



南 华 大 学
UNIVERSITY OF SOUTH CHINA

毕业论文文献综述

题 目 基于最优传输理论图像匹配算法文献综述

学院名称 数理学院

指导教师 高有

职 称 讲师

班 级 信计 1802 班

学 号 20184390213

学生姓名 朱柳承

2022 年 1 月 30 日

基于最优传输理论图像匹配算法文献综述

摘要：图像匹配是虚拟图像重建的一个关键步骤，是视觉识别的一个核心过程，是图像检索的中心操作模块。视觉算法在过去几十年来飞速发展，基于图像匹配的应用更是层出不穷，从卫星遥感图像配准到纳米级零件配准，在科学研究、社会需求、工业制造等各个方面，图像匹配这一基本问题无处不在。而对于图像匹配其定义一般被描述如下：同一摄影项目的图像如医学图像，可以在遮挡、多姿态等条件下从任何的光照强度和频率以各种角度拍摄，这些同一项目的图像之间存在内容、结构、特征、色彩及纹理等对应关系，而图像匹配则致力于这些关系中的一致性和相似性分析。

尤其随着计算机运算能力的提升、图像处理加速芯片的发展，图像匹配算法数量越来越多，其类型也越来越丰富。随之产生了两个问题：新方法 with 旧方法的取舍；如何在现有理论指导下设计更适用、准确、鲁棒的高性能匹配算法。为了回答这两个问题，我们有必要系统的回顾和评估过去和现在的图像匹配算法。我在检索相关的文献时，发现了一个奇怪的现象：如今大热的并且在理论上对于图像匹配算法的发展具有巨大的推进潜力的最优传输理论方面的内容常常被一些重要的文献综述 [2] 所忽略。因此，本文主要讨论基于最优传输理论的图像匹配算法。

本文首先回顾了遵循着从手工设计特征到深度学习的图像匹配算法的发展路线，并简要分析了其中各个阶段的算法的特点。然后，详细地介绍最优传输理论在视觉算法上的进展，其中涉及计算最优传输映射的方法、计算 Wasserstein 距离的方法的概述；接着，列举几个基于最优传输理论的图像匹配应用，并且与经典的解决方案相比较，以此了解 OT 在视觉算法中的意义。最后，我们总结了图像匹配技术的现状，并对未来的工作进行了富有洞察力的讨论和展望。本调查可作为（但不限于）图像匹配及相关领域的研究人员和工程师参考。

关键词：图像匹配，共形映射，最优传输理论，最佳质量传输映射，Wasserstein 距离，曲面配准，手工设计特征，深度学习

Graduation Thesis Literature Review

Graduation Thesis Topic: Literature review of image matching algorithms
based on optimal transmission theory

Student name: Zhu Liucheng Student number: 20184390213

Tutor name: Gaoyou Professional qualifications: Lecturer

Abstract : Image matching is a key step in virtual image reconstruction, a core process of visual recognition, and a central operation module of image retrieval. With the rapid development of visual algorithms in the past decades, image matching-based applications have emerged one after another. From satellite remote sensing image registration to nano-scale parts registration, the basic problem of image matching is ubiquitous in scientific research, social needs, industrial manufacturing and so on. The definition of image matching is generally described as follows: images of the same photographic project, such as medical images, can be taken from any angle of light intensity and frequency under occlusion, multi-posture, etc. There are corresponding relationships among the images of the same project, such as content, structure, feature, color and texture, while image matching focuses on consistency and similarity analysis in these relationships.

Especially with the improvement of computer computing ability and the development of image processing acceleration chip, there are more and more image matching algorithms and their types are more and more abundant. Two problems arise: the choice between the new method and the old one; How to design a more applicable, accurate and robust high performance matching algorithm under the guidance of the existing theory. In order to answer these two questions, it is necessary to systematically review and evaluate past and present image matching algorithms. When I retrieve the relevant literature, I found a strange phenomenon: the hot and theoretically promising aspects of optimal transmission theory for the development of image matching algorithms are often ignored by some important literature reviews[2]. Therefore, this paper mainly discusses image matching algorithms based on optimal transmission theory.

This paper first reviews the development of image matching algorithms that follow the path from manual design features to in-depth learning, and briefly analyses the characteristics of the algorithms at each stage. Then, the progress of optimal transmission theory in visual algorithms is introduced in detail, including an overview of the methods for calculating optimal transmission mappings and Wasserstein distances. Next, several image matching applications based on optimal transmission theory are listed and compared with classical solutions to understand the significance of OT in visual algorithms. Finally, we summarize the current status of image matching technology, and make insightful discussions and prospects for future work. This survey can be used as a reference for (but not limited to) researchers and engineers in image matching and related fields.

Keywords: Image matching, Conformal mapping, Optimal transmission theory, Optimal mass transport mapping, Wasserstein distance, Surface registration, Handcrafted feature, Deep learning

1 引言

元宇宙概念在近年来吸引了大量企业与研究者的关注，主要是因为其在技术上，特别是与虚拟图像、计算机辅助技术领域的三大突破有关。其一是英伟达发布了世界首款实时光线追踪 GPU，我们知道高质量的 3D 渲染的核心算法是基于几何光学的光线追踪法。二十年前，该算法只能在昂贵的 Sun 或者 SGI 工作站上运算。依随岁月的流逝，越来越多的物理定则被加入到算法流程之中，渲染效果愈发逼真。几乎所有的电影特效都是基于光学追踪法，一部电影往往需要数千台 Linux 服务器计算数年。长久以来，大家都将实时光线追踪计算作为一个梦想。终于，英伟达的 GPU 技术积累到达了临界点。

其二便是 Epic Game 发布的虚幻引擎五，它具备两大全新核心技术：Nanite 虚拟微多边形几何技术和 Lumen 动态全局光照技术。Nanite 虚拟几何技术的出现意味着由数以亿计的多边形组成的影视级艺术作品可以被直接导入虚幻引擎，Nanite 几何体可以被实时流送和缩放，因此无需再考虑多边形数量预算、多边形内存预算或绘制次数预算了；也不用再将细节烘焙到法线贴图或手动编辑细节层次（LOD），这必定是图形学领域革命性的飞跃。

其三便是 AI 的 GAN model，对抗生成网络（Generative Adversarial Network GAN）获得了爆炸式的增长，其应用范围几乎涵盖了图像处理和机器视觉的绝大多数领域。其精妙独到的构思，令人拍案叫绝；其绚烂逼真的效果，令众生颠倒。一时间对抗生成网络引发了澎湃汹涌的技术风潮，纳什均衡的概念风靡了整个人工智能领域。GAN 的核心思想是构造两个深度神经网络：判别器 D 和生成器 G，用户为 GAN 提供一些真实货币作为训练样本，生成器 G 生成假币来欺骗判别器 D，判别器 D 判断一张货币是否来自真实样本还是 G 生成的伪币；判别器和生成器交替训练，能力在博弈中同步提高，最后达到平衡点的时候判别器无法区分样本的真伪，生成器的伪造功能炉火纯青，生成的货币几可乱真。这种阴阳互补，相克相生的设计理念为 GAN 的学说增添了魅力。

1.1 全局准则

对于可预见的未来，设计图像检索工具的主要限制是我们对视觉的理解十分有限。尽管理解不充分，我们也能构造有用的工具，就如 IBM 的已经出现在大量的市场广告中的图像搜索产品 QBIC，以及看上去生意兴隆的 Virage 公司的

图像搜索引擎等等，但是我们仍然很难去评价怎样算成功。表示图像的方式粗略地分有三种：在像素级，人们对具体的箱数值感兴趣；在组合级，人们关心图像的整体外观；或是在对象语义级，人们关注图像所描述的事务。

1.2 局部特征准则

对于可预见的未来，设计图像检索工具的主要限制是我们对视觉的理解十分有限。尽管理解不充分，我们也能构造有用的工具，就如 IBM 的已经出现在大量的市场广告中的图像搜索产品 QBIC，以及看上去生意兴隆的 Virage 公司的图像搜索引擎等等，但是我们仍然很难去评价怎样算成功。表示图像的方式粗略地分有三种：在像素级，人们对具体的箱数值感兴趣；在组合级，人们关心图像的整体外观；或是在对象语义级，人们关注图像所描述的事务。

1.3 特征学习准则

对于可预见的未来，设计图像检索工具的主要限制是我们对视觉的理解十分有限。尽管理解不充分，我们也能构造有用的工具，就如 IBM 的已经出现在大量的市场广告中的图像搜索产品 QBIC，以及看上去生意兴隆的 Virage 公司的图像搜索引擎等等，但是我们仍然很难去评价怎样算成功。表示图像的方式粗略地分有三种：在像素级，人们对具体的箱数值感兴趣；在组合级，人们关心图像的整体外观；或是在对象语义级，人们关注图像所描述的事务。

2 基于最佳质量运输理论的匹配算法

对于可预见的未来，设计图像检索工具的主要限制是我们对视觉的理解十分有限。尽管理解不充分，我们也能构造有用的工具，就如 IBM 的已经出现在大量的市场广告中的图像搜索产品 QBIC，以及看上去生意兴隆的 Virage 公司的图像搜索引擎等等，但是我们仍然很难去评价怎样算成功。表示图像的方式粗略地分有三种：在像素级，人们对具体的箱数值感兴趣；在组合级，人们关心图像的整体外观；或是在对象语义级，人们关注图像所描述的事务。

2.1 Wasserstein 距离

对于可预见的未来，设计图像检索工具的主要限制是我们对视觉的理解十分有限。尽管理解不充分，我们也能构造有用的工具，就如 IBM 的已经出现在大量的市场广告中的图像搜索产品 QBIC，以及看上去生意兴隆的 Virage 公司的图像搜索引擎等等，但是我们仍然很难去评价怎样算成功。表示图像的方式粗略地分有三种：在像素级，人们对具体的箱数值感兴趣；在组合级，人们关心图像的整体外观；或是在对象语义级，人们关注图像所描述的事务。

2.2 前置处理

对于可预见的未来，设计图像检索工具的主要限制是我们对视觉的理解十分有限。尽管理解不充分，我们也能构造有用的工具，就如 IBM 的已经出现在大量的市场广告中的图像搜索产品 QBIC，以及看上去生意兴隆的 Virage 公司的图像搜索引擎等等，但是我们仍然很难去评价怎样算成功。表示图像的方式粗略地分有三种：在像素级，人们对具体的箱数值感兴趣；在组合级，人们关心图像的整体外观；或是在对象语义级，人们关注图像所描述的事务。

2.3 后置处理

对于可预见的未来，设计图像检索工具的主要限制是我们对视觉的理解十分有限。尽管理解不充分，我们也能构造有用的工具，就如 IBM 的已经出现在大量的市场广告中的图像搜索产品 QBIC，以及看上去生意兴隆的 Virage 公司的图像搜索引擎等等，但是我们仍然很难去评价怎样算成功。表示图像的方式粗略地分有三种：在像素级，人们对具体的箱数值感兴趣；在组合级，人们关心图像的整体外观；或是在对象语义级，人们关注图像所描述的事务。

2.4 OT 匹配算法

对于可预见的未来，设计图像检索工具的主要限制是我们对视觉的理解十分有限。尽管理解不充分，我们也能构造有用的工具，就如 IBM 的已经出现在大量的市场广告中的图像搜索产品 QBIC，以及看上去生意兴隆的 Virage 公司的图像搜索引擎等等，但是我们仍然很难去评价怎样算成功。表示图像的方式粗略地分有三种：在像素级，人们对具体的箱数值感兴趣；在组合级，人们关心图像

的整体外观；或是在对象语义级，人们关注图像所描述的事务。

3 OT 匹配算法的应用

对于可预见的未来，设计图像检索工具的主要限制是我们对视觉的理解十分有限。尽管理解不充分，我们也能构造有用的工具，就如 IBM 的已经出现在大量的市场广告中的图像搜索产品 QBIC，以及看上去生意兴隆的 Virage 公司的图像搜索引擎等等，但是我们仍然很难去评价怎样算成功。表示图像的方式粗略地分有三种：在像素级，人们对具体的箱数值感兴趣；在组合级，人们关心图像的整体外观；或是在对象语义级，人们关注图像所描述的事务。

4 匹配算法优化与评估方法

对于可预见的未来，设计图像检索工具的主要限制是我们对视觉的理解十分有限。尽管理解不充分，我们也能构造有用的工具，就如 IBM 的已经出现在大量的市场广告中的图像搜索产品 QBIC，以及看上去生意兴隆的 Virage 公司的图像搜索引擎等等，但是我们仍然很难去评价怎样算成功。表示图像的方式粗略地分有三种：在像素级，人们对具体的箱数值感兴趣；在组合级，人们关心图像的整体外观；或是在对象语义级，人们关注图像所描述的事务。

常见问题

1. 模板使用说明请见 [ucasthesis](#)：中国科学院大学学位论文 LaTeX 模板。
2. 填表说明和模板说明不是开题报告的一部分，请删除。
3. 开题报告样式设计导致对题目换行与不换行难以兼容，排版十分困难。推荐采用当前设置，尽量避免将精力花在这些无关紧要的细节上。

5 总结与展望

对于可预见的未来，设计图像检索工具的主要限制是我们对视觉的理解十分有限。尽管理解不充分，我们也能构造有用的工具，就如 IBM 的已经出现在

大量的市场广告中的图像搜索产品 QBIC，以及看上去生意兴隆的 Virage 公司的图像搜索引擎等等，但是我们仍然很难去评价怎样算成功。表示图像的方式粗略地分有三种：在像素级，人们对具体的箱数值感兴趣；在组合级，人们关心图像的整体外观；或是在对象语义级，人们关注图像所描述的事务。

参考文献

- [1] Haker S, Zhu L, Tannenbaum A, et al. Optimal mass transport for registration and warping [J]. International Journal of computer vision, 2004, 60(3): 225-240.
- [2] Ma J, Jiang X, Fan A, et al. Image matching from handcrafted to deep features: A survey [J]. International Journal of Computer Vision, 2021, 129(1): 23-79.
- [3] Ma M, Lei N, Su K, et al. Surface-based shape classification using wasserstein distance [J]. Geometry, Imaging and Computing, 2015, 2(4): 237-255.
- [4] Ma M, Lei N, Chen W, et al. Robust surface registration using optimal mass transport and teichmüller mapping [J]. Graphical models, 2017, 90: 13-23.
- [5] Su Z, Wang Y, Shi R, et al. Optimal mass transport for shape matching and comparison [J]. IEEE transactions on pattern analysis and machine intelligence, 2015, 37(11): 2246-2259.
- [6] De Goes F, Cohen-Steiner D, Alliez P, et al. An optimal transport approach to robust reconstruction and simplification of 2d shapes [C]//Computer Graphics Forum: volume 30. Wiley Online Library, 2011: 1593-1602.
- [7] Mérigot Q. A multiscale approach to optimal transport [C]//Computer Graphics Forum: volume 30. Wiley Online Library, 2011: 1583-1592.
- [8] Ma M, Marino J, Nadeem S, et al. Supine to prone colon registration and visualization based on optimal mass transport [J]. Graphical Models, 2019, 104: 101031.
- [9] 曹杰彰, 莫朗元, 杜卿, 等. 基于最优传输理论的联合分布匹配方法及应用 [J]. 计算机学报, 2021, 44(06): 1233-1245.
- [10] Villani C. Optimal transport: old and new: volume 338 [M]. Springer, 2009.
- [11] Peyré G, Cuturi M, et al. Computational optimal transport [J]. Center for Research in Economics and Statistics Working Papers, 2017(2017-86).