指导教师： 杨涛　　提交时间： 2013/3/19

CVPR2015 Paper Translation

No: 01

姓名： 刘冰

学号： 2013302628

班号： 10011305

通过视图合成的 24/7 位置识别

摘要

我们解决大规模的视觉识别的现场发生重大变化的情况下的问题,例如,由于照明(白天/晚上),季节的变化,随时间老化,或结构的修改,如建筑的建造或销毁等。这种情况代表了当前大规模的位置识别方法的一项重大挑战。这项工作有以下三个主要贡献。首先,我们证明了当查询图像和数据库图像描绘的场景都是相同的视角时，匹配发生大变化的场景变得容易了。其次,在此基础上观察,我们开发一个新地方相结合的识别方法和一个高效合成新颖的观点，一个紧凑的可转位图像表示。第三,我们引入一个新的具有挑战性的数据集(1125 个东京的查询图像)包含重大变化的照明(白天,日落,夜晚)以及场景中的结构变化。在这个具有挑战性的数据中我们证明了我们提出的方法明显优于其他大规模的识别技术。

1. 介绍

近年来在大型视觉位置识别问题［27、36]取得了巨大进展［3、 6、 7、 8、10、 14、 24、 28、 34、 35、 36、40、 44]。现在可以在一个由 1M 图像描述或由三维点云重建的城市中获得查询照片的摄像机的精确位置。这些表示法是建立在局部不变特征如 SIFT［29］以便识别可以由其他对象一起观点、 规模或部分遮挡的适度变化。

（a） 查询图像 （b） 街景

（c） 合成视图 (d) 在地图上的位置

图 1。匹配整个场景外观主要变化是类似的观点更容易。（a） 查询图像。（b）由于现场外观发生的重大变化及角度的变化,原始的数据库映像无法匹配查询。（c） 匹配更多类似的合成视图是可能的。（d） (a-c) 在地图上的位置图。点和箭头指示相机位置和查看方向。

文件［33、 39］或索引技术的量化产品［20］。尽管这进展，整个场景外观照明 （昼/夜），由于主要变化确定同一地点改变的季节，老化，或结构上的修改时间［12，30］，如图 1 所示仍然是一项重大挑战。解决这一问题会有，然而，重大的实际意义。想象一下，例如，自动搜索公共档案馆找到所有意象描绘的相同的地方，分析针对应用程序的体系结构、 考古学和城市规划；随着时间推移的变化或者想象的相同的地方在不同照明条件，季节或落后的时间。

在本文中，我们证明，匹配整个场景外观变化较大时，更容易查询图像和数据库图像描绘的场景从大约相同的观点。我们通过合成在地图上的密被采样网格的虚拟视图来实现这一想法。这会带来以下三个主要挑战。

首先，我们要怎样有效地综合虚拟观点为整个城市？第二，我们如何处理增加的数据库的大小增加了额外的综合意见吗？最后，我们如何代表现场外观大变化的鲁棒性的方式合成的意见？

为了解决这些问题，第一，发展可以呈现虚拟视图直接从谷歌街景视图全景图和他们的关联近似深度地图视图合成方法不需要重建现场精确 3D 模型。而所产生的图像往往是喧闹和包含工件，我们表明，这种表示是足够的大规模的地方识别任务。这种方法的主要优势是街道视图数据是可用全世界开放升值可能性为真正的行星尺度［23］的位置识别。其次，以应付大量的合成数据 — — 就像九倍更多的图片比在原始的街景 — — 我们使用紧凑 VLAD 编码［2，21］的局部图像描述符，适于进行高效的压缩、 存储和索引。最后，我们代表图像跨多尺度使用密集采样局部梯度基于的描述符 (SIFT［29］在我们的例子)。我们发现这种表示方法具有更强的照明、 老化、等外观变化较大。因为它并不依赖可重复检测的局部不变特征，如高斯拉普拉斯［29］。虽然局部不变特征已成功地用于近两年来简明地表示图像匹配的整个观点和规模［41］他们往往非可重复整个外观由于，例如非模型化变化。强烈的透视效果或夜景照明［4，9］的重大变化。不依靠局部不变特征检测是要付出代价的减少不变性到几何变换。然而，我们发现这其实是一种优势，而不是一个问题，因为生成的表示是更有特色，从而更好地应对虚假的正面形象，由于很多大型数据库扩张合成视图的增加率。

2. 相关工作

地方与局部不变特征的识别。大型地方识别经常制定作为图像检索［22，33］在哪里查询照片本地化的匹配对大型数据库的地理标记的图像，如谷歌街景［6、 8、 10、 14、 24、 35、 36、 40、 44］变异。可以事先也重建环境的三维结构和查询直接与重建点云［28，34］然后相匹配而不是单个图像。这些方法的基本外观表示基于局部不变特征［41］，聚合成一个映像级可转位刀片表示［8、 10、 14、 24、 40、 44］，或向个别重建 3D 点［28，34］相关联。这些方法为大型匹配整个规模和观点的由局部不变特征探测器建模的适度变化方面有突出表现。但是，匹配跨非模型化的外观变化，如照明的主要变化，老化或季节还有很多挑战。

我们调查基于描述符浓密采样在图像，而不是基于局部不变特征的紧凑表示。密被采样的描述符长用于类别级别识别［5、 11、 26、 32］包

括类别级别定位［14］，但由于其有限的实例级识别介绍了对几何变换不变性。我们表明，结合虚拟视图合成致密交涉可用于大型地方识别整个场景出现的重大变化。

为实例级匹配的虚拟视图。有关我们的工作还有方法生成某种形式的虚拟数据为实例级的匹配，但是通常他们专注于扩大的可识别的观点［17、 37、 43］或跨域［4，38］匹配范围和不考虑紧凑的表示法，为大规模的应用。［17］生成包的视觉词描述符为虚拟位置在地图上的现有视图中提取到更好的模型现场能见度。山等[37] 使用三维结构合成虚拟视图以匹配整个空域图像的对齐方式的极端观点变化。吴等人[43] 纠正基于底层的 3D 结构，延长局部不变特征（SIFT）的观点不变性的图像。他们的方法成功申请位置识别[8]，但需要已知的三维结构或在查询方面的整改。最近，呈现虚拟视图还探讨了跨域匹配对齐画到 3D 模型［4］或匹配 SIFT 描述符之间的图像和激光扫描［38］。

模拟场景照明位置识别。在位置识别建模室外照明相关的工作集中于估计的地点和时间戳从观察到的照明效果［13，18］。相比之下，我们专注于跨的光照变化认识相同的场景。然而，如果照明效果可以可靠地合成［25］由此产生的图像可以用于进一步扩大图像数据库。

3. 在发生重大变化的外观上匹配局部描述符

在本节中，我们探讨使用局部不变特征的图像匹配跨越的重大变化的挑战在场景中的变化。

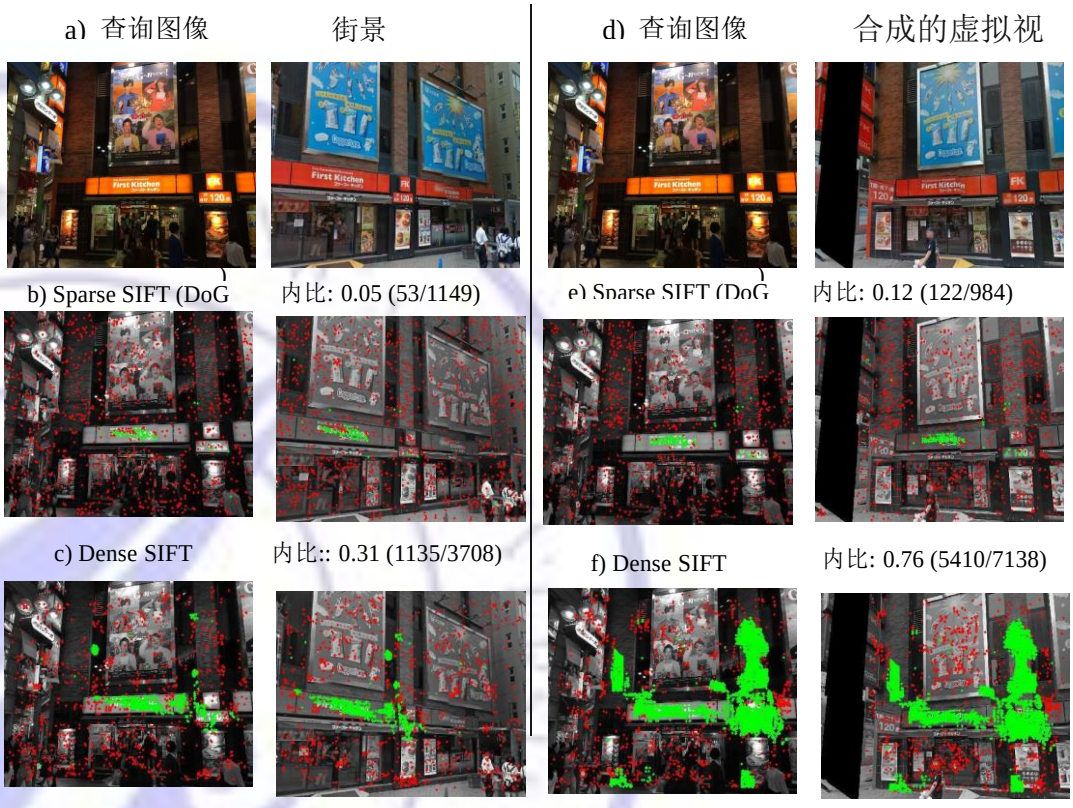


图 2。匹配整个照明和场景中的结构变化。第一行：相同的查询图像与街景图

像描绘了相同地方从不同的角度（a）和合成的虚拟视图，描述查询到地方从相同的角度（d）相匹配。第二行 ： 说明跨照明重大改变采样的 SIFT 描述符匹配困难为相同（e）不同的（b）的观点。第三行 ： 密被采样的描述符可以匹配整个照明（c）发生大的变化和匹配时容易得多的观点是相似的（f）。在所有情况下暂定未匹配显示为红色和几何验证的匹配以绿色显示。请注意如何基于虚拟视图的合成，再加上密被采样描述符的拟议的方法（f）获得显著高于内比（0.76）对此具有挑战性图像两人的主要照明和结构

在现场到场景中的白天、夜晚和结构的变化。我们首先说明局部不变特征基于高斯特征探测器的区别不是在这种条件下可靠地重复。然后我们显示密被采样的描述符结果中更好地匹配，但有限的不变性妇女遭受到几何变换 （规模和观点）。最后，我们证明 该匹配显著提高当我们匹配到大约相同的视角来合成的虚拟视图。在本节中，我们说明以上各点上一个匹配的例子，图 2 所示。我们验证这些定量研究中第 5 条的地方识别任务。

在图 2 中的所有示例中我们建立暂定比赛由发现相互最近的描述符。暂定比赛红色所示。我们然后几何验证匹配项，它是通过反复查找使用 RANSAC 的几个重点。以绿色显示几何一致匹配（相）。我们认为作为正确的所有几何验证的匹配 （虽然仍然可能不正确的几场比赛）。匹配的质量被衡量内比率，*即*。几何一致匹配比例。内比之间 0 和 1 的一个完美的分数与 1 暂定的所有匹配项时几何一致。

首先，我们匹配的直立 SIFT 描述符[1] 在狗要点 [29] 采样之间查询图像和街景视图图像描绘查询地方（图 2(a)）从不同的观点。比赛图 2 （b）所示，导致只有 0.05，内比清楚地显示匹配的高斯要点跨外观变化较大的困难。

第二，我们重复相同的步骤为合成视图（图 2(d)），捕获查询从相同的地方作为查询图像的观点。结果如图 2（e）所示。只有 0.12 产生内的比值表明，匹配整个外观变化较大高斯要点是困难的尽管这两种观点有了相同的观点。

第三，我们提取的宽度为 SIFT 描述符 40 像素（在 640 × 480 图像）在规则的密被采样网格与步幅的 2 像素。描述符匹配进行提取在疏生检测要点的描述符的方式一样。已经跨 不同视角和光照条件匹配的密被采样的描述符显示改善，对稀疏的关键点，相比内比增加 0.05 至 0.31 （图 2(c)）。描述符 (SIFT) 是完全相同的两种采样方法的事实表明，存在的主要问题是非重复性支撑的稀疏采样的方法，而不是本身的描述符的高斯局部不变特征的差异。

最后，我们应用密被采样的描述符图像对不同照明条件，但类似的观点（图 2(d)）。图 2 （f） 所示的比赛。内比进一步增大到 0.76 清楚地显示虚拟视图合成致密描述符匹配的好处。

4. 从街道级图像视图合成

在本节中我们描述我们的视图合成方法，扩大地理标记图像数据库与其他观点在规则的网格采样。综合其他意见，我们使用现有的全景图像，以及与每个全景图，相关联的智者平面深度映射图 4 所示。分段平面深度映射提供只有非常粗糙的三维结构的场景，这经常会导致在合成图像中有明显的失真。然而，在第 5 我们证明这种品质是不足以显著改善地方识别性能。此外，此数据是本质上是提供世界各地 [15]，因此开放行星尺度视图合成和地方识别 [23] 的可能性。视图合成收益两个步骤。我们合成候选

人虚拟相机的位置，所遵循的合成的个人意见。接下来讨论了两个步骤。

我们生成候选相机位置定期5 *米*× 5 *米*网格覆盖原来的街景相机位置在地图上。我们只需要生成内的相机位置 20*m*距离从原始的街景视图轨迹，轨迹通过连接邻近的街道视图相机位置。我们发现那远比 20*m* 往往产生重大文物在合成视图中的。我们还使用可用深度映射丢弃会躺在建筑物内部的相机位置。综合意见的相机位置介绍了地图上，图 3 所示。

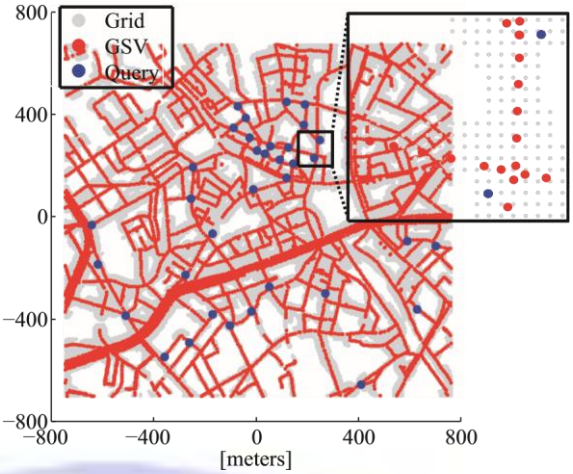
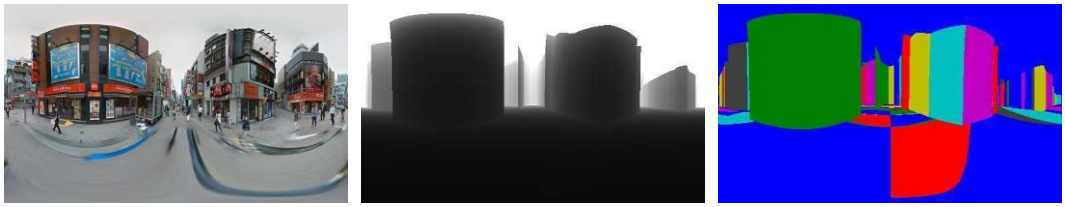


图 3。结合街道视图图像与合成的意见。该图显示相机位置为 24/7-东京数据集的一部分。原始的街景图片的位置显示为红色，综合意见的立场 (5×5 m 网格) 灰色和查询图像显示为蓝色的位置所示。插图 （右上） 显示一个道路交叉口的一个特写镜头。地理标记图像数据库包括 75,984 从原始生成视图 6,332 街景视图全景图和 597,744 合成在生成视图 49,812 虚拟摄像机位置。

综合虚拟的意见，我们使用从 Google 上下载的全景和深度地图中的特定虚拟相机位置映射 [15]。每个全景捕捉 360^7 由 180^7 水平和垂直可视角度，分别，和大小 13,312 × 6,656 像素，如图 4 （a） 所示。深度地图是一套 3D 平面参数（正常和每个平面的距离）和 512 × 256 指数指向，为每个像素，一架飞机，如图 4 （c） 所示的图像。我们使用此索引可以查找相应平面的每个像素，这使我们能够生成的实际深度映射为全景图，如图 4 （b） 所示。在一个特定的虚拟摄像机位置的所有视图是从最近的街道视图图像全景图和深度图都合成的。虚拟视图合成标准射线追踪与双线性插值。详细，每个像素在合成的虚拟视图中，我们从中心的虚拟摄像机用光线投射、 相交与平面的 3D 结构，从最接近的街景视图全景深度图获得、 项目到街景视图全景，交叉和插值输出像素值从相邻的像素。对于每个虚拟摄像机位置我们生成 12 透视图的 1,280 × 960 像素（对应于 60 度的水平视角）与螺距方向 12 和以下 12 偏航方向$[0°, 30°, ..., 360°]$。此全景视图采样是类似于如 [8，40]。合成的虚拟视图的例子所示 1、 8 和 9 的数字。虽然合成的意见，亦缺少信息和工件（例如错误地呈现人或对象），我们发现这个简单的渲染是已经不足以改善地方识别性能。高质量合成可以潜在地通过结合来自多个全景图的信息。呈现一个虚拟视图需要大约一秒钟。我们生成相同的街景视图的原始图像的全景视图集，并将真实的和虚拟的意见合并到单个位置识别数据库。注意，虚拟视图只需要用于提取紧凑密集 VLAD 描述符中第 3 条所述，可以之后丢弃。



（a） 街景视图全景图 (b) 有关深度映射 （c） 个人现场模型

图 4。视图合成的输入的数据。（a） 街景视图全景图。（b）分段相关平面深度映射。亮度指示距离。（c） 个别场景模型所示不同的颜色。

(a) 查询1. (b) 查询2. (c) 查询3. (d) 数据库图像

图 5。从新收集的示例查询图像 24/7 东京数据集。每个地方的查询集捕获在一天的不同时间 ： 白天 （a）、 (b) 日落和 （c） 晚上。为比较，在关闭的位置数据库-街景图像 (d) 所示。(D) 在数据库图像和查询图像 (a、 b、 c) 之间注意外观 （场景中的光照变化） 的主要变化。

5。 实验

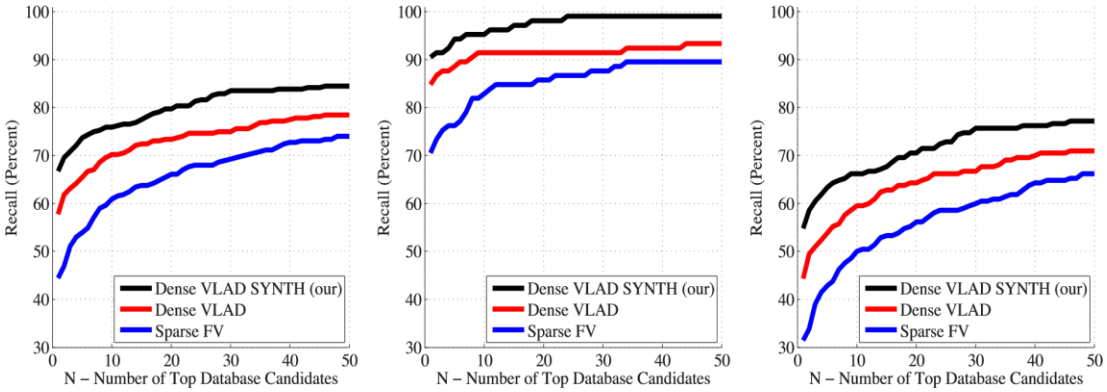
在本节中我们描述新收集的 24/7 东京数据集、 给地方识别性能的措施和概述我们相比几个基线的方法定量和定性的结果。

24/7 东京数据集。我们收集了一套新的测试的 1,125 查询图像。我们捕获图像在 125 不同地点。在每个位置我们捕获图像在 3 个不同的查看方向和在 3 个不同时期的日子，如图 5 所示。在每个位置的地面真相 GPS 坐标录由手动定位在最好的缩放级别的地图上的观察者的位置。我们估计的地面真相位置误差小于 5m。数据集是可在 [16]。在以下的评估中，我们使用的一个子集 315 查询图像内面积约 1,600m × 1,600m 由我们地理标记数据库。

评价指标。如果至少一个查询地方视为正确公认的顶部 N 检索的数据库映像是内 d = 25 米从地面真相位置的查询。这是一个共同的地方识别度量用于例如 [8、 35、 40]。正确识别查询 （召回） 百分比然后绘制的不同值 N .

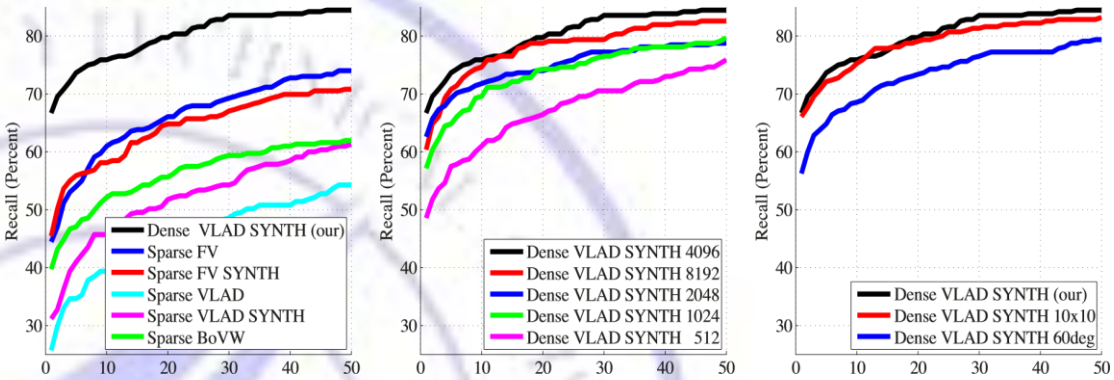
执行详细信息。为了计算密集 VLAD 描述符，我们每个调整图像的大小有最大尺寸 640 像素。这是有益的计算效率和限制提取描述符的小的规模。我们提取 SIFT [29] 描述符在 4 个尺度对应区域的 16、 24、 32、 40 像素的宽度。描述符被提取定期的密被采样网格与步幅的 2 个像素。当使用了合成图像时，我们删除与 （如黑色合成图像中所示） 没有图像数据的图像区域重叠的描述符。我们使用 [42] 跟着 SIFT 正常化 [1]，即中可用的 SIFT 实现。其次是面向元素平方根的 L1 正常化的视觉词汇 128 视觉单词（质心）由 25M 描述符随机抽样从使用 k-均值聚类的图像数据库。我们保留了原始维度的 sift 不同 [22]。每个图像然后笔下，其后 PCA 压缩至 4,096 的尺寸、 美白和 L2 正

常化［19］聚合内归一化［2］VLAD 描述符。测试查询和数据库图像之间的相似度测量的归一化的点积，可使用有效率地履行［20，31］。之后［6］，我们通过在地图上，在那里我们联想到分数由每个虚拟视图到最接近的街景视图全景执行空间非最大抑制多样化返回的入围名单。



（a） 所有查询 （b） 白天查询 （c） 日落和晚上查询

图 6。24/7-东京数据集上的评价。正确识别查询 （召回，y 轴） 与顶部的数的分数 *N* 检索数据库映像 （x 轴） 为比较基准方法 （密集弗拉德，稀疏 FV） 的拟议方法 (密集 VLAD SYNTH)。性能评价为所有测试查询图像 （a）、 以及分别白天查询 （b）、 和日落/夜查询 （c）。(致密 VLAD SYNTH) 方法的好处是最突出的困难灯饰 (c)。



N-多的顶级数据库候选人 N-数目顶级数据库候选人 N-顶级数据库候选人

（a） 与基线比较 （b） 描述符维度 （c） 视图采样密度

图 7。放在识别性能 24/7-东京数据集。每个图形显示与顶部的数正确识别查询 （召回，y 轴） 的分数 *N* 检索数据库映像 （x 轴）。

基准方法。我们向以下基线的结果进行比较。首先，我们评估基于高斯局部不变特征［29，42］差异的弗拉德描述符 (稀疏 VLAD)。在这里我们

使用直立的 SIFT 描述符在高斯要点取样，否则描述符是相同的方式构建，作为我们密被采样的弗拉德。第二，我们比较与标准稀疏费舍尔向量[22]（稀疏 FV），已被证明能很好执行位置识别 [40]。费舍尔矢量构造使用相同的直立 RootSIFT 描述符作为稀疏 VLAD 基线。[22]，经提取的 SIFT 描述符都降至 64 尺寸主成分分析法。A 256-组件混合高斯模型然后训练从 25M 随机抽样从数据库图像描述符。在 [22]，造成 256 × 64 维费舍尔向量减少到 4,096 尺寸使用 PCA，其次是美白和 L2 正常化 [19]。最后，我们也比较结果到袋视觉单词的基线。我们构建包描述符（稀疏 BoVW）使用相同的直立 RootSIFT 描述符所使用的稀疏 VLAD 基线。

200000 视觉字词汇，建立了近似 k-均值聚类 [31，33]。由此产生的袋-视觉-词向量重新加权使用自适应分配 [40]。

密集的描述符和合成的视图的好处。首先，在图 6 我们评估 （i） 密集描述符（密集 VLAD）和 （ii） 其他合成的视图（密集 VLAD SYNTH）的好处。我们比较性能与标准的费舍尔矢量描述符基于局部不变特征（稀疏 FV），被发现的地方识别 [40] 运作得很好。我们显示结果为所有查询（图 6(a)），但要清楚地说明的差异，我们也分开查询图像到白天（图 6(b)）和日落/夜查询（图 6(c)）。

图 8。示例的地方识别结果为我们的方法（密集 VLAD SYNTH）的相比，使用只有稀疏采样的特征点（稀疏 FV）的基线。（左）查询图像。（第 2 列）用我们的方法 （正确） 的最佳匹配合成视图。（第三列）最佳匹配的街道视图图像由基线 （稀疏费舍尔向量没有合成视图）。（第四列）原始的街道视图图像到查询中的最近位置。请注意，我们的方法可以匹配难查询与改变光照条件。

同时有密集的描述符（密集 VLAD）已经提高了性能相对于基线（稀疏 FV），它是组合的密集的描述符与合成虚拟视图（密集 VLAD SYNTH）带来重大改进查询与困难灯饰（图

6(c))，清楚地说明我们的方法这两个组件的重要性。

比较稀疏的基线。在图 7 （a），我们显示我们的方法（密集 VLAD SYNTH）到几个比较基准比较使用稀疏采样局部不变特征。弗拉德从关键点狗 （稀疏） 计算，为添加合成虚拟视图（稀疏 VLAD SYNTH）帮助（相比稀疏弗拉德）。与此相反的是，向费舍尔矢量匹配（稀疏 FV SYNTH）添加合成的虚拟视图，不能提高在没有虚拟视图（稀疏 FV）标准 FV。总体来看，我们的方法可大大提高在稀疏的全部基线。

描述符维数分析。 在图 7（b）我们探讨如何地方识别性能变化与进行降维的密集 VLAD 描述符从 4,096 到 2 048, 1,024 和 512 尺寸。我们观察专为最低级别的维度的性能下降。这表明，有足够丰富的表示形式是重要的匹配整个外观变化较大。

如何得到许多虚拟视图？图 7 （c） 在我们的虚拟视图所需的取样进行评估。首先，我们的虚拟空间从 5 × 5 米的网格 （到目前为止我们方法中使用） 为 10 × 10 米网格。空间采样到 10 × 10 可以减少由虚拟视图的数量 75%地方识别性能跌幅相对较小。然后我们的偏航方向只数 6 相机的位置，每一每个 60^7 相比 12 偏航方向，一个每个 30^7 在我们的方法中使用。在这个实验中我们保持空间采样 5 × 5 米。虽然角采样减少只有 50%的数字合成意见它会导致性能，尤其是在顶部 1 位置相当显著下降。

可伸缩性。24/7 的东京数据集，我们的方法合成 597,744 虚拟视图相比， 75,984 透视街景视图图像在同一地区。因此，我们的方法需要对索引 9 倍更多图片相对于基线没有虚拟视图的合成。我们相信，加大对地方识别整个城市中可以通过标准的压缩技术等产品量化 (PQ) [20]。

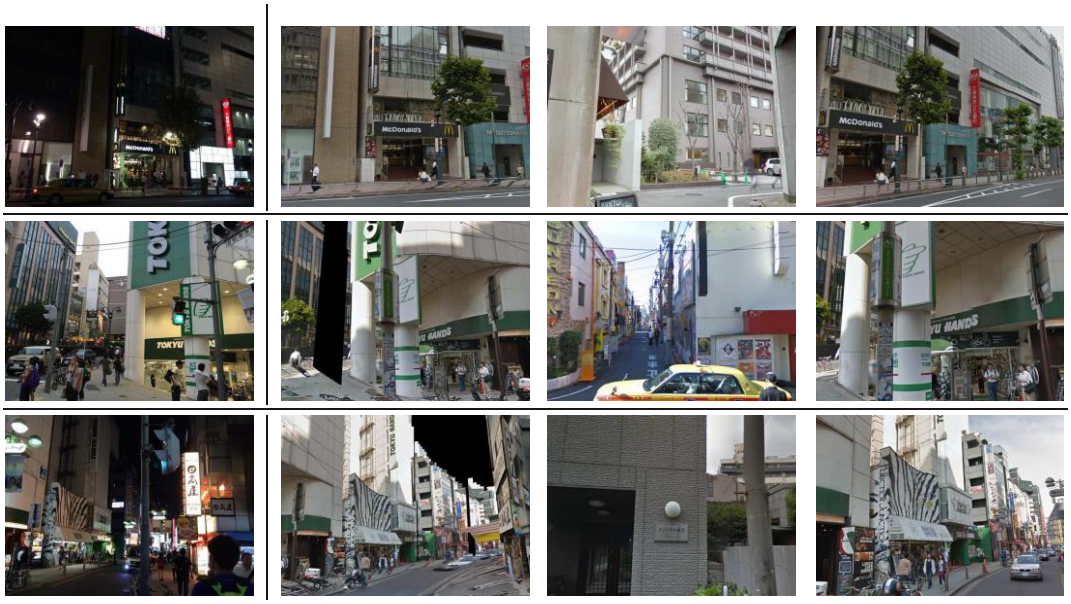


图 9。示例的地方识别结果与合成视图（我们的方法）使用只有原始的谷歌街景图像相比。（左）查询图像。注意到困难的照明.(第 2 列）最佳匹配图像（正确）我们法 （密集 VLAD 描述符与数据库扩展的综合意见）(第三列）最佳匹配图像（不正确） 的密集 VLAD 匹配但使用只有原始的街道视图图像。（第四列）原始的街景视图数据库图像到查询中的最近位置。我们的方法 （第 2 列） 使用虚拟视图非常相似的观点，到查询可以本地化查询与困难 （夜间） 照明，从而使真正的 24/7 的本地化。这是不可能使用原始的街道视图图像 （最后一列），它描述了同一个地方，而是来自完全不同的观点。请参阅其他结果上项目网页 [16]

图 10。具挑战性仍然难以本地化的查询图像的例子。

涵盖重大部分的东京市包含 1 M 透视图像。我们估计数据库的大小，从那开始但生成 9 时代更多的虚拟视图，与我们的合成方法，和压缩的结果描述符与 PQ，将仅需 2.9 GB。

定性的结果。图 8 和图 9 显示的地方示例识别结果。注意查询图像（左列） 观点和光照相比，同样的地方 （右栏） 为可用的街道视图包括较大的改变。合成的意见 （第 2 列）在新的职位大大减少观点的变化，从而使匹配整个大的光照变化，如第 3 条所述。

限制。图 10 显示仍然是非常难以进行本地化的查询的示例。典型故障模式是 （i） 非常黑暗的夜晚时间图像与有限的动态范围、 （ii） 地方与植被，这是很难唯一描述使用当前表示和 （iii） 地方视图合成会经常失败由于复杂基础三维结构不好被近似深度映射可用与街景视图图像。

6. 结论

我们描述了一种新的虚拟视图的合成结合密集采样但紧凑的图像描述符的地方识别方法。该方法使真正的 24/7 的地方识别横跨整个白天和夜晚的夜景照明的主要变化。我们新收集的地方识别数据集——*24/7 东京*——捕捉不同照明条件中的相同位置，实验展示了它的好处。我们的工作是另一个例子，在最近的趋势显示的 3D 结构为视觉识别的优势。作为我们的基础上广泛可用的谷歌街景视图图像我们工作证明了行星尺度 24/7 的地方识别的可能性。

引用

[1] R. Arandjelovic´and A. Zisserman. Three things everyone should know to improve object retrieval. In *CVPR*, 2012.

[2] R. Arandjelovic´and A. Zisserman. All about VLAD. In *CVPR*, 2013.

[3] R. Arandjelovic´and A. Zisserman. DisLocation: Scalable descriptor distinctiveness for location recognition. In *Asian Conference on Computer Vision*, 2014.

[4] M. Aubry, B. C. Russell, and J. Sivic. Painting-to-3d model alignment via discriminative visual elements. *ACM Transactions on Graphics (TOG)*, 33(2):14, 2014.

[5] A. Bosch, A. Zisserman, and X. Munoz. Image classification using random forests and ferns. In *ICCV*, 2007.

[6] S. Cao and N. Snavely. Graph-Based Discriminative Learning for Location Recognition. In *CVPR*, 2013.

[7] S. Cao and N. Snavely. Minimal Scene Descriptions from Structure from Motion Models. In *CVPR*, 2014.

[8] D. Chen, G. Baatz, et al. City-scale landmark identification on mobile devices. In *CVPR*, 2011.

[9] O. Chum, J. Philbin, J. Sivic, M. Isard, and A. Zisserman. Total recall: Automatic query expansion with a generative feature model for object retrieval. In *ICCV*, 2007.

[10] M. Cummins and P. Newman. Highly scalable appearanceonly SLAM - FAB-MAP 2.0. In *Proceedings of Robotics: Science and Systems*, Seattle, USA, June 2009.

[11] N. Dalal and B. Triggs. Histograms of Oriented Gradients for Human Detection. In *CVPR*, 2005.

[12] D. Hauagge and N. Snavely. Image matching using local symmetry features. In *CVPR*, 2012.

[13] D. Hauagge, S. Wehrwein, P. Upchurch, K. Bala, and N. Snavely. Reasoning about photo collections using models of outdoor illumination. In *BMVC*, 2014.

[14] J. Hays and A. Efros. im2gps: estimating geographic information from a single image. In *CVPR*, 2008.

[15] http://maps.google.com/help/maps/streetview/.

[16]http://www.ok.ctrl.titech.ac.jp/~torii/project/247/.

[17]A. Irschara, C. Zach, J. Frahm, and H. Bischof. From structure-from-motion point clouds to fast location recognition. In *CVPR*, 2009.

[18]N. Jacobs, S. Satkin, N. Roman, R. Speyer, and R. Pless. Geolocating static cameras. In *ICCV*, 2007.

[19]H. Je ́gou and O. Chum. Negative evidences and cooccurrences in image retrieval: the benefit of PCA and whitening. In *ECCV*, Firenze, Italy, 2012.

[20]H. Je ́gou, M. Douze, and C. Schmid. Product quantization for nearest neighbor search. *PAMI*, 33(1):117–128, 2011.

[21]H. Je ́gou, M. Douze, C. Schmid, and P. Perez. Aggregating local descriptors into a compact image representation. In *CVPR*, 2010.

[22]H. Je ́gou, F. Perronnin, M. Douze, J. Sa ́nchez, P. Pe ́rez, and C. Schmid. Aggregating local image descriptors into compact codes. *PAMI*, 34(9):1704–1716, 2012.

[23]B. Klingner, D. Martin, and J. Roseborough. Street view motion-from-structure-from-motion. In *ICCV*, 2013.

[24]J. Knopp, J. Sivic, and T. Pajdla. Avoding Confusing Features in Place Recognition. In *ECCV*, 2010.

[25]P.-Y. Laffont, Z. Ren, X. Tao, C. Qian, and J. Hays. Transient attributes for high-level understanding and editing of outdoor scenes. *ACM Trans. Graphics*, 33(4), 2014.

[26]S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *CVPR*, pages 2169–2178, 2006.

[27]F. Li and J. Kosecka. Probabilistic location recognition using reduced feature set. In *Proc. Int. Conf. on Robotics and Automation*, 2006.

[28]Y. Li, N. Snavely, D. Huttenlocher, and P. Fua. Worldwide Pose Estimation Using 3D Point Clouds. In *ECCV*, 2012.

[29]D. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.

[30]K. Matzen and N. Snavely. Scene chronology. In *ECCV*, 2014.

[31]M. Muja and D. Lowe. Fast approximate nearest neighbors with automatic algorithm configuration. In *VISAPP*, 2009.

[32]A. Oliva and A. Torralba. Modeling the shape of the scene: a holistic representation of the spatial envelope. *IJCV*, 42(3):145–175, 2001.

[33]J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman. Object retrieval with large vocabularies and fast spatial matching. In *CVPR*, 2007.

[34]T. Sattler, B. Leibe, and L. Kobbelt. Improving Image-Based Localization by Active

Correspondence Search. In *ECCV*, 2012.

[35]T. Sattler, T. Weyand, B. Leibe, and L. Kobbelt. Image Retrieval for Image-Based Localization Revisited. In *BMVC*, 2012.

[36]G. Schindler, M. Brown, and R. Szeliski. City-Scale Location Recognition. In *CVPR*, 2007.

[37]Q. Shan, C. Wu, B. Curless, Y. Furukawa, C. Hernandez, and S. M. Seitz. Accurate geo-registration by ground-to-aerial image matching. In *3DV*, 2014.

[38]D. Sibbing, T. Sattler, B. Leibe, and L. Kobbelt. SIFTRealistic Rendering. In *3DV*, 2013.

[39]J. Sivic and A. Zisserman. Video Google: A text retrieval approach to object matching in videos. In *ICCV*, 2003.

[40]A. Torii, J. Sivic, T. Pajdla, and M. Okutomi. Visual Place Recognition with Repetitive Structures. In *CVPR*, 2013.

[41]T. Tuytelaars and K. Mikolajczyk. Local invariant feature detectors: a survey. *Foundations and Trends in Computer Graphics and Vision*, 3(3):177–280, 2008.

[42]A. Vedaldi and B. Fulkerson. VLFeat: An open and portable library of computer vision algorithms. `http://www.vlfeat.org/`, 2008.

[43]C. Wu, B. Clipp, X. Li, J.-M. Frahm, and M. Pollefeys. 3D model matching with viewpoint-invariant patches (VIP). In *CVPR*, pages 1–8, June 2008.

[44]A. R. Zamir and M. Shah. Accurate Image Localization Based on Google Maps Street View. In *ECCV*, 2010.

[45]W. Zhao, H. Je ́gou, and G. Gravier. Oriented pooling for dense and non-dense rotation-invariant features. In *BMVC*, 2013.