点。

#### 4. 试验结果

为了试验性地证实我们的提议，我们测量了描述器匹配可以在一个标准匹配管道中用我们的森林随机算法由更替手工制作的基于显著性分析的探测器提高的性能。

为了落实我们的提议，我们用由 OpenCV 实现的随机森林算法和 PCL 中包含的 SHOT 算法。至于基于显著性的探测器，我们选择那些 PCL 中可用的，也就是 ISS，Harris3D 和 NARF[21]，以及 Mian 等人[15]提出的在[24]的评估中被称为 KPQ 的探测

器。结果是，和 ISS 一样，是尺度不变探测器的最优算法之一。我们也考虑像基线 3D 探测器一样均匀采样，考虑到它是正常地由 3D 社区快速展开，虽然很少使用，也是供替代的选择。传说中，我们参考了我们的提议关键点学习(KPL)。

我们用所有三个公共[24]中用到的可获得的数据集测试了所有探测器来比较它们：Laser Scanner 数据集由 Mian 等人提出；Random Views 数据集基于 Stanford 3D 扫描知识库，最初在探测器评估中提出[24]；[19]中介绍的 Kinect 数据集。每个数据集包括一个有趣的模型的清单，我们用它来训练我们的探测器，以及几个场景，我们在那里做关键点检测。当模型是全 3D 的时，把它当做 Laser Scanner 和 Random Views 的实例，为了应用我们的学习算法，我们简化了 42 个等距 2.5D 视图作为二十面体的顶点，随着[1]中的方法论的提出，它的实现在 PCL 中也可以完成。场景是 2.5D 视图独立地由模型和模型的描绘安排和其他目标(杂波)捕获。因此，从模型中提

取的训练探测器的视图与场景中出现的模型视图无关。此外，这个评估方法模拟了很可能付诸实用的技术，任何一个模型都是先预知或在初始化时取得，但是场景是不可见的的数据。

在测试和训练中用在每一个数据集上的参数在表 1 中显示。由于可以从图中看出，我们用同样对 SHOT 经过所有数据集支持的大小，缩放适应每一个数据集的计量单位(即 Kinect 和 Random Views 的米，Laser Scanner 的毫米)。同样地，我们用相同支持的大小计算我们探测器的特征值和其他方法的显著性函数。

#### 4.1 随机森林的调整

随机森林可以学习复杂的函数，虽然只使用几个自我解释的参数。然而，对它的参数比如树的深度和节点处停止分裂的样本数的极大值，性能和泛化能力的作用效果还不明显：如，Breiman[3]提议让树一直生长直到一个节点只有一个样本为止，然而 Criminisi 等人[6]通常用大量样本来估算节点的后部。

因此，我们研究树的数量的作用，最大的树的深度，以及停止分裂的样本数。为了估算一般的分类器的效果，我们在视图水平上完成了三褶皱交叉验证，即，我们用  $2/3^N$  的模型视图定义训练集，其余的视图用来测试，并且我们关心训练和测试的错误。

就树的数目而言，我们让数值在

10-100 范围中变化，我们检查树的数

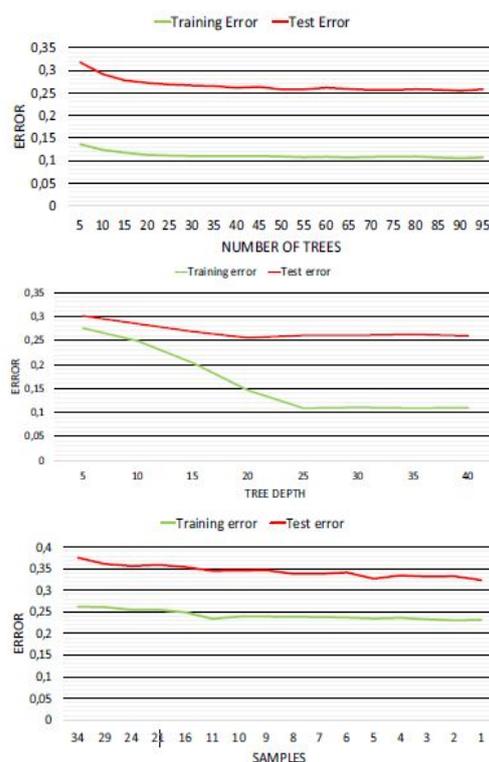


Figure 5. Random Forest parameters tuning over the Kinect dataset.

目在 10-40 之间时深度的作用，我们也让样本数从 1% 的训练点集停止分裂变化，直到 1。图 5 展示了当我们考虑到 50 颗以上的树时分类器的性能没有提高。更显然的是，让树生长的深度达到 25 级之后并没有什么好处，我们通过 5 个以下样本在节点左侧时停止分裂得到了最好的一般化的结果，因为有限的过度拟合可以用更小的数值观察到，即，即使训练错误减少，测试错误也在增加。对于 Laser Scanner 数据集，训练需要更多数据的特征 (100K 正例对 Kinect 数据集的 8000 正例)，我们反而从 100 棵树中得到了最好的性能，同样的深度，平均一个样本节点停止分裂。

## 4.2 Laser Scanner 数据集

这个数据集包含了 4 个全 3D 模型和 50 个场景，模型相互咬合，也存在不包含在模型列表的对象，所以有些杂乱。

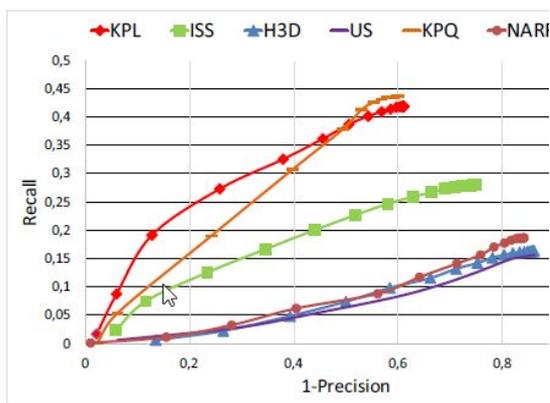


Figure 6. Results on the *Laser Scanner* dataset.

为了进行描述器匹配试验，我们首先在所有模型的所有视图上探测关键点，计算相应的 SHOT 描述器，并创造唯一的基于所有模型描述器的随机树。我们之后在场景上运行探测器，对于每一场景关键点我们把它和随机树中 SHOT 描述器里和它最近的点进行匹配。对于每一个匹配，我们相应地检测它是否是正确匹配，以及是真阳性还是假阳性增长。通过改变 SHOT 描述器的最大值的临界值来完成匹配，我们为探测器确定它的精率，见图 6。

KPQ 是最好的基于显著性特征的提议的探测器。这让人有些不那么惊讶，因为这个探测器最初是为 Laser Scanner 数据集设计的。ISS 也给出了

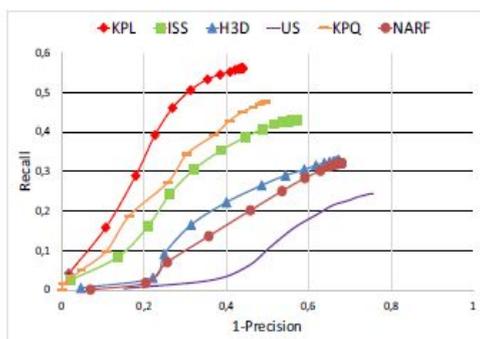
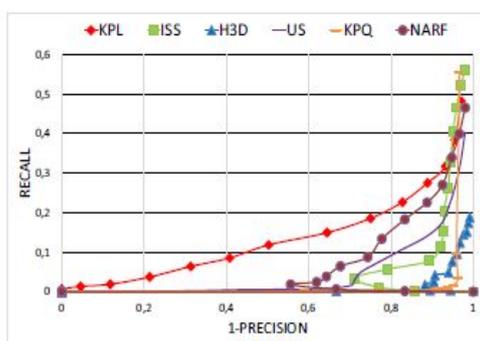
合理的结果，鉴于 NARF 和 Harris3D 执行结果和基线均匀采样相似。我们的提议能鉴别最佳的 SHOT 描述器检测区域，甚至比专门为这种数据集调整的探测器更好，这肯定了我们的直觉：基于显著性特征的探测器不能选择最好的让所有探测器-描述器管道性能达到最优的区域。

## 4.3 随机视图的迁移学习

这个数据集包含 6 个全 3D 模型和 36 个场景，模型在场景中相互遮挡但是并不杂乱。模型和场景有很高的分辨率，但综合的随机噪声也被加入了场景中。这里我们考虑到有  $\sigma=0.1$  的网眼分辨率单元的高斯噪声的场景。

用在 Laser Scanner 数据集上的测试协议是一样的。显然，因为数据集提供了一个比得上 Laser Scanner 数据集的细节和噪声的层次，我们不用为这个数据集训练一个随机森林，而是用已经在先前的数据集上学习好的随机森林。这可以检测我们的方法对于不可见物体的泛化能力。

结果在图 7 中有体现。总的来说，探测器的相对命令比得上之前的数据集。然而，我们的建议和 KPQ 之间的差别在于：基于显著性特征的探测器在对于不同数据集维持相似特征级别更加困难。此外，这个结果展示了我们为分类器选择训练集的方法、我们提取的特征，和选择的参数在学习分类函数上有效并有很高的泛化能力。

Figure 7. Results on the *Random Views* dataset.Figure 8. Results on the *Kinect* dataset.

#### 4. Kinect 数据集

这个数据集包含 7 个 2.5D 视图的模型和 17 个模型在高混乱度和重叠下的场景。数据质量不高，因为它由廉价的普通相机获得。

为了在数据集上比较已评估的探测器，我们跟随已经在最初的 SHOT 论文里提出的方案。尤其，关键点在场景的每一个模型中被提取和描述。之后提取描述器，在模型关键点的位置和从混乱中提取的额外关键点的位置。每一个场景中的描述符之后匹配模型描述符的比例标准[14]和检查几何正确性。通过改变比率检定的临界值，我们从图 8 获得 Precision-Recall

曲线。

这个数据集很具有挑战性。结果大部分都比之前的数据集获得的数据更差。标记出手动探测器变化的相对命令很重要：一方面，基线统一取样执行得比 KPQ, ISS 和 Harris3D 好；另一方面，在先前的数据集中特性不那么令人满意的 NARF，结果是最好的基于显著性特征的探测器。我们的 KPL 探测器很明显有更好的性能，显示出适应不同感觉通道的能力和对被提出的有旋转不变特征的鲁棒性。

#### 5. 结论及后续工作

关键点检测的问题是为了依据 3D 描述器算法提供特殊区域，来提高全部匹配管道的性能达到能成功依据自动学习解决问题的目的。对符合正确 3D 匹配描述空间中最近邻域的正向训练样本的定义产生了一个相对于标准人工得出的显著性更适合用描述器特殊编码的鉴别区域的分类函数。一个适当调整过的随机森林能够很好地从训练数据中用提出的旋转不变几何特征学习这样的函数。我们的特有描述符探测器完美适应不同获得物形式的数据监测。当在数据监测中用同样的形式，也会展现出好的从一个数据集到另一个的迁移学习能力，所以训练一个新的分类器也许不是那么必要。

因此，也许在这项工作未来的发展中，学习对特定 3D 数据有普遍性的探测器是可能的。此外，随机森林完美

地处理多类分类的能力让学习一个自适应尺度探测器变为可能，这可能会进一步推进管道的匹配性能。一个级联的方法也可能被设计用做迅速排除输入点云中不太可能得到好关键点的区域，只用几个树就可以，从而提高测试中运行时间的性能。最后，由于基础的原理和方法论很具有一般性，我们计划根据其他最先进的 3D 描述器，比如[8,18,11]，应用我们的学习框架获得能鉴别好的关键点的探测器。

## 参考文献

- [1] A. Aldoma, F. Tombari, L. Di Stefano, and M. Vincze. A global hypotheses verification method for 3d object recognition. In European Conf. on Computer Vision (ECCV), volume 7574 of Lecture Notes in Computer Science, pages 511 - 524. Springer Berlin Heidelberg, 2012.
- [2] P. Bariya and K. Nishino. Scale-hierarchical 3d object recognition in cluttered scenes. In Computer Vision and Pattern Recognition (CVPR), IEEE Conf. on, pages 1657 - 1664, June 2010.
- [3] L. Breiman. Random forests. *Machine Learning*, 45(1):5 - 32, 2001.
- [4] U. Castellani, M. Cristani, and S. Fantoni. Sparse points matching by combining 3D mesh saliency with statistical descriptors. *Proc. Computer Graphics Forum*, pages 643 - 652, 2008.
- [5] C. Creusot, N. Pears, and J. Austin. A machine-learning approach to keypoint detection and landmarking on 3d meshes. *International Journal of Computer Vision*, 102(1-3):146 - 179, 2013.
- [6] A. Criminisi, J. Shotton, and E. Konukoglu. Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. *Foundations and Trends in Computer Graphics and Vision*: Vol. 7: No 2-3, pp 81-227, 2012.
- [7] H. Dutagaci, C. Cheung, and A. Godil. Evaluation of 3d interest point detection techniques via human-generated ground truth. *The Visual Computer*, 28(9):901 - 917, 2012.
- [8] Y. Guo, F. Sohel, M. Bennamoun, M. Lu, and J. Wan. Trisi: A distinctive local surface descriptor for 3d modeling and object recognition. *Computer Graphics Theory and Applications (GRAPP)*, 8th International Conference on, 2013.
- [9] W. Hartmann, M. Havlena, and K. Schindler. Predicting matchability. In *Computer Vision and Pattern Recognition (CVPR)*, 2014 IEEE Conference on, pages 9 - 16, June 2014.
- [10] S. Holzer, J. Shotton, and P. Kohli. Learning to efficiently detect

repeatable interest points in depth data. In Computer Vision (ECCV), 2012 IEEE European Conference on, 2012.

[11] A. Johnson and M. Hebert. Using spin images for efficient object recognition in cluttered 3D scenes. *IEEE Trans. On Pattern Analysis and Machine Intelligence*, 21(5):433 – 449,1999.

[12] S. Leutenegger, M. Chli, and R. Siegwart. BRISK: Binary robust invariant scalable keypoints. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 2548 – 2555, Nov 2011.

[13] Y. Li, S. Wang, Q. Tian, and X. Ding. A survey of recent advances in visual feature detection. *Neurocomputing*, 149, Part B(0):736 – 751, 2015.

[14] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91 – 110, nov 2004.

[15] A. S. Mian, M. Bennamoun, and R. A. Owens. On the repeatability and quality of keypoints for local feature-based 3D object retrieval from cluttered scenes. *Int. Journal of Computer Vision*, 89(2-3):348 – 361, 2010.

[16] E. Rosten, R. Porter, and T. Drummond. Faster and better: a machine learning approach to corner detection.

*IEEE Trans. Pattern Anal. Mach. Intell.*, 32(1):105 – 119, jan 2010.

[17] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski. Orb: An efficient alternative to sift or surf. *Computer Vision, IEEE International Conference on*, 0:2564 – 2571, 2011.

[18] R. B. Rusu, N. Blodow, and M. Beetz. Fast Point Feature Histograms (FPFH) for 3D registration. In *International Conference on Robotics and Automation*, pages 3212 – 3217, 2009.

[19] S. Salti, F. Tombari, and L. D. Stefano. Shot: Unique signatures of histograms for surface and texture description. *Computer Vision and Image Understanding*, 125(0):251 – 264, 2014.

[20] J. Shi and C. Tomasi. Good features to track. In *1994 IEEE Conf. on Computer Vision and Pattern Recognition(CVPR' 94)*, pages 593 – 600, 1994.

[21] B. Steder, R. B. Rusu, K. Konolige, and W. Burgard. Point feature extraction on 3d range scans taking into account object boundaries. In *Robotics and automation (icra), 2011 iee international conference on*, pages 2601 – 2608. IEEE, 2011.

[22] C. Strecha, A. Lindner, K. Ali, and P. Fua. Training for task specific keypoint detection. In J. Denzler, G.

Notni, and H. Se, editors, Pattern Recognition, volume 5748 of Lecture Notes in Computer Science, pages 151 - 160. Springer Berlin Heidelberg, 2009.

[23] L. Teran and P. Mordohai. 3D interest point detection via discriminative learning. In ECCV 2014, volume 8689 of Lecture Notes in Computer Science, pages 159 - 173. Springer International Publishing, 2014.

[24] F. Tombari, S. Salti, and L. DiStefano. Performance evaluation of 3d keypoint detectors. *Int. J. of Computer Vision*, 102(1-3):198 - 220, 2013.

[25] T. Tuytelaars and K. Mikolajczyk. Local invariant feature detectors: a survey. *Found. Trends. Comput. Graph. Vis.*, 3(3):177 - 280, jul 2008.

[26] Y. Verdie, K. M. Yi, P. Fua, and V. Lepetit. TILDE: A Temporally Invariant Learned DETector. In *Proceedings of the Computer Vision and Pattern Recognition*, 2015.

[27] P. Viola and M. J. Jones. Robust real-time face detection. *Int. J. of Computer Vision*, 57(2):137 - 154, 2004.

[28] A. Zaharescu, E. Boyer, K. Varanasi, and R. Horaud. Surface feature detection and description with applications to mesh matching. *Proc. Int. Conf. on Computer Vision and Pattern*

Recognition (CVPR), pages 373 - 380, 2009.

[29] Y. Zhong. Intrinsic shape signatures: A shape descriptor for 3D object recognition. In *Proc. Int. Conf. on Computer Vision Workshops*, pages 1 - 8, 2009.