Improving Keyppoint Matching Using a Landmark-Based Image Representation

方法介绍

本方法有以下四步组成：

1. 提取两张图片中匹配的关键点作为地标
2. 计算两张图中的ConvNet特征用来检测和匹配地标
3. 在每对匹配的地标中匹配关键点，并使用修剪技术修剪匹配的关键点（以生产高质量的推定匹配项）
4. 将地标的所有匹配关键点组合为两个图像之间的推定匹配，并使用MVG和RANSAC识别内部匹配

选取作为地标的对象

使用BING方法提取地标，因为BING方法曾经呗看作为state-of-the-art物体检测算法。相比于其他方法，BING的执行速度更快，在作者的笔记本上能做到24ms/帧，在实验中，每张图片提取100个地标。

ConvNet特征作为地标描述子

使用预先训练的卷积神经网络就如AlexNet为每个地标生成特征向量或者描述子。

AlexNet包括五个卷积层，每个层都followed一个非线性激活函数。AlexNet同时有三个全连接层和最终的一个soft-max层。研究表明，较为中间的层中就像AlexNet的第三卷积层上表现出对图片中形状改变时提取出特征具有不变性。因此使用Conv3去提取地标的描述子。对于每个地标提取区域，resize it to 224 * 224 * 3，将其输入到AlexNet网络（ConvNet3），从而获得一个64890维的特征向量。

计算两个地标表述子的余弦距来确定两个地标相同：

$$d_{ij} = \frac{< l_i^a, l_j^b >}{|l_i^a||l_j^b|}$$

当两个地标满足互为最近邻时，则可认为这是一对真实的匹配。

另外使用形状相似去来过滤地标对。匹配的地标对的两个相应的边界框必须在几何意义上足够的相似：

$$max(\omega_i^a, \omega_j^b) \leq r \times min(\omega_i^a, \omega_j^b)$$

$$max(h_i^a, h_j^b) \leq r \times min(h_i^a, h_j^b)$$

w_i^a：宽；h_i^a：高，r是一个阈值，文中给出为$r = 1.3$

使用匹配的地标对来匹配关键点

由于更大的边界框会导致更多的匹配错误，所以需要首先限制一下边界框的大小从而限制地标对的匹配：

$$w_i^a \leq s \times w^a$$

$$h_i^a \leq s \times h^a$$

本文中$s = 0..6$，w^a是图片I^a的宽，h^a是图片I^a的高。

对于每对匹配的地标，将它们视为由其边界框定义的图像块。随后，将其关键点进行匹配，以生成针对地标对的一组假定关键点匹配。真正的积极关键点匹配必然来自匹配的地标，并且图像的非标志性区域往往质地较差并且不太可能产生易于匹配的独特关键点。

实验部分

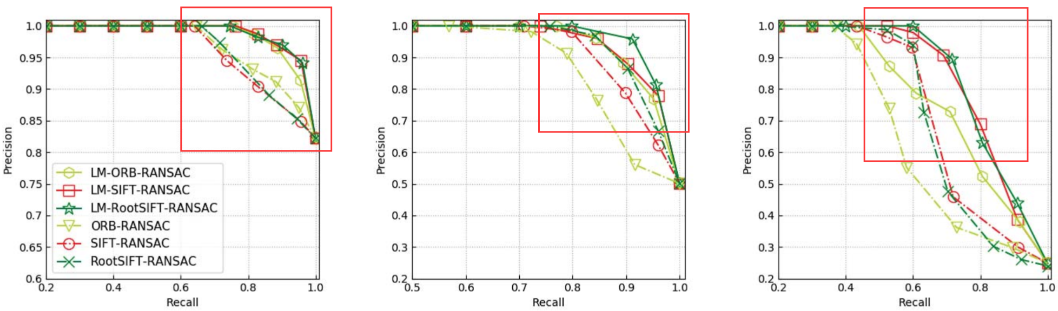
回环检测

- 使用数据集：UACampus，Mapillary
- 回环检测算法：CNN-based，GIST-based
- 实验设置：

TABLE I: Main properties of testing datasets used in our experiment. **Number** means the number of images in the corresponding subset. Note that **Potential loop closure dataset** indicates the loop closure image pairs generated by GIST-based and CNN-based loop closure detection algorithms on UACampus and Mapillary dataset.

Dataset	Number	Potential loop closure dataset		Main changes	
				Illmination	Viewpoint
UAcampus [14]	647 647	UACampus-GIST	UACampus-CNN	large	minor
Mapillary [23]	1300 1300	Mapillary-GIST	Mapillary-CNN	moderate	large

- 结果：



(a) UACampus-GIST (b) Mapillary-CNN (c) Mapillary-GIST

Fig. 4: Comparisons of competing methods in loop closure verification accuracy in terms of the precision-recall curve on four datasets of loop closure hypotheses. Note that the result of UACampus-CNN is not shown due to the almost perfect result and that the range of the y-axis of Fig. 4(a) is different from that of (b) and (c).

TABLE II: Loop closure verification accuracy of **LM-ORB-RANSAC** vs. **ORB-RANSAC**, **LM-SIFT-RANSAC** vs. **SIFT-RANSAC** and **LM-RootSIFT-RANSAC** vs. **RootSIFT-RANSAC** in terms of maximum recall at 100% precision (Re. at 100% Pr.) and average precision (AP). The highest value with respect to each metric on each dataset is highlighted in bold. The middle values are the differences between **LM-ORB-RANSAC/ORB-RANSAC**, **LM-SIFT-RANSAC/SIFT-RANSAC** and **LM-RootSIFT-RANSAC/RootSIFT-RANSAC**.

Method	UAcampus-CNN		UAcampus-GIST		Mapillary-CNN		Mapillary-GIST	
	Re. at 100% Pr.	AP	Re. at 100% Pr.	AP	Re. at 100% Pr.	AP	Re. at 100% Pr.	AP
LM-ORB-RANSAC	**93.16%**	**99.81%**	**75.37%**	**98.10%**	**77.09%**	**96.16%**	**43.87%**	**78.64%**
	+13.83%	*+0.14%*	*+10.92%*	*+1.32%*	*+20.26%*	*+5.07%*	*+7.10%*	*+11.34%*
ORB-RANSAC	79.33%	99.67%	64.45%	96.78%	56.83%	91.09%	36.77%	67.30%
LM-SIFT-RANSAC	**89.60%**	**99.79%**	**76.23%**	**98.87%**	**74.01%**	**96.52%**	**52.26%**	**84.51%**
	+14.08%	*+0.12%*	*+12.20%*	*+2.91%*	*+3.08%*	*+1.96%*	*+9.03%*	*+10.04%*
SIFT-RANSAC	75.52%	99.67%	64.03%	95.96%	70.93%	94.56%	43.23%	74.47%
LM-RootSIFT-RANSAC	**94.44%**	**99.81%**	**74.52%**	**98.81%**	**79.74%**	**97.38%**	**60.00%**	**85.15%**
	+7.16%	*+0.10%*	*+8.35%*	*+2.68%*	*+3.97%*	*+2.29%*	*+22.59%*	*+11.57%*
RootSIFT-RANSAC	87.28%	99.71%	66.17%	96.13%	75.77%	95.09%	37.41%	73.58%

可明显看出LM-系列算法鲁棒性更强

与传统方法对比

本文方法：LM-ORB-RANSAC、LM-SIFT-RANSAC、LM-RootSIFT-RANSAC

对比方法：ORB-RANSAC、SIFT-RANSAC、RootSIFT-RANSAC

结果：同上

总结

提出了一种有效的关键点匹配方法。简单来说，使用BING从两个关键点匹配的图像中提取地标方案。然后，为检测到的界标计算ConvNet特征，并在两个图像之间匹配界标。该方法在多个数据集上进行验证能有明显优于标准方法。