

密级：公开 保密期限：

北京邮电大学

硕士学位论文



题目： 基于局部特征的图像重建算法研
究

学 号： 2012110191

姓 名： 王继哲

专 业： 信号与信息处理

导 师： 李学明

学 院： 信息与通信工程学院

二〇一四年十二月

独创性（或创新性）声明

本人声明所呈交的论文是本人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢中所罗列的内容以外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得北京邮电大学或其他教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

申请学位论文与资料若有不实之处，本人承担一切相关责任。

本人签名：_____ 日期：_____

关于论文使用授权的说明

学位论文作者完全了解北京邮电大学有关保留和使用学位论文的规定，即：研究生在校攻读学位期间论文工作的知识产权单位属北京邮电大学。学校有权保留并向国家有关部门或机构送交论文的复印件和磁盘，允许学位论文被查阅和借阅；学校可以公布学位论文的全部或部分内容，可以允许采用影印、缩印或其它复制手段保存、汇编学位论文。（保密的学位论文在解密后遵守此规定）

本学位论文不属于保密范围，适用本授权书。

本人签名：_____ 日期：_____

导师签名：_____ 日期：_____

基于局部特征的图像重建算法研究

摘 要

中、英文摘要位于声明的次页，摘要应简明表达学位论文的内容要点，体现研究工作的核心思想。重点说明本项科研的目的和意义、研究方法、研究成果、结论，注意突出具有创新性的成果和新见解的部分。

关键词是为文献标引工作而从论文中选取出来的、用以表示全文主题内容信息的术语。关键词排列在摘要内容的左下方，具体关键词之间以均匀间隔分开排列，无需其它符号。

关键词：重建 局部特征 图像分割 图像配准 图像融合

RESEARCH ON IMAGE RECONSTRUCTION ALGORITHM BASED ON LOCAL FEATURES

ABSTRACT

The Chinese and English abstract should appear after the declaration page. The abstract should present the core of the research work, especially the purpose and importance of the research, the method adopted, the results, and the conclusion.

Key words are terms selected for documentation indexing, which should present the main contributions of the thesis. Key words are aligned at the bottom left side of the abstract content. Key words should be separated by spaces but not any other symbols.

KEY WORDS: T_EX L^AT_EX x_eCJK template typesetting thesis

目 录

第一章 绪论	1
1.1 中文信息处理软件的国内外发展现状	2
1.2 本说明的主要内容	3
参考文献	3
第二章 基于局部特征的图像重建算法概述	5
2.1 传统的图像重建算法	5
2.2 基于局部特征的图像重建算法	6
2.3 图像的局部特征	6
2.3.1 局部特征概述	6
2.3.2 SIFT	7
2.4 图像配准	9
2.4.1 2D 几何变换	9
2.4.2 RanSAC	10
2.5 图像分割	10
2.5.1 基于图的图像分割算法	10
2.6 图像融合	14
参考文献	14
第三章 大规模近似重复图像搜索算法概述	17
3.1 传统的大规模图像检索方法	17
3.1.1 近似重复图像搜索算法概述	17
3.1.2 图像表示与相似性度量	17
3.1.3 基于局部特征的相似图像搜索算法	17
3.2 改进的相似搜索算法	17
3.2.1 sim-hash	18
3.2.2 最小哈希的相似性比较	18
3.2.3 LSH	19
3.3 空间信息匹配搜索算法	20

3.3.1 随机抽样一致算法 20

3.3.2 视觉词组 22

3.4 适用于旅游景点图像的相似图像搜索技术 22

参考文献 22

附录 A 不定型 (0/0) 极限的计算 25

附录 B 缩略语表 27

致 谢 29

攻读学位期间发表的学术论文目录 31

符号对照表

$(\cdot)^*$	复共轭
$(\cdot)^T$	矩阵转置
$(\cdot)^H$	矩阵共轭转置
\mathbf{X}	矩阵或向量
\mathcal{A}	集合
$\mathcal{A} \times \mathcal{B}$	集合 \mathcal{A} 与集合 \mathcal{B} 的 Cartesian 积, 即 $\mathcal{A} \times \mathcal{B} = \{(a, b) : a \in \mathcal{A}, b \in \mathcal{B}\}$

第一章 绪论

北京邮电大学北京邮电大学 (Beijing University of Posts and Telecommunications, BUPT) 研究生院培养与学位办公室于 2010 年 3 月 1 日颁布了最新的《北京邮电大学关于研究生学位论文格式的统一要求》(以下简称“要求”)[1], 对原有研究生学位论文的格式要求做出了新的修订。但是迄今为止, 研究生院尚未发布统一的论文模板。对于已经、正在或者即将撰写学位论文的同学都只能按照该要求的规定自行调整其学位论文的格式, 一方面给大家增加了繁重的排版工作, 另一方面也不利于统一全校的论文格式。

2007 年 9 月, 北京邮电大学无线新技术研究所无线新技术研究所 (Wireless Technology Innovation Institute, WTI) 的王旭博士制作并发布了 latex-bupt——北京邮电大学博士毕业论文 L^AT_EX 模板 (非官方版)[2]。该模板可以满足旧版官方论文格式要求[3], 但是在一些细节上的处理还有待改进, 例如:

- 参考文献不能分列在各章末尾;
- 不能利用 BiBTeX 处理发表学术论文列表;
- 参考文献的格式上不能完全满足学校要求等。

2009 年, 张煜博士发布了 bupththesis——北京邮电大学研究生学位论文 L^AT_EX 文档类 (非官方版)[4]。该模板解决了 latex-bupt 中存在的问题, 并且同样可以满足旧版官方论文格式要求[3], 但是仍然存在以下一些问题可以改进:

- 论文格式与最新版的官方论文格式要求[1]有细微出入;
- 中文解决方案采用旧式 CJK 宏包, 需要用户自行生成字体;
- 缺乏详细的用户使用文档, 用户撰写论文过程中遇到的问题基本都需要登陆北邮人论坛发问, 由张博逐一解答。

本模板在 bupththesis[4] 的基础上, 增加了 XeTeX 编译引擎, 使用 xeCJK 宏包作为中文解决方案。同时, 本模板还根据北京邮电大学发布的最新的论文格式要求 [1] 进行模板格式的修改。本模板还提供了较为细致的用户使用文档, 可以帮助初级用户快速上手使用本模板。

1.1 中文信息处理软件的国内外发展现状

中文信息处理软件可以分为字处理软件和排版软件两大类。字处理软件包括以下功能：字体、字号设定，英文断字，拼写和语法检查等。通常字处理软件处理文档的规模比较小，一般是作为办公自动化套件的一个重要组成部分，目前广泛使用的中文字处理软件主要包括微软 Office 套件中的 Word、金山公司的 WPS，以及开源社区的 OpenOffice 等。排版软件则是针对大规模专业出版印刷而设计的一类软件，其主要功能是文字图像定位，基本图形绘制等。排版软件相对于字处理软件其专业针对性更强，目前广泛使用的中文排版软件主要包括北大方正的书版系列软件，飞腾系列软件，蒙泰桌面出版系统，Adobe 公司的 PageMaker，FrameMaker，以及 QuarkPress 公司的 PassPort 等。除此而外，由 D. E. Knuth 编写的 $\text{T}_{\text{E}}\text{X}$ 和由 L. Lamport 编写的 $\text{L}_{\text{A}}\text{T}_{\text{E}}\text{X}$ 也是学术界广泛的应用排版软件。

微软公司的 Word 是目前国内最为普及的字处理软件之一，也是大多数学校规定的学位论文编辑排版工具。不容否认，Word 在简单文书（例如：通知、简报等）编辑排版方面具有方便快捷的优势，而且其对多人协同编辑的支持也给文字修订工作带来了极佳的用户体验。但是从实际使用的情况看，尽管 Word 已经经历了第 12 个版本的改进，但是其对于处理大型文书文稿（例如：书籍、学位论文等）的能力仍然有待进一步完善和提高。由于 Word 版本不兼容造成的来回反复，也是使用 Word 编辑文字稿件的烦事之一。另外，由于 Word 对数学公式编辑的支持一直延续其“对象链接与嵌入”（Object Linking and Embedding, OLE）的设计理念，这也使得每位使用 Word 排版过理工类的文字资料的人都有一段或多段刻骨铭心的痛苦经历，往往花在调整格式这种 dirty work 上的时间和花在编写文章内容上的时间差不多或者甚至更多。

北大方正的书版系列软件是专业中文出版领域的权威，国内几乎所有的大型出版社、报社、政府机关几乎都使用书版系列软件对其出版的书籍、报纸和公文进行编辑排版。但是，书版软件作为方正电子出版流程中的一个主要组成部分，主要定位于印前排版环节，面向专业排版工作人员。因此，学习和使用使用书版软件需要花费较长的时间来熟悉复杂的排版命令，发排后需要使用专用的 RIP 软件或者方正的专用打印机才能输出样张等。

美国 Stanford 大学的荣誉退休教授 D. E. Knuth 在 197x 年独自一人开发了 $\text{T}_{\text{E}}\text{X}$ 排版系统，随后，L. Lamport 为 $\text{T}_{\text{E}}\text{X}$ 编写了一系列的宏包使得 $\text{T}_{\text{E}}\text{X}$ 的使用更加方便，这些宏包被称为 $\text{L}_{\text{A}}\text{T}_{\text{E}}\text{X}$ 。自从 $\text{T}_{\text{E}}\text{X}/\text{L}_{\text{A}}\text{T}_{\text{E}}\text{X}$ 问世以来它们就受到了学术界的青睐，目前

几乎所有的国外出版社都接受或指定使用 $\text{T}_\text{E}\text{X}/\text{L}_\text{A}\text{T}_\text{E}\text{X}$ 对稿件进行排版编辑。19xx 年, 中国科学院的张林波研究员开发了 CCT 使得 $\text{L}_\text{A}\text{T}_\text{E}\text{X}$ 可以用于中文文稿的处理。德国的 W. Lemberg, 编写了 CJK 宏包为 $\text{L}_\text{A}\text{T}_\text{E}\text{X}$ 提供了中日韩三国语言的解决方案。使用 $\text{T}_\text{E}\text{X}/\text{L}_\text{A}\text{T}_\text{E}\text{X}$ 排版学术论文的最大优势在于, 它让作者可以不用为排版输出的具体格式操心, 而全心投入文章、书稿内容的编写上, 最大程度的降低作者从事排版 dirty work 的工作量。

目前, 我国的清华大学、哈尔滨工业大学、西安电子科技大学、西安交通大学等都已经纷纷制作了本校学位论文的 $\text{L}_\text{A}\text{T}_\text{E}\text{X}$ 模板, 并接受使用 $\text{L}_\text{A}\text{T}_\text{E}\text{X}$ 排版的学位论文。

1.2 本说明的主要内容

本说明全面介绍了如何使用 BUPTGraduateThesis 来排版符合 [1] 规定的北京邮电大学学位论文。全文内容安排如下:

1. 第二章介绍……
2. ……

参考文献

- [1] 北京邮电大学研究生院培养与学位办公室. 关于研究生学位论文格式的统一要求 [EB/OL]. 2010 [2013-07-06]. <http://www.bupt.edu.cn/>.
- [2] 王旭. latex-bupt: $\text{L}_\text{A}\text{T}_\text{E}\text{X}$ style for BUPT thesis [EB/OL]. 2007 [2009-01-14]. <http://code.google.com/p/latex-bupt/>.
- [3] 北京邮电大学研究生院培养与学位办公室. 关于研究生学位论文格式的统一要求 [EB/OL]. 2010 [2013-07-06]. <http://www.bupt.edu.cn/>.
- [4] 张煜. 北京邮电大学研究生学位论文 $\text{L}_\text{A}\text{T}_\text{E}\text{X}$ 文档类 [EB/OL]. 2009 [2013-07-06]. <http://code.google.com/p/buptthesis/>.

第二章 基于局部特征的图像重建算法概述

图像重建可以被概括的定义为这样一个基本问题：从一个退化版本的二维物体估算实际的二维物体^[1]。退化过程的数学形式取决于图像重建算法实际的应用场景。

2.1 传统的图像重建算法

传统的图像重建算法所使用的场景一般是指图像修复（Image Restoration），原始“物体”由于经历了某种退化过程，不能直接由观测信息判断出来，为了消除退化过程的影响，必须根据观测到的数据进行重建来还原得到原始信息。在图像修复中，引起退化的原因叫做失真，其定义如下：

$$y = A(X) \bullet b \quad (2-1)$$

其中 $A(\cdot)$ 是退化函数，可以看做是一个滤波器， b 表示的是噪声， \bullet 表示的叠加方式。失真通常包含对 X 的卷积或者模糊，加性噪声或者乘性噪声。而图像修复的解决方案是通过对观测信息进行退化模型的数学建模，利用约束条件来推导出退化过程的逆过程，对观测信息进行逆过程得到原始图像。

另一类图像重建场景是超分辨率重建，在近年来得到飞速的发展，是炙手可热的研究领域，它的基本思想是通过多张连续的低分辨率图像序列得到一张高分辨率的图像。很多数字图像应用中都需要高分辨率的图像，高分辨率的图像能够提供更好的视觉体验，提供更丰富的信息，比如高分辨率的医学图像能够让医生更好的进行病情诊断，高分辨率的卫星图像能够进行更准确的模式识别任务。从1970年代以来，CCD和CMOS传感器被大规模的使用，获得了大量的数字图像，但是很多图像的分辨率较低，不能满足日益增长的业务需求，超分辨率重建是在这样的背景下诞生的^[2]。

那么，我们如何通过多张低分辨率图像获得一个高分辨率图像呢？如果一个场景下有多张低分辨率图像，而且这些图像从不同的角度来“描述”这个场景，那么这些低分辨率的图像可以看做是该场景的子采样和子像素精度的位移。如果这些低分辨率图像是以整数像素为单位进行的位移，那么多张低分辨率图像没有提供任何“新的信息”，但是如果位移单位是子像素单位的，序列中的每一个图像不能够由其

他图像得出，换言之每个图像都提供了子像素精度的不同信息，我们可以利用这些信息重建一个高分辨率的图像。一般来说，SRR 算法分为基于重建和基于学习的两大类：基于重建的算法如频域重建法利用图像序列的交叠关系，凸集投影（POCS）等利用一些先验知识来约束求解过程，以达到增加细节信息的目的；基于学习的算法则使用多种机器学习的概率模型，包括基于流形学习、基于支持向量机和基于独立分量的超分辨率重建技术。基于学习的方法采用大量的高分辨率图像构造学习库来训练学习模型，在对低分辨率图像进行重建的过程中引入由学习模型获得的先验知识，进而得到图像的高频细节，获得较好的图像重建效果。

总体而言，超分辨率重建的整个流程包括三个基本环节：（1）低分辨图像的预处理，包括降噪和裁剪等基本图像数据处理。（2）配准过程，利用像素的空间信息估算低分辨率序列图像之间的运动矢量和空间位置关系。（3）完成重建，使用图像分割和融合等技术，利用多帧低分辨率图像的信息完成超分辨率重建。

2.2 基于局部特征的图像重建算法

从另一个角度来看，本文所采用的图像重建的部分流程可以看成是多幅图像的全景图拼接问题。与文献^[3]中的流程类似，主要包含以下几个环节：（1）使用具有不变性的特征来描述图像；（2）自动的找到图像之间的空间位置关系，进行图像配准；（3）图像融合，消除不同图像之间的光照差别，去除边缘噪声。与全景图拼接不同的是，本文提出的图像重建系统每幅小图（Patch）块大小不一，导致图像空间位置关系可能存在不准确，多幅图像之间有大量重叠，整体思路是先融合大块，后融合小块，分别介绍如下。

2.3 图像的局部特征

2.3.1 局部特征概述

图像的局部特征是计算机视觉领域一个基本问题，它能够反映图像某一局部的特性，对寻找图像对应的局部单元以及特征描述中有着重要作用。通常意义的局部特征包含两个方面，特征检测子（Detector）和特征描述子（Descriptor）。检测子能够检测出我们“感兴趣”的点或者局部区域，而一个好的局部特征描述子反映出图像的局部特性能够帮助找到图像与图像点集合对应关系，进而建立图像之间的空间对应关系。局部图像特征描述的核心问题是不变性（invariant）和可区分性

(discrimination)。

目前人们提出的众多图像局部特征中算子中，由 Lowe 提出的尺度不变特征变换（Scale Invariant Feature Transform，简称 SIFT）应用最为广泛。1999 年首次提出，至 2004 年得到完善^[4] 的 sift 算子是图像局部特征研究领域的一项重大突破。sift 算子具有很强的可区分性，同时对尺度、旋转以及一定视角和光照变化等图像变化都具有不变性。在其之上衍生出来的 SURF（Speeded Up Robust Features）是对 SIFT 的改进版本，它利用 Haar 小波来近似 SIFT 方法中的梯度操作，同时利用积分图技术进行快速计算，SURF 的速度是 SIFT 的 3-7 倍。

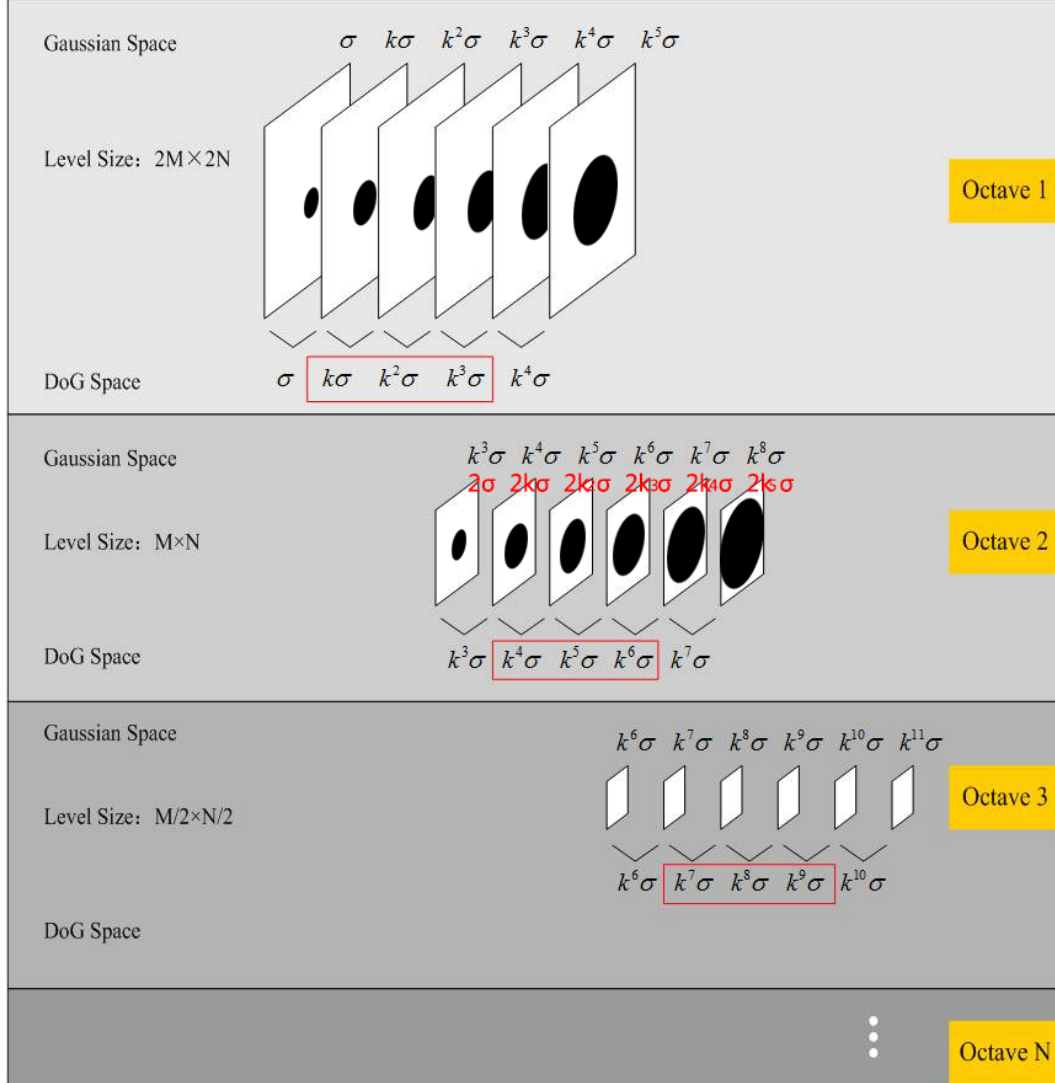
除此以外，常见的特征检测子包括 Harris 角点，ANMS 等，描述子还包括 DAISY，ASIFT，MROGH，BRIEF 等，分别适用于不同的图像应用场景下，本文提出的系统采用适用性最广泛的 sift 算子，下面我们对其进行简要的介绍。

2.3.2 SIFT

（1）尺度空间理论尺度空间理论目的是模拟图像数据的多尺度特征，尺度空间中各尺度图像的模糊程度逐渐变大，能够模拟人在距离目标由近到远时目标在视网膜上的形成过程。我们可以把两幅图像想象成是连续的，分别以它们作为底面作四棱锥，就像金字塔，那么每一个截面与原图像相似，那么两个金字塔中必然会有包含大小一致的物体的无穷个截面，但应用只能是离散的，所以我们只能构造有限层，层数越多当然越好，但处理时间会相应增加，层数太少不行，因为向下采样的截面中可能找不到尺寸大小一致的两个物体的图像。一个图像的尺度空间， $L(x,y,\sigma)$ 定义为一个变化尺度的高斯函数 $G(x,y,\sigma)$ 与原图像 $I(x,y)$ 的卷积。

$$L(x,y,\sigma) = G(x,y,\sigma) \otimes I(x,y) \quad (2-2)$$

下面这幅图反映了图像金字塔的情况：



图中的黑色圆盘是我加上去的，表示的是该图像所在的尺度的特征覆盖的范围，其特点是不同组同一层上的特征覆盖范围一样，同一组不同层上的特征覆盖范围逐步增大。关键点的尺度坐标就是按关键点所在的组和组内的层，利用下面这个公式计算而来：

$$\sigma(o, s) = \sigma_0 2^{o+s/S}, \quad o \in o_{\min} + [0, \dots, O-1], \quad s \in [0, \dots, S-1] \quad (2-3)$$

(2) SIFT 检测子 sift 算法有两个主要环节，一个是检测“感兴趣”的关键点，另一个是描述这个“关键点”。SIFT 关键点是精心选择的一组在高斯差分尺度空间 (Difference of Gaussians scale space, DoG) 尺度空间上的极值点，该关键点包含三个关键信息，分别是 (1) 亚像素精度的 (x, y) 位置信息；(2) 尺度大小，反映关键

点局部的大小，同时决定了特征的覆盖范围，对后文局部块的提取起到至关重要的作用；（3）所在高斯尺度空间上的主方向，该主方向是有一个高斯窗口函数计算得来，反映的是关键点所在局部的方向信息。其中差分高斯尺度空间表示为

$$D(x, \sigma(s, o)) \doteq G(x, \sigma(s+1, o)) - G(x, \sigma(s, o)). \quad (2-4)$$

关于尺度空间和描述子的具体讲解在 Lowe 的论文中^[4] 已有详细的介绍，这里不再详述。

（3）SIFT 描述子 SIFT 描述子反映关键点局部的信息，是高斯尺度空间上某一局部和方向上的梯度信息，以直方图的形式对信息做统计，最终每一个描述子是一个 128 的特征。

2.4 图像配准

先匹配，根据匹配的特征点对进行配准。

2.4.1 2D 几何变换

（1）旋转和平移变换, 也叫 2D 刚体运动即 2D 欧式变换（因其保持欧式距离），写作 $x = Rx + t$ 或者写作

$$\begin{bmatrix} R & t \end{bmatrix} \bar{x} \quad (2-5)$$

其中

$$\begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \quad (2-6)$$

是一个正交旋转矩阵，有 $RR^T = I$ 和 $|R| = 1$

（2）放缩旋转，也叫做相似变换，该变换可以表示为 $\bar{x} = sRx + t$ ，其中 s 是一个任意的尺度因子。它也可以写作

$$x = \begin{bmatrix} sR & t \end{bmatrix} \bar{x} = \begin{bmatrix} a & -b & t_x \\ b & a & t_y \end{bmatrix} \bar{x} \quad (2-7)$$

其中我们不再需要 $a^2 + b^2 = 1$ 。相似变换保持直线间的夹角。各种 2D 变换如下表所示：

当我们使用 sift 算法得到匹配到的特征点后，有两种方法，一种是直接写出变换矩阵，另一种是使用 RANSAC 方式多次迭代找到最准确的变换矩阵，两种变换方法 TODO

2.4.2 RanSAC

2.5 图像分割

本文主要采用的是基于图的图像分割算法

2.5.1 基于图的图像分割算法

1、背景介绍

主要参考的是这篇文章《Efficient Graph-Based Image Segmentation》Pedro F.Felzenszwalb

文章首先自己定义一种区域边界的度量方法，其度量方法是在基于图的图像表示法之上去定义的。在这种度量方法之上，我们衍生出来比较高效的图像分割算法。该算法是一种贪婪算法，并且分割结果满足全局属性。

通过上述逻辑关系可以看出，文章定义的度量方法需要比较准确，有一定的物理意义在里面，不然即使算法再高效，度量本身有问题，那么分割出来的图像区域也是不准确的。那么文章自然的可以分为两个部分 1) 区域度量方法。2) 高效分割算法

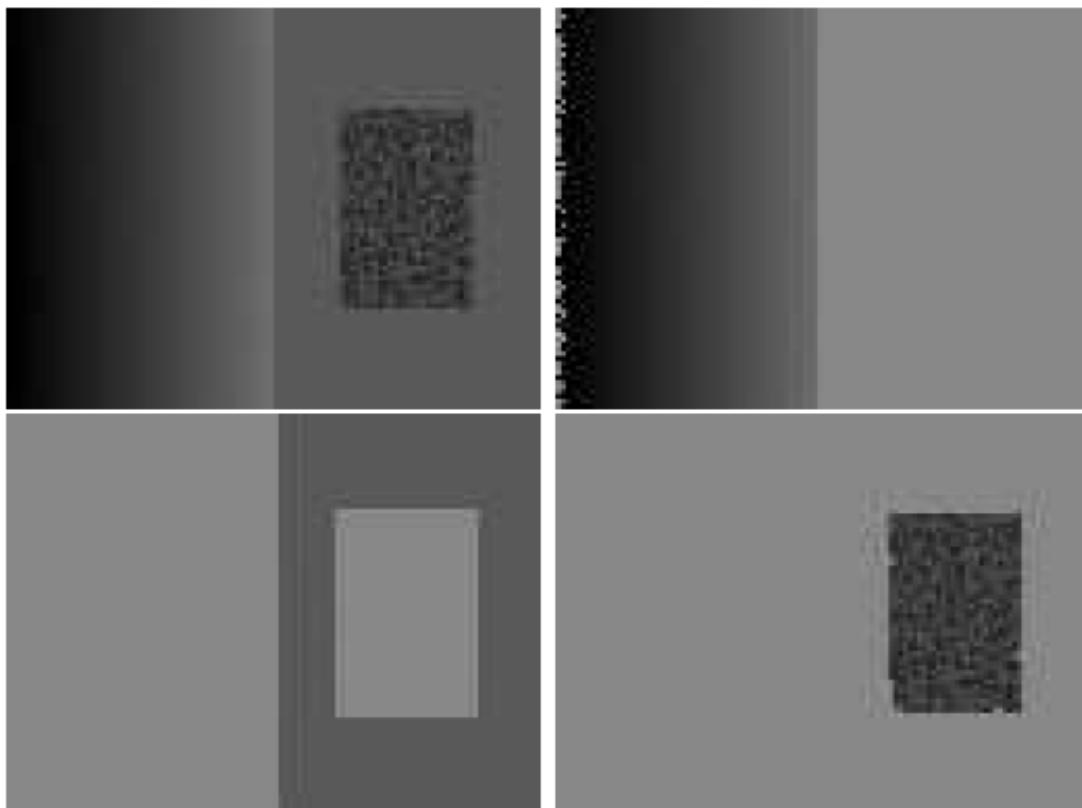
图像分割的在很多应用中非常重要，是很多高层应用的前提，比如识别、索引等，我们不具体举例。我们认为图像分割的方法有下面这样的特性：

- 能够捕捉到感知上比较重要的区域，这通常体现在图像的全局特性方面。这里有两个关键点，一方面要提供感知重要的精准属性，另一方面能够确定给定的分割技术是做什么的。我们认为应该有对分割结果属性的经确定以，这样的方法才能够更好的被理解，进而与其他的方法进行比较。
- 高效，接近图像像素点数量的线性时间复杂度。为了能够实际使用，我们认为分割方法应该与边缘检测或者其他 low-level 图像处理技术有着相似的时间复杂度，意味着时间复杂度是线性，而且常系数也比较小。比如每一秒能对几帧图像进行分割的算法就能够处理实时的视频数据。

然而，近几年的一些方法并不能够达成上述两方面要求，哪些方法太慢以致不能实践中使用。相比较而言，本文提到的方法已经因公在大尺度图像数据集应用上。

有一些其他的方法可以比较快速的进行图像分割，但是这些方法不能捕捉感知上重要的非局部特性，下文会提到。总而言之，本文在保证效率的同时考虑到了图像全局属性上的感知重要区域。

首先我们来看一幅人造图像：



我们人眼会认为这幅图像有三个区域，这个例子能够解释什么是感知重要属性 (conceptually important property)。首先，亮度的变化不应该单独的座位分割区域的衡量标准。比如图像中左侧渐变区域和右侧的高频噪声区域都有较大的亮度变化，但是我们他们应该被分割成多个区域。因此，假设一个区域有着接近恒定的或变化很小的亮度是不正确的。

第二个感知重要属性是有意义的区域不能单纯的依靠局部划分标准。还是在图上我们可以看到原因，渐变图像与常量区域的边界上的亮度差值比很多高频区域的差值要小，因此我们得出结论，为了分割一幅图像，我们需要引入一些适应性的或者非局部的衡量标准。

我们在下一节提出的衡量标准会比较两个属性：

- 边界的亮度差值

- 区域内部的邻居像素间的亮度差值

直观上，两个区域的边界上的亮度差值如果比比两个区域中至少一个区域的内部像素差值大的话，那么边界亮度差值会更多的影响我们的感知，这个时候我们说边界亮度差是感知重要的。

2、基于图的图像表示

好下面我们来进入正题，基于图的图像分割（Graph-Based Segmentation）。我们使用基于图的方法来做图像分割，令 $G = (V, E)$ 表示一个无向图，点集 $v_i \in V$ ，待分割的元素集合。边 $(v_i, v_j) \in E$ 有一个相应的权重 $w((v_i, v_j))$ ，是一个非负值，描述两个相邻元素 v_i 和 v_j 的不相似度。在图像分割，也就是本文的语境下， V 中的元素就是像素点，边就是它的两个像素点（这两个像素点是相邻的）不相似性的某种度量（例如亮度，颜色，运动，位置或者其他局部属性）。在文章的最后我们会讨论比较特殊的边集合和权重函数，不过这里的公式和不相似性度量的方法是独立的，我们可以按照自己的需求定制度量方案，这里讨论的是大框架。

在基于图的方法中，一个分割方案 S 是 V 的一个划分，每一个区域 (region or component) $C \in S$ 对应着图 $G' = (V, E')$ 的一个连通区域，其中 $E' \subseteq E$ 。有许多方法来衡量一个分割的好坏，大体上我们希望一个区域内部的元素尽可能相似，不同区域之间的像素尽可能不同。这意味着同一区域内，相邻两个点的有相对来说比较小的权值，不同区域的相邻两个点的边有大的权值。

3、成对的区域比较预测，内部不相似度与外部不相似度

这一节我们首先定义一个预测， D ，来估计是否存在一个显著的证据表明有一个边界能将两个区域分割开。就像上文说的，就是对外部的不相似性与内部不相似进行比较，也就是比较 inter-component 和 within component 的差值。

我们定义内部不相似性为该区域最小生成树的最大边， $MST(C, E)$ ，即：

$$Int(C) = \max_{e \in MST(C, E)} w(e) \quad (2-8)$$

这个方法潜在的直觉是一个区域 C ，它保持连通的最低要求是 $Int(C)$ 这个 edge 所决定的。

定义两个区域的不同：区域 $C_1, C_2 \subseteq V$ ，连接这两个区域的所有边的权值中，最小的那个权值。即，

$$Dif(C_1, C_2) = \min_{v_i \in C_1, v_j \in C_2, (v_i, v_j) \in E} w((v_i, v_j)) \quad (2-9)$$

如果两个区域没有连接的边，则令 $Dif(C_1, C_2) = \infty$ 这个定义理论上可能会有问题，因为它只反映了（或者说只考虑到了）两个区域间权值最小的那条边。在实践中我们发现尽管有显著的局限，但这种度量方式结果颇佳。值得一提的是，改变这个衡量标准也是可以的，比如采用中位数或者其他分位点，提升对异常值的鲁棒性，但这种改变会使问题编程 NP-hard 问题。因此一个小小的分割标准的改变会大大改变解决问题的难度。

区域比较预测法通过比较 $Dif(C_1, C_2)$ 和 $Int(C_1)$ 与 $Int(C_2)$ 中较小的一个，来判断这两个区域是否有一个边界（换言之这两个区域是否有足够的理由保持两个区域）。

$$f(n) = \begin{cases} true, & \text{if } Dif(C_1, C_2) > MInt(C_1, C_2) \\ false, & otherwise \end{cases} \quad (2-10)$$

我们引入了一个阈值函数来控制我们希望的外部不相似度与内部不相似度的相差程度。

$$MInt(C_1, C_2) = \min(Int(C_1) + \tau(C_1), Int(C_2) + \tau(C_2)) \quad (2-11)$$

对于比较小的区域， $Int(C)$ 并不能够较好的反应局部特性，比如最极端的情况下，当 $|C| = 1$ 时， $Int(C) = 0$ 。因此我们需要一个跟区域大小相关的阈值函数

$$\tau(C) = \frac{k}{|C|}$$

其中 $|C|$ 表示的是区域 C 的大小， k 是一个常数。越是小的区域，我们越希望较大的外部不相似性。在实际中，我们可以调整 k 的取整来获得不同的效果。当 k 值很大时，算法倾向于分割出来较大的块，当 k 值较小时，算法倾向于更细的划分。

本节最后我们探讨一个比较有趣的话题，就是 τ 函数的选取，如果我们改变这个函数，不会对算法的大框架造成影响，而会对分割结果的倾向性有影响。比如我们可以让分割倾向于某一种形状 A ，令 τ 函数在区域不是形状 A 的时候较大即可。这种形状上的倾向可以比较简单，比如希望正方形的或者扁平状的，也可以比较复杂，是一种特殊的形状。

4、分割算法

本节讲解主要的算法部分，怎样利用上述的定义，在基于图的表示方法下，做

出高效而准确的分割。算法的核心：

输入是一个图 $G = (V, E)$ ，有 n 个点和 m 个边。输出是一个分割 V ，分割成 $S = (C_1, \dots, C_2)$ 。

1. 对 E 进行排序，生成非递减的序列 $\pi = (o_1, \dots, o_m)$
2. 从初始分割 S^0 开始，每一个点 v_i 自己就是一个区域
3. 对于每一个 $q = 1, \dots, m$ 重复步骤 3
4. 通过 S^{q-1} 构建 S^q ，使用如下的方式：令 v_i 和 v_j 表示按顺序排列的第 q 条边的两个点，比如 $o_q = (v_i, v_j)$ 。如果 v_i 和 v_j 在 S^{q-1} 中连个不同的区域下，并且 $w(o_q)$ 比两个区域的内部不相似度都小，那么合并这连个区域，否则什么也不做。用公式来表达就是：令 C_i^{q-1} 是 S^{q-1} 的一个区域，它包含点 v_i ；令 C_j^{q-1} 是 S^{q-1} 的一个区域，它包含点 v_j 。如果 $C_i^{q-1} \neq C_j^{q-1}$ 并且 $w(o_q) \leq MInt(C_i^{q-1}, C_j^{q-1})$ ，那么通过合并 C_i^{q-1} 和 C_j^{q-1} 我们得到了 S^q ；否则的话 $S^q = S^{q-1}$
5. 返回 $S = S^m$

分割结果如图所示：



2.6 图像融合

参考文献

- [1] Demoment G. Image reconstruction and restoration: Overview of common estimation structures and problems [J]. Acoustics, Speech and Signal Processing, IEEE Transactions on, 37 (12), 1989: 2024–2036.
- [2] Park S C, Park M K, Kang M G. Super-resolution image reconstruction: a technical overview [J]. Signal Processing Magazine, IEEE, 20 (3), 2003: 21–36.

- [3] Brown M, Lowe D G. Automatic Panoramic Image Stitching using Invariant Features [J]. International Journal of Computer Vision, 74 (1), 2006: 59–73.
- [4] Lowe D G. Distinctive image features from scale-invariant keypoints [J]. International Journal of Computer Vision, 60 (2), 2004: 91–110.

第三章 大规模近似重复图像搜索算法概述

随着多媒体业务的日益增长，近似重复图像搜索（Near Duplicate Image Retrieval）或部分重复图像搜索（Partial Duplicate Image Retrieval）技术得到了愈加广泛的应用。在我们的图像重建系统中的相似图像搜索环节，我们希望找到尽可能多的与用户拍摄图像相似的图像，将其作为后续重建环节的候选图像。因此我们面临的三个技术难点是：（1）相似搜索是在图像的局部进行的，而不是整幅图像，所以使用全局特征进行相似图像搜索的传统方案并不适用，是否有能表述局部特性的图像表示方法；（2）图像的局部特征信息较少，如何充分利用特征之间的几何位置关系进行图像局部匹配来提高搜索精度；（3）云端图像数据库是 Web 规模的（Web-Scale），图像数据量极大，对算法的时空复杂度限制较大。如何在使用图像局部特征和其空间位置关系的同时尽量不增加搜索算法的复杂度，是本系统需要解决的难题。

本章首先介绍传统的图像搜索算法，再介绍利用局部特征的空间信息的相似图像搜索算法，最后针对本论文的应用场景，提出一种结合多种技术的新的相似图像搜索技术。

3.1 传统的大规模图像检索方法

3.1.1 近似重复图像搜索算法概述

3.1.2 图像表示与相似性度量

3.1.3 基于局部特征的相似图像搜索算法

BoW

聚类 tf-idf

3.2 改进的相似搜索算法

文献^[1]对近期的大规模相似图像搜索技术做了总结，提到了 Partial-Duplicate Image Retrieval via Saliency-Guided Visual Matching^[2]技术，通过视觉显著性（saliency）模型进行比较，消除背景中的噪声。这种方法使得索引和匹配都集中在显著性区域，更能够符合用户的预期。显著值和空间约束都能够被用来进行相似性度量，并且能

够高效的进行二级索引，对于大规模的 partial duplicate search 非常有利，但是内存开销比较大。

Web-Scale Image Retrieval Using Compact Tensor Aggregation of Visual Descriptors^[3] 描述了目前存在的各种视觉描述子的概况，介绍了相关的索引技术，包括哈希、词袋以及基于树的表示方法。(hashing, bag-of-words, and tree-based representation) 引出内存开销问题并提出一种生成高度压缩签名 (highly compact signatures) 的方法，包括张量聚合，PCA，kernel PCA 等一些列算法。它改进了 Fisher Vector 族描述子，提高它的可区分性，以及特征签名的大小 (feature discriminative power and the size of feature signature)。

对于相似性视频搜索，它的特点是特征维数特别大，有研究提出了稀疏投影方式进行特征降维，并且使用数据挖掘的知识使用一些 metadata 来共同进行搜索^[4]。使用机器学习技术，学习稀疏投影矩阵 (sparse projection matrices)。这种学习方法可以选择性的使用外部信息，比如 Wikipedia 上的知识和 Google 搜索结果中的摘要，创建一个语义相关的投影矩阵，生成一个压缩签名，以满足手机媒体检索的诸多限制。手机内存空间小，计算资源有限，传统的将高维特征映射到低维的投影矩阵在手机内存是放不下的。而我们的稀疏投影矩阵是能够在手机上使用的。

下面我们简单介绍各种性能改进的方法。

3.2.1 sim-hash

文本的去重算法中常见的有余弦夹角算法、欧式距离、Jaccard 相似度、最长公共子串、编辑距离等，但是只适合于小数据集。simhash 传统的用来判断两篇文章的相似度，将两篇文章映射到低维空间上，并且保持它们互相之间的相似度，但是它很难应用在图像比较上，因为图像的特征是用实数来表示的，尽管可以将其量化，但是两幅相似图像量化后的特征集合交叠的比率仍旧很小，远远小于文档，因为两幅图像不相似的区域噪声特征非常大。但是如果使用视觉词组，那么如果两个相似区域的视觉词组会非常相同，我们就可以使用 simhash 了。所以，min-hash 的使用场景是特征比较多，相似度比较显著的情况下。

3.2.2 最小哈希的相似性比较

最小哈希方法是一种广泛应用在相似性查找领域的算法。

在生成视觉词带之后，比较两幅图像或者两篇文章的相似度问题转化为比较两

个只包含 0, 1 元素的集合的相似度, 集合的相似度是 Jaccard 相似度。

$$JS(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

我们首先定义一组随机哈希函数 $f_j: \mathcal{V} \rightarrow R$, 每一个哈希函数是独立的, 将一个视觉词映射为一个实数。两个不同的视觉词 X_a 和 X_b , 哈希函数需要满足两点: (1) $f_j(X_a) \neq f_j(X_b)$; (2) $P(f_j(X_a)) < (P(f_j(X_b))) = 0.5$ 。

注意到函数 f_j 能够反映出视觉词集合的一个顺序排列情况, 我们定义 min-hash 就是这个排列中排在最前面的视觉词, 因此有

$$m(\mathcal{A}_i, f_j) = \arg \min_{X \in \mathcal{A}_i} f_j(X)$$

根据上述定义, 我们会发现以下这个事实:

$$Pr(m(A, f_j) = (B, f_j)) = \frac{|A \cap B|}{|A \cup B|} = sim_s(A, B)$$

如果 r 是随机变量, 当 $m(A, f_j) = (B, f_j)$ 时值为 1, 其它情况下值为 0 的, 那么 r 可认为是 $J(A, B)$ 的无偏估计。因此我们可以使用 min-hash 函数将一幅图片或者一篇文章转化为一个数 (对该文章中的每一个单词 id 使用 hash 函数后得到一个新的 id 序列, 这个序列中的第一个出现 1 的行号, 就是 min-hash 的值), 当我们使用 k 个 hash 函数, 得到 k 个值, 将原本的高维向量映射到了低维。min-hash 在压缩原始集合的情况下, 保证了集合的相似度没有被破坏。文献^[5]中提出将词频-逆文档频率 (term frequency-inverse document frequency, 简称 TF-IDF) 融合在传统的 min-hash 算法中, 实验表明能够提高搜索的准确率。

3.2.3 LSH

使用 min-hash 对数据降维后, 可以使用 LSH 缩小查找范围, 其基本思路是将相似的集合聚集到一起, 减小查找范围, 避免比较不相似的集合。

对每一列 c (即每个集合) 我们都计算出了 n 行 min-hash 值, 我们把这 n 个值均分成 b 组, 每组包含相邻的 $r=n/b$ 行。对于每一列, 把其每组的 r 个数都算一个 hash 值出来, 把此列的编号记录到 hash 值对应的 bucket 里。如果两列被放到了同一个 bucket 里, 说明它们至少有一组 (r 个) 数的 hash 值相同, 此时可认为它们有较大可能相似度较高 (称为一对 candidate)。最后在比较时只对落在同一个 bucket 里的集合两两计算, 而不是全部的两两比较。

3.3 空间信息匹配搜索算法

3.3.1 随机抽样一致算法

随机抽样一致 RANdom SAmple Consensus (RANSAC) 是一种空间匹配算法。该算法将数据分成两类，局内点 (inlier) 和局外点 (outlier) 它可以从一组包含局外点的观测数据集中，通过迭代方式估计数学模型的参数。

这是一种不确定的算法，有一定的概率得出一个正确的或者说是可接受的合理结果；一般情况下，迭代次数的增加可以提升结果的准确性。该算法由 Fischler 和 Bolles 于 1981 年提出，在图像检索中，RANSAC 可以作为检索后的后续处理，对图像的中目标进行空间一致验证。

RANSAC 算法对数据集做了三个假设：

- 数据由局内点组成，局内点的数据的分布符合某一特定的概率模型；
- 与局内点相对的是局外点，他们不能够适应该模型；
- 局内点与局外点之外的数据属于噪声

RANSAC 有以下几个步骤：

- 随机选择数据集合的一个子集
- 使用选择的自己拟合一个数学模型
- 确定该模型下局外点的个数
- 重复步骤 1 3 若干次，以最好的一次结果最为最终拟合出来的数学模型

RANSAC 算法迭代次数的选取取决于我们期望的准确率与样本数量。设 p 为任意给定对应点合法的概率，即

$$p = \frac{\text{局内点的数量}}{\text{数据集全部数据的数量}}$$

而 P 是经过 S 次试验后成功的总体概率。设我们需要 k 个随机样本来估计模型，那么在一次试验中，该 k 个样本都是局内点的可能性为 p^k 。因此， S 次试验失败的可能性是

$$1 - P = (1 - p^k)^S$$

两边去对数，得到最少需要的试验次数是

$$S = \frac{\log(1-P)}{\log(1-p^k)}$$

随着 k 的增大，需要的最少试验次数增多，在实际中，我们应该尽可能的选择小的 k 值。在模型确定以及最大迭代次数允许的情况下，RANSAC 总是能找到最优解。对于含有较大误差的数据集，RANSAC 的效果远优于直接的最小二乘法。

当对两幅图像进行匹配的时候，所以相互匹配的局部特征作为数据全集，我们要估算的模型是一个变换矩阵 H ，能够将图像 I 投影到图像 I' 。每次迭代过程中，随机的选择四对匹配的特征点，根据这四个特征点的位置信息解得变换 H ，利用 H 计算其它匹配对的位置信息中有哪些属于局外点，记录局外点的个数。局外点的个数越少，变换矩阵 H 越准确。反复迭代多次得到一个相对准确的透视变换模型。

上述提到的用四对匹配点拟合出的变换矩阵叫做单应矩阵（Homography），最简单的求解单应性矩阵的算法叫做直接线性变换法（Direct Linear Transform, DLT）^[6]，其具体算法如下：

假设我们相匹配的一对点分别是 x 和 x' ，单应性矩阵是 H ，那么有如下等式：

$$c \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = H \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad (3-1)$$

其中 c 是一个非零常数， $(u \ v \ 1)^T$ 代表 x' ， $(x \ y \ 1)^T$ 代表 x ，而 $H = \begin{pmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{pmatrix}$

将公式 (3-1) 展开，分别用第一行和第二行除以第三行，得到

$$-h_1x - h_2y - h_3 + (h_7x + h_8y + h_9)u = 0 \quad (3-2)$$

$$-h_4x - h_5y - h_6 + (h_7x + h_8y + h_9)u = 0 \quad (3-3)$$

公式 (3-2) 和 (3-3) 可以写成矩阵的形式：

$$A_i h = 0 \quad (3-4)$$

其中 $A_i = \begin{pmatrix} -x & -y & -1 & 0 & 0 & 0 & ux & uy & u \\ 0 & 0 & 0 & -x & -y & -1 & vx & vy & v \end{pmatrix}$, 而 $h = (h_1 \ h_2 \ h_3 \ h_4 \ h_5 \ h_6 \ h_7 \ h_8 \ h_9)^T$ 。

因为每一对匹配的点可以提供两个等式, 对于解决 8 自由度的矩阵 H , 只需要四对 (任意三点不能共线) 匹配的特征点。

DLT 算法依赖于坐标系的原点和尺度, 所以该算法并不稳定, 在实际中更多的使用多个匹配点得到更多的方程, 将求单应矩阵的问题转化为求解最小二乘的问题, 用矩阵奇异值分解 (Singular value decomposition, SVD) 的方法来求解等式 $Ah = 0$ 。

3.3.2 视觉词组

BoW 的一个被人诟病的缺点是没有利用任何的图像空间信息, 而这正是图像搜索与文本搜索最显著的区别所在, 目前, 有很多研究成功的利用局部特征的空间位置信息进行更加精确的, 文献^[7]深入研究 SIFT 描述子。提出了一个非常优雅的方法: 生成 sift 组, 嵌入几何信息, 最终讲一个 group 压缩到一个 64 比特的二维签名中, 叫做 Nested-SIFT。

它的优点是 Nested-SIFT 使用 sift 描述子的嵌套关系, 很自然的将不同尺度的局部关键点组合在一起, 生成一个特征签名。嵌入空间信息的 Nested-SIFT 可区分性更强。使用 SimHash 进行压缩后, 在视觉搜索中效率更高。实验结果表明这种方法提高搜索的准确度, 减少了内存消耗, 提高搜索速度。其缺点是生成 Nested-SIFT 会有一定的计算消耗。

3.4 适用于旅游景点图像的相似图像搜索技术

参考文献

- [1] POLICY N. Web-Scale Near-Duplicate Search: Techniques and Applications [J]. IEEE MultiMedia, 2013.
- [2] Li L, Jiang S, Zha Z-J, et al. Partial-Duplicate Image Retrieval via Saliency-Guided Visual Matching [J]. MultiMedia, IEEE, 20 (3), 2013: 13–23.
- [3] Negrel R, Picard D, Gosselin P. Web scale image retrieval using compact tensor aggregation of visual descriptors [J], 2013.
- [4] Wu G-L, Kuo Y-H, Chiu T-H, et al. Scalable mobile video retrieval with sparse projection learning and Pseudo label mining [J], 2013.
- [5] Chum O, Philbin J, Zisserman A. Near Duplicate Image Detection: min-Hash and tf-idf Weighting. [J], 810, 2008: 812–815.
- [6] Dubrofsky E. Homography estimation [J], 2009.

- [7] Xu P, Zhang L, Yang K, et al. Nested-SIFT for Efficient Image Matching and Retrieval [J], 2013.

附录 A 不定型 (0/0) 极限的计算

定理 A.1 (L'Hospital 法则) 若

1. 当 $x \rightarrow a$ 时, 函数 $f(x)$ 和 $g(x)$ 都趋于零;
2. 在点 a 某去心邻域内, $f'(x)$ 和 $g'(x)$ 都存在, 且 $g'(x) \neq 0$;
3. $\lim_{x \rightarrow a} \frac{f'(x)}{g'(x)}$ 存在 (或为无穷大),

那么

$$\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = \lim_{x \rightarrow a} \frac{f'(x)}{g'(x)}. \quad (\text{A-1})$$

证明: 以下只证明两函数 $f(x)$ 和 $g(x)$ 在 $x = a$ 为光滑函数的情形。由于 $f(a) = g(a) = 0$, 原极限可以重写为

$$\lim_{x \rightarrow a} \frac{f(x) - f(a)}{g(x) - g(a)}.$$

对分子分母同时除以 $(x - a)$, 得到

$$\lim_{x \rightarrow a} \frac{\frac{f(x) - f(a)}{x - a}}{\frac{g(x) - g(a)}{x - a}} = \frac{\lim_{x \rightarrow a} \frac{f(x) - f(a)}{x - a}}{\lim_{x \rightarrow a} \frac{g(x) - g(a)}{x - a}}.$$

分子分母各得一差商极限, 即函数 $f(x)$ 和 $g(x)$ 分别在 $x = a$ 处的导数

$$\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = \frac{f'(a)}{g'(a)}.$$

由光滑函数的导函数必为一光滑函数, 故 (A-1) 得证。□

附录 B 缩略语表

BUPT Beijing University of Posts and Telecommunications, 北京邮电大学

WTI Wireless Technology Innovation Institute, 无线新技术研究所

致 谢

感谢 Donald Ervin Knuth.

攻读学位期间发表的学术论文目录

期刊论文

- [1] **Zhang San**, Newton I, Hawking S W, et al. An extended brief history of time [J]. Journal of Galaxy, 1234 (4), 2079: 567–890. (SCI 收录, 检索号: 786FZ) .

会议论文

- [1] McClane J, McClane L, Gennero H, et al. Transcript in Die hard [C]. In Proc. HDDD 100th Super Technology Conference (STC 2046), Eta Cygni, Cygnus, September 21–24, 2046: 123–456. (EI 源刊) .