

学校编码: 10384

分类号 _____ 密级 _____

学号: 22120051302347

UDC _____

厦 门 大 学

硕 士 学 位 论 文

Hausdorff 距离多核并行技术及其应用研究

**The Implementation and Application of Parallel Hausdorff
Distance Algorithm**

林萌

指导教师姓名: 李翠华 教授

专 业 名 称: 计算机应用技术

论文提交日期: 2008 年 月

论文答辩时间: 2008 年 月

学位授予日期: 2008 年 月

答辩委员会主席: _____

评 阅 人: _____

2008 年 月

厦门大学学位论文原创性声明

兹提交的学位论文，是本人在导师指导下独立完成的研究成果。本人在论文写作中参考的其他个人或集体的研究成果，均在文中以明确方式标明。本人依法享有和承担由此论文产生的权利和责任。

声明人（签名）：

年 月 日

厦门大学学位论文著作权使用声明

本人完全了解厦门大学有关保留、使用学位论文的规定。厦门大学有权保留并向国家主管部门或其指定机构送交论文的纸质版和电子版，有权将学位论文用于非赢利目的的少量复制并允许论文进入学校图书馆被查阅，有权将学位论文的内容编入有关数据库进行检索，有权将学位论文的标题和摘要汇编出版。保密的学位论文在解密后适用本规定。

本学位论文属于

1. 保密（ ），在 年解密后适用本授权书。
2. 不保密（ ）

（请在以上相应括号内打“√”）

作者签名：

日期： 年 月 日

导师签名：

日期： 年 月 日

摘要

Hausdorff 距离是匹配点特征的一种重要方法，在图像处理、天文、数学、网络应用、医学、经济预测等众多领域中有重要应用，特别是在图像处理的匹配识别中应用十分广泛。传统的 Hausdorff 距离算法计算复杂度高，计算效率低，因此提高算法效率有十分重要的。另一方面，多核计算技术是当前计算机领域的研究热点，它使计算机的计算能力显著提升，将成为一种广泛普及的计算模式。然而，要真正地凸显多核处理器的优势，软件的发展必须紧跟硬件的步伐，如何开发与多核相适应的软件日益成为计算机技术研究的热点。

本文根据当前计算机软硬件技术的发展趋势，围绕着 Hausdorff 距离算法并行化展开研究，旨在寻求多核平台上高效简捷的并行化支持方案。本文首先对多核体系结构、常用开发环境和适用软件工具进行比较分析，探讨适合于发挥多核性能的编程技术和解决方案；在详细剖析 Hausdorff 距离算法的基础上对其进行了一定的改进，设计了基于多核架构的并行算法，并成功应用于侧视图像中建筑物目标的匹配识别系统，同时能够适用于印刷板检测系统；接下来使用 OpenMP 共享存储编程，结合 Intel VTune Performance Analyzer、Intel Thread Checker 和 Intel C++ Compiler 等工具和解决方案测试其性能，根据代码在多核架构上的性能表现做出相应的调整，并从代码并行化和编译器优化两方面进行优化；最后根据 Amdahl 定律和 Gustafson 定律做出扩展性分析和客观性能评价。本文研究的特色与创新一是将 Hausdorff 距离算法由传统的串行运算改造为 IA 多核架构上的高性能多核并行算法，并成功地应用于侧视建筑物识别定位系统和印刷板检测系统。二是采用崭新技术和解决方案进行并行代码的性能分析，实现计算软件的算法并行优化、编译优化，提高其运算效率和适应硬件发展的可扩展性。本文采用的技术路线和方法带有普遍性，可以推广到其它图像处理函数的并行化改造。

关键词： Hausdorff 距离；多核并行；性能优化；建筑物识别

Abstract

Hausdorff distance is an important point characteristics matching method. Widely used in image processing, astronomy, mathematics, network applications, medical, economic forecasts, and other fields, especially have a very wide range of applications in the image matching. The traditional Hausdorff distance algorithm has complex calculation and low efficiency, improves the efficiency of the algorithm is useful and important. On the other hand, multi-core technology is the current hot researching field of computer, and significantly improves the computer's calculation ability. It will be a widely popular computing model. However, to really highlight the advantages of multi-core processors, software development must keep up with the pace of hardware, how to develop in line with the multi-core software is increasingly becoming a hot research field of computer technology.

This paper first studies the current computer hardware and software technology development trends, then studies the paralleled way around the Hausdorff distance algorithm, aiming at finding a simple and efficient parallel support scheme on multi-core platform. This paper first introduces multi-core technology, multi-core architecture, common developing environment and application of software tools for comparative analysis, then finds the technologies and solutions suitable for play on the performance of multi-core programming. After a detailed analysis of the Hausdorff distance algorithm, this paper designs a Multi-core structure of parallel algorithm, and successfully applied to the building matching and locating system, the algorithm also can apply to the printing plate detection system; Next we use OpenMP shared memory programming and the combination of Intel VTune Performance Analyzer, Intel Thread Checker and Intel C++ Compiler, and other tools and solutions for developing and

testing, then according to the performance in the framework of multi-core platform to make corresponding adjustments, and make two aspects of optimization from the code Parallel and compiler optimization; Finally, according to Amdahl's and Gustafson's Law to make expansion analysis, and make an objective evaluation. One characteristic innovate is making the Hausdorff distance algorithm from the traditional serial method to multi-core parallel computing way suitable for the multi-core structure of the IA's high-performance structure, and successfully applied to the buildings matching and locating system and the printing plate detection system. The other is the performance analysis of parallel code by the newest technology, which realizes algorithm optimization, code level optimization and compiler level optimization, enhances the operation efficiency and the adaptation of hardware development. The technical route used in the subject has the universality, may promotes to the parallelization of other function libraries.

Key Words: Hausdorff Distance; Multi-core Parallel Computing; Performance Optimization; Building Recognition

目 录

第一章 绪论	1
1.1 研究背景和意义	1
1.2 研究现状	3
1.3 研究内容和技术难点	4
1.4 论文结构安排	5
第二章 并行计算技术与多核架构	7
2.1 并行计算机	7
2.1.1 并行计算机的分类.....	7
2.1.2 存储方式.....	9
2.2 并行计算	9
2.2.1 并行算法设计原则和方法.....	9
2.2.2 并行编程模型.....	10
2.3 多核架构	12
2.3.1 处理器体系结构基础.....	13
2.3.2 多核与超线程.....	14
2.4 多核平台并行程序设计流程	16
2.5 多核并行程序设计的关键问题及解决方法	17
2.6 OpenMP 多线程编程及性能优化	18
2.6.1 OpenMP 并行编程模型	19
2.6.2 OpenMP 并行优化技术	19
2.6.3 OpenMP 多线程应用程序性能分析	20
2.7 本章小结	21
第三章 Hausdorff 距离算法及其在图像匹配中的应用	23
3.1 Hausdorff 距离	23
3.1.1 Hausdorff 距离定义	23
3.1.2 部分 Hausdorff 距离算法	25
3.1.3 考虑模板变换的 Hausdorff 距离	26
3.2 Hausdorff 距离算法在建筑物匹配中的应用	27

3.2.1 使用 Hausdorff 距离算法进行图像匹配	28
3.2.2 相关技术	29
3.3 Hausdorff 距离算法在其它领域的应用	33
3.4 本章小结	34
第四章 多核并行 Hausdorff 算法构建优化与性能评价.....	35
4.1 并行方案和并行编程模式	35
4.2 多核并行 Hausdorff 算法的优化	38
4.2.1 性能数据的收集与分析	39
4.2.2 并行优化与正确性检查	42
4.2.3 编译器优化	46
4.3 并行 Hausdorff 算法的性能评价	49
4.4 本章小结	52
第五章 实验结果与分析.....	55
5.1 实验环境	55
5.2 测试结果与分析	55
第六章 总结与展望.....	75
参考文献.....	77
研究生期间发表的论文.....	82
致谢.....	83

Contents

Chapter 1 Introduction	1
1.1 Research background	1
1.2 Research actuality	3
1.3 Main job and difficulty	4
1.4 Organization of this paper	5
Chapter 2 Parallel computing and multi-core architecture ..	7
2.1 Parallel computer.....	7
2.1.1 Parallel computer classification.....	7
2.1.2 Storage way.....	9
2.2 Parallel computing.....	9
2.2.1 Parallel algorithm design methods.....	9
2.2.2 Parallel programming pattern.....	10
2.3 Multi-core architecture	12
2.3.1 Processor architecture.....	13
2.3.2 Multi-core and Hyper-Threading.....	14
2.4 Parall programming design process.....	16
2.5 Key issues of parall programming design and resolution	17
2.6 OpenMP multi-threading programming and optimization	18
2.6.1 OpenMP parallel programming.....	19
2.6.2 OpenMP parallel optimization technology.....	19
2.6.3 OpenMP programme performacen analysis.....	20
2.7 Summary.....	21
Chapter 3 Hausdorff distance algorithm and its application in	
image matching	23
3.1 Hausdorff distance.....	23
3.1.1 Definition of Hausdorff distance.....	23
3.1.2 Partial Hausdorff distance algorithm.....	25
3.1.3 model transformation in Hausdorff distance.....	26

3. 2 Hausdorff distance algorithm in image matching	27
3.2.1 Image matching using Hausdorff distance algorithm....	28
3.2.2 Related technology.....	29
3. 3 Application of Hausdorff distance in other fields	33
3. 4 Summary.....	34
Chapter 4 Hausdorff distance multi-core parallel computing algorithm optimization and performace evaluation	35
4. 1 Parallel programme and parallel programming pattern.....	35
4. 2 Hausdorff distance multi-core parallel computing algorithm optimization	38
4.2.1 Performance data collection and analysis.....	39
4.2.2 Parallel optimization and accuracy check.....	42
4.2.3 Compiler Optimization.....	46
4. 3 Parallel Hausdorff distance algorithm evaluation and analysis.....	49
4. 4 Summary.....	53
Chapter 5 Experiment and analysis.	55
5.1 Experiment circumstance.....	55
5.2 Experiment result and analysis.....	55
Chapter 6 Conclusions	75
References	77
Researches	82
Acknowledgments	83

第一章 绪论

1.1 研究背景和意义

Hausdorff 距离作为一种重要的距离测度,常用于衡量两个点集合之间的相似程度,它是两个点集之间距离的一种定义形式[1]。使用 Hausdorff 距离作为距离测度时无需考虑两个点集中的点与点之间的对应关系,在方法上具有简单性和有效性,因此具有适应性强的特点,已被广泛应用于图像处理、网络应用、机械设计、经济预测、气象、天文、数学、控制、生物、医学等众多领域中,特别是在图像处理与计算机视觉领域的经典问题之一——图像匹配中使用广泛。

但是,传统的 Hausdorff 距离匹配,其计算复杂性相当高[1],计算量很大,耗费大量时间。而当今信息时代,计算机所需处理的数据量日益庞大,面对海量数据,特别是一些对速度要求较高的系统,如何缩短运行时间,提高算法的运算效率便成为亟待解决的问题。

另一方面,随着计算机技术的不断发展,提升主频引发散热和功耗的大幅增加等问题,提升 CPU 主频的难度越来越大,于是 Intel、AMD 和 Sparc、PowerPC 等主要的厂商调整研究方向,向超线程和多核架构靠拢。随着操作系统及应用软件对多核处理器的进一步支持及优化、芯片制造工艺的成熟、AMD 及 Intel 为代表的低功耗技术的发展、芯片级虚拟化技术的成熟等诸多因素,将推动服务器处理器多核化趋势的进一步彰显,多核技术将成为服务器技术的重要技术支点,多核时代真正来临[2, 3]。

应用需求的不断提高是计算机发展的根本动力。多核架构的优势十分明显,它能满足人们对更高性能的需求,特别是在一些大型企业的 ERP、CRM 等复杂应用,科学计算、政府的大型数据库管理系统、数字医疗领域、电信、金融等都需要高性能计算,多核技术可以满足这些应用的需求。多个执行内核同时运算,因此计算能力显著提升,而且允许每个内核主频比以前低,性能发挥到极致的同时,功耗也得到很好的控制,总体功耗增加不大[4]。其次,与多 CPU 相比,多核处理器采用与单 CPU 相同的硬件架构,用户提升计算能力的同时无需改变任何硬

件, 使用非常方便。正是由于这些优点, 多核很快被用户接受, 并得以普及。根据 IDC 的一份报告, 到今年年底 80%~90% 出货的 PC 其 CPU 都是多核的[1]。

不难看出, 多核处理器是处理器发展的必然趋势[4]。无论是移动/嵌入式应用、桌面应用还是服务器应用, 都将采用多核的架构, 众多软件供应商如 Google、Microsoft、Oracle 以及 Linux 的软件供应商也都积极着手开发软件向多核时代过渡。但并不是所有的操作系统和应用软件都做好了迎接多核平台的准备, 微软软件架构师 Herb Sutter 曾指出: 软件开发者对多核处理器时代的来临准备不足。他说, 软件开发社区认识到处理器厂商被迫采用多核设计以应对处理器速度提升带来的发热问题, 但却没有清楚地了解这样的设计为软件开发带来多少额外的工作[5]。过去单核架构下, 开发者只需对现有软件程序作轻微改动就能坐观其性能在随着硬件性能的上升而不断提升。多核设计概念的出现迫使软件世界不得不直面并行性(将单个任务拆分成多个小块以便分别处理之后再重新组合的能力)问题。如果不针对多核进行软件开发, 不仅多核提供的强大计算能力得不到利用, 相反还可能不如单核 CPU 高效。多核处理器最迫切需求配套的是系统软件支持和应用软件开发, 软件针对处理器进行性能调优才能使基于多核处理器的软件实现最佳性能。就已有的串行应用软件而言, 那些在单核处理器上开发的不支持多线程多内核技术的应用, 尽管仍然可以在新一代处理器的体系结构上运行, 但显然不能充分利用多核处理器架构带来的性能优势, 这部分软件的并行处理需要解决。从软件角度来看, 多线程、多内核意味着相同的编程问题, 开发者要编写的是能同时完成多个任务的软件[6, 7]。一方面要求软件开发者(ISV)换种思路, 摒弃传统的串行编程方法, 去掌握多线程并行的编程理念; 另一方面则是对并行编程环境还有待完善的渴求, 当前比较流行的并行编程环境主要有消息传递、共享存储和数据并行三大类, 其中的典型代表是 MPI 和 PVM, OpenMP 和 HPF 也占一定比例[8]。基于对新的高生产率并行编程语言的需求, 国际上对新一代并行程序设计语言的研究正日渐升温, 如 IBM 的 X10、Sun 公司的 Fortress 以及 Cray 公司的 Chapel 等语言已经出现, 而且正在快速发展, 有待进行大规模推广。

针对多核架构开发的软构件在这一机遇中崭露头角, 软构件库运行于形形色色的多核架构之上, 对新一代基于 Web 应用与传统软件应用提供无缝支持, 满足大量应用的需要, 具有通用性、复用性、高效灵活等优势。针对多核和多线程的

软件开发将是未来十年软件开发的主要挑战，基于多核架构的并行计算将是软件开发史上的又一个重大变革。

1.2 研究现状

自从 Hausdorff 距离提出之后，由于它的计算的简单有效性，已经广泛应用于多个领域，特别是应用于图像处理中的图像匹配中。但 Hausdorff 距离是一个最大最小距离，易受突发噪声影响。因此，1993 年，Huttenlocher D. P. 等人研究了基于 Hausdorff 距离的图像之间的相似性，提出了局部 Hausdorff 距离算法 [1]，该算法对干扰点或背景点相对不敏感，并且允许形状的一部分与另一个图像进行匹配。对于匹配有严重遮掩和退化的图像中的部分图像，也有较好的效果；Dubuisson 和 Jain 提出了基于平均距离值的 Hausdorff 距离 (MHD)，它采用了对所有距离求平均值的运算，将单个元素代表整个集合的思想扩展为集合内所有元素的平均作用，很大程度上提高了算法对噪声的适应性，可避免由于部分噪声像素点的干扰所带来的偏差。这些方法虽然比较有效地解决了 Hausdorff 距离对噪声敏感的缺陷，但是，对 Hausdorff 距离算法计算量大，耗费大量时间这一重要缺陷没有解决。

改进 Hausdorff 距离算法效率的研究大多都是上在单核机器上进行的。例如 1993 年，Huttenlocher D. P. 等人 [1] 提出的使用动态规划裁剪空间的方法，通过一定的判断，剔除不满足条件的空间，一定程度上提高了 Hausdorff 距离的计算速度。1997 年，Rucklidge [10] 提出的子空间分解、盒距离变换算法极大加速了 Hausdorff 距离的计算速度，使 Hausdorff 距离有了一定的实用价值。

1992 年，Gualtieri J. A 等人通过使用 MasPar Parallel 并行处理机实现 Hausdorff 的并行化 [11]，这种方法需要特定的硬件设备，并且对需要专业的编程人员才能实现。

1995 年，Jane you 等人在分布式系统上采用 RPC（远程过程调用）实现了 Hausdorff 算法的并行移植 [12]，这种方法无需特定的软件。

Jinbo Xu 等人利用 FPGA 芯片在嵌入式系统中实现了 Hausdorff 距离算法的并行处理 [13]。

You J 等人将 Hausdorff 距离算法在 PVM (Parallel Virtual Machine) 网

络环境下实现了分布式并行[14],这种方法将图像分割成多个子图像让它们在网络上的不同节点中分别进行计算,以此来提高计算效率。

上述 Hausdorff 距离算法大多是使用分布式并行计算机实现的,需要特定的硬件环境,并且对编程人员的要求较高,入门较难。而且,随着多核技术的发展和普及,多核下的并行计算必将成为一种趋势,如何在多核系统下采用共享存储的方式实现算法并行是一个需要研究的课题。

1.3 研究内容和技术难点

本文的研究目标是针对 Hausdorff 距离算法计算复杂度高,计算时间大这一缺陷,对近来流行的多核计算技术,在详细剖析多核体系结构和并行算法的基础上,分析影响多核并行计算性能的主要因素,探讨适合于发挥多核性能的算法和编程技术,使用 OpenMP 等工具实现了 Hausdorff 距离并行算法,有效提高算法效率。

本文的主要工作有:

(1) 对多核体系结构、常用开发环境和适用软件工具进行比较分析,探讨适合于发挥多核性能的编程技术和解决方案。

(2) 详细剖析 Hausdorff 距离算法基础上对其进行一定的改进,设计基于多核架构的并行算法,并成功应用于侧视图像中建筑物目标的匹配识别系统,同时能够适用于印刷板检测系统。

(3) 通过对并行方案和编程模式的比较分析,采用适合多核架构下的 Hausdorff 并行策略,使用 OpenMP 共享存储编程,结合 Intel VTune Performance Analyzer、Intel Thread Checker 和 Intel C++ Compiler 等工具和解决方案测试其性能,根据代码在多核架构上的性能表现做出相应的调整,并从代码并行化和编译器优化两方面进行优化。

(4) 根据 Amdahl 定律和 Gustafson 定律做出扩展性分析,得出了客观性能评价。

研究难点有:

(1) 目前多核架构上的研究较少,并局限于学术领域高性能计算,存在多种标准和不同体系,相适应的应用软件技术研究并不多见,可以借鉴的相关资

料和成果较少，因此，有关多核并行及其相关解决方案需要查找相关资料并进行大量探索和实验，同时，许多研究还有待进一步拓展和深入。

(2) 寻找并行性，这种并行性在多核架构和实现中的表现如何以及基于这种可行的并行性我们的系统加速比是多少等问题是多核并行 Hausdorff 距离算法设计中面临的问题。对并行算法代码的分析、优化、测试是一个循环渐进的过程，工作量大，需花费一定精力研究试验。同时将其应用于建筑物识别和其它应用领域，如：印刷板检测也需要一定工作。

(3) 数据相关性是多核编程中的一大难题。当两个线程对同一个变量进行操作，并且有一个操作为写操作时，这两个线程就存在数据竞争。此时，读出的数据不一定是前一次写操作的数据，而写入的数据也可能并不是程序所需要的。为了保证程序的正确性，需要仔细检查程序并采取适当的方法使程序并行化后两个线程之间不出现数据竞争的同时保证结果的正确性。

1.4 论文结构安排

本文的结构将以如下方式组织：

第一章是绪论，主要介绍课题的研究背景和意义，Hausdorff 距离算法的发展现状、课题的研究内容和技术难点。

第二章主要探讨并行计算技术、多核架构以及多核平台下的编程技术。首先介绍并行计算机体系结构、并行算法设计原则和方法，比较了消息传递、共享存储、数据并行三大类并行计算模型的主要特征，介绍多核处理器架构，然后对比概述多核平台并行程序设计流程，详细介绍 OpenMP，最后讨论多核并行程序设计的关键问题和解决方法。

第三章首先介绍了 Hausdorff 距离以及部分 Hausdorff 距离等相关内容，然后介绍了侧视建筑物识别系统以及 Hausdorff 距离算法在其中的一定改进和应用，最后将 Hausdorff 距离算法在印刷板检测系统中进行了尝试应用。

第四章通过对并行方案和编程模式的比较分析，采用适合多核架构下的 Hausdorff 并行策略，利用了 Intel 公司的一整套成熟的多核软件解决方案对算法进行了并行和优化。首先使用 Intel (R) VTune(TM) Performance Analyzer 工具分析算法代码，查找代码瓶颈，然后利用 Intel (R) Thread Checker 开发工具

分析并行性，讨论程序的并行性能优化方法，根据这种并行性构建基于多核并行计算下的 Hausdorff 算法。最后采用 Intel C++ Compiler 工具对并行化后的算法进行了分析和优化算法。

第五章是详细的实验数据及其分析，针对侧视建筑物识别应用进行了一定的测试，同时也在印刷板检测的应用中进行了简单尝试。实验表明，本文的方法有较好的效果。

第六章是总结和展望，主要是对本文工作的总结，以及目前正在进行的工作，下一步需要的研究的重点和将来工作的展望。

最后是研究生期间发表的工作总结和致谢。

Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.

厦门大学博硕士论文摘要库