南華大學
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：新方法与旧方法的取舍；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 learing

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 技术积累到达了这个临界点。

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 Adersarial 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

图像的一般描述方法有纹理、统计、基于模型以及基空间方法。纹理作为一个关键度量，是图像处理中的重要主题，他通常分为结构方法和统计方法，结构方法寻找边缘和形状等特征，而统计方法关心的是像素值的关系和统计矩。

Fourier 空间等基空间方法也可以用于特征描述。在 20 世纪 60 80 年代，人们为了在高分辨率的彩色图像上做一些匹配等任务，往往只有在内存足够的情况下才能进行。这一时期主要是整体目标方法，它用特征度量来描述几乎整个目标、较大的区域或图像。大型目标的模式匹配采用 FFT 谱方法和其他方法，识别方法包括目标、形状以及纹理等度量，并使用简单的集合元素进行目标组合。NTSC 制、PAL 制和 SECAM 制等低分辨率图像比较常见，而且主要是灰度图像。

基于上述的全局特征与一些几何方法，通过计算两幅图像的特征向量之间的欧氏距离来进行匹配。这种方法本质上对光照变化具有鲁棒性，但有一个巨大的缺点：即使使用最先进的算法，标记点的准确配准也很复杂。在 [2] 中进行了一些关于几何人脸匹配的工作. 使用了 22 维特征向量，并且对大型数据集的实验表明，仅几何特征可能无法携带足够的信息来进行人脸匹配。简而言之这类图像匹配算法缺乏泛化能力。于是人们在分析图像空间时考虑了降维方法，认为高维的图像空间如 100×100 的图像即 10000 维的图像空间中，只有部分像素是我们关心的，因此主成分分析法很自然地引入。在图像匹配过程中，该分析识别具有最大方差的轴，即确定一组正交基从而得到描述图像的主成分向量，接着把两幅图像的主成分之间的距离作为相似性度量进行图像匹配操作。

然而，PCA 方法从重建的角度来看，这种转换是最佳的，但它没有考虑任何类标签。比方说对一个人脸进行差异化分类，可分为脸间差异和脸内差异。脸内差异表示同一个人脸的各种可能变形。脸间差异表示不同人的本质差异。对于同一个人，不同表情会使匹配效果不稳定，因此一般 PCA 方法会使用同一项目的各种姿势下的平均图像作为匹配依据，但是当一个方差是由外部来源产生的情况下，具有最大方差的轴不一定包含任何判别信息，此时分类变得不可能。对于这一类情况，人们进一步提出了将具有线性判别分析的特定类别投影应用于人脸匹配 [3]，其基本思想是最小化类内的方差，同时最大化类间的方差。

上述的图像匹配思想都是基于全局特征进行的，这有一个不足之处便是当图像的热点区域若出现遮挡或缺损等情况，上述算法则会不稳定。因此，人们进一步提出了各种基于局部特征提取的方法。

1.2 局部特征准则

20 世纪 90 年代初期（部分目标方法），人们越来越多地使用局部特征和兴趣点来描述图像中较小的目标、目标的部件和图像区域，例如 Shi 和 Tomasi[149] 改进了 Harris 检测器，kitchen 和 Rosenfeld[200] 提出了灰度角点检测方法，Khotanzad 和 Hong[268] 与 1990 年提出使用 Zernike 多项式来计算多边形形状的图像矩等。20 世纪 90 年代中期（局部特征方法）：特征描述子从每个特征周围的窗口上添加更多细节，通过特征搜索和匹配来进行目标识别。先是搜索特征集并使用更复杂的分类器来匹配描述子。描述子包括梯度、边缘和颜色。20 世纪 90 年代后期，开发了各种具有局部不变性的特征描述子，这些描述子对尺度、亮度、旋转和仿射变换等具有不变性。Schmid 和 Mohr[340] 详细介绍了局部特征描述方法。特征就像字母，是拼写出复杂的特征描述子或向量的基础，这些向量将用于特征匹配。

21 世纪初期 Lowe 提出的 SIFT[153] 算法和 Bay 等人提出的 SURF[152] 算法都采取了不同的方式来使用 HAAR 特征而不仅是特度特征。2010 年以后：多模态特征度量融合，这一时期人们更多地使用深度传感器信息和深度图来分割图像、描述特征，比如 Rusu 和 Bradski 等人创建了 VOXEL 度量 [380] 和 2D 纹理度量在 3D 空间中的表示。人们开发出更快、更好的二值模式特征描述子，这种描述子使用汉明距离进行快速匹配，如由 Alahi 等人提出的 FREAK[122]、由 Rublee 等人提出的 ORB[112]。多模态和多变量描述子由图像特征和其他传感器（比如加速度计传感器、位置传感器等）信息构成。

只描述图像的局部区域，提取的特征对部分遮挡、光照和小样本量更鲁棒。用于局部特征提取的算法有 Gabor Wavelets ([4])、离散余弦变换 ([5]) 和局部二进制模式 ([6])。在应用局部特征提取时，什么是保留空间信息的最佳方法仍然是一个开放的研究问题，因为空间信息是潜在的有用信息。换句话说选择一些具有鲁棒性的特征算子仍然是困难的问题。人们对此的解决方案，是采用机器学习的思想来选取更好的特征，

1.3 特征学习准则

21 世纪初期，这一时期采用有良好形势的描述子将场景和目标建模为特征组件或模式集合；为了进行特征匹配，需度量特征之间的空间关系；新的复杂的

分类和匹配方法会采用 Boosting 及相关方法，这种方法结合强弱特征来进行更优效的识别。Viola 和 Jones 方法 [486] 使用 HAAR 特征和基于 Bossting 的学习方法来进行分类，从而加快了匹配速度。21 世纪头十年中期（较细粒度的特征和度量组合方法），Czuka 等人 [226] 提出描述场景和目标的各种特征与度量的组合，而 Sivic 采用关键点方法来描述场景。此外人们还开发了特征学习和稀疏特征码本方法，以减少模式空间，加快搜索速度、提高准确性。

特征学习方法会会创建一组平均的、被压缩（或稀疏）的分层特征集，这就是训练集中的主要特征。机器学习过程可以表述为如下 3 种方式：(1) 特征提取：可用局部特征描述子或通过深度神经网络来学习特征。(2) 特征编码：可保留所有特征集或仅保留一个稀疏特征集。(3) 分类器的设计和训练。

特征学习架构包括以下两大类：(1) 统计学习方法。这类方法包括在特征描述子中广泛使用的方法、学习方法、稀疏编码和统计分类器。(2) 神经网络方法。它是受神经生物学概念的启发而建立的，比如局部感受野与人工神经元的连接、深度特征层次。

其中基于 Haar 特征的 Adaboost 算法取得目前在室内外环境中位姿估计任务中最稳定的结果。该类算法展示了基于注意力的图神经网络对局部特征匹配的强大功能。SuperGlue 的框架使用两种注意力 ([7])：（i）自我注意力，可以增强局部描述符的接受力；（ii）交叉注意力，可以实现跨图像交流，并受到人类来回观察方式的启发进行匹配图像。文中方法通过解决最优传输问题，优雅地处理了特征分配问题以及遮挡点。实验表明，SuperGlue 与现有方法相比有了显着改进，可以对室内和室外的图像进行高精度的相对姿势估计。此外，SuperGlue 可以实时运行，并且可以同时使用经典和深度神经网络去学习特征。如今，深度学习在图像匹配中应用地更为广泛，但是它的不足之处也是显然的，模型的可解释性差、模型坍塌等等。

1.4 总结

深度学习得到的特征为什么效果会好呢? 其原因有：(1) 特征的绝对数量；(2) 特征的层次性，即特征能表示低级概念、中级概念和高级概念。这表明用层次化方法来创建局部特征集（比如 SIFT 和 FREAK）有可能得到与卷积神经网络（采用简单相关性模板特征）相当（甚至可以超过）的性能。

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素产品 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素产品 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素产品 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素产品 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素产品 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素产品 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素产品 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素产品 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).