

分类号\_\_\_\_\_

UDC \_\_\_\_\_

密级\_\_\_\_\_

编号\_\_\_\_\_

# 中国科学院研究生院

## 博士学位论文

面向立体视觉的匹配算法研究

刘志花

指导教师\_\_\_\_\_焦建彬教授\_\_\_\_\_

中国科学院研究生院

申请学位级别\_\_\_\_\_工学博士\_\_\_\_\_ 学科专业名称\_\_\_\_\_计算机应用技术\_\_\_\_\_

论文提交日期\_\_\_\_\_2011.4\_\_\_\_\_ 论文答辩日期\_\_\_\_\_2011.5\_\_\_\_\_

培养单位\_\_\_\_\_工程教育学院\_\_\_\_\_

学位授予单位\_\_\_\_\_中国科学院研究生院\_\_\_\_\_

答辩委员会主席\_\_\_\_\_周东华\_\_\_\_\_



## 中国科学院研究生院直属院系

### 研究生学位论文原创性声明

本人郑重声明：所提交的学位论文是本人在导师的指导下独立进行研究工作所取得的成果。尽我所知，除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的研究成果。对论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明或致谢。

作者签名：

日 期：

## 中国科学院研究生院直属院系

### 学位论文授权使用声明

本人完全了解并同意遵守中国科学院有关保存和使用学位论文的规定，即中国科学院有权保留送交学位论文的副本，允许该论文被查阅，可以公布该论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存、汇编本学位论文。

涉密的学位论文在解密后适用本声明。

作者签名：

日 期：

导师签名：

日 期：



## 摘要

图像匹配是计算机视觉和图像理解领域的基础研究问题之一，已广泛应用于多源图像数据融合、目标识别、三维重建等诸多领域。论文主要针对双目立体视觉中的匹配问题进行了研究，主要研究成果和创新之处有以下几点：

1) 在图像校正系统中，提出了一种由粗到精的特征点匹配算法。在粗匹配阶段，本文提出了互相似度算法求解特征点相似度约束模型，利用该模型可以得到初始解并将多对多的特征点匹配问题转化为一对一的匹配问题。在精匹配阶段，本文利用改进遗传算法求解特征点空间约束匹配模型，改进策略包括：整数编码、优化初始种群、稳态的选择算子、多点交叉算子等。更重要的是，本文将粗匹配的结果作为改进遗传算法的一个初始解，这将使得匹配结果更加准确。在三个标准测试集上对提出的算法进行了实验和比较，实验结果表明本文提出的算法对于不同视角、不同光照、不同尺度下的宽基线图像具有更高的匹配精度。

2) 针对经典置信传播算法中存在计算冗余、低精度等问题，提出了一种自适应的置信传播算法，该算法将有限消息序列收敛和低纹理区域检测这两个思想融入到置信传播算法中。在标准数据集上的实验表明，与经典置信传播算法相比，所提出算法的计算时间减少了大约 50%，同时，计算精度也有明显提高。

3) 提出了一种基于视差校正和视差优化的立体匹配算法。该算法将整个匹配过程分为三个阶段：初始匹配、视差校正和视差优化。在视差校正阶段，提出了一种区分遮挡像素和误匹配像素的分类方法。在视差优化阶段，提出了一种渐进的区域融合算法，该算法以大区域和置信系数高的区域为种子区域，然后逐步与其相邻的小区域进行比较与合并，通过不断优化视差平面，得到比较精确的视差图。

**关键词：**遗传算法，置信传播算法，自适应置信传播算法，视差校正，视差优化

## Abstract

### Matching Algorithm Research for Stereo Vision

**Liu Zhihua** (Computer Application)

Directed by: **Jiao Jianbin** (Professor)

Image matching is a hot topic in the field of computer vision and image understanding, which has been widely used in multi-sensor image fusion, target recognition and 3D reconstruction. In this paper, we investigate the matching problem in binocular stereo vision. The main research results and innovation are as follows:

1) In image rectification procedure, a new coarse-to-fine feature point matching algorithm is proposed. In the coarse matching stage, a Mutual Similar (MS) algorithm is proposed to acquire the initial solution of feature point similarity constraint model and many-to-many correspondence is converted to one-to-one correspondence. In the refine matching stage, an improved genetic algorithm is used to solve the spatial constraint model and some improvements include: integer encoding, initial optimized population, steady-state selection operator, multi-points crossover operator are proposed. More importantly, coarse matched result is considered as a solution in the initial population of genetic algorithm, which is more reliable to the solution. Experiments on three image data sets demonstrate that the proposed algorithm has better performance under different viewpoint, large scale, illumination change and wide separation.

2) For the problems, such as redundancy in computation, low accuracy, etc. in belief propagation algorithm, an adaptive belief propagation (ABP) algorithm for stereo matching is proposed in this paper. In the algorithm, the definition of finite message sequence's convergence with the detection of textureless regions is proposed and the two ideas are added into BP algorithm, which will not only reduce redundancy in computation but also make all points in the image approximately convergence to the global optimal solution. The proposed ABP algorithm is evaluated and compared on the Middlebury data sets with a BP algorithm. Experimental results show that it can reduce about 50% computational costs with higher accuracy compared to the BP algorithm.

3) A stereo matching algorithm based on disparity rectification and disparity optimization is proposed in the paper. The matching algorithm is divided into three phases: initial matching, disparity correction and disparity optimization. In disparity correction phase, a new classification

method is proposed to classify occluded pixels with mismatched ones. In disparity optimization phase, a progressive region merging algorithm is proposed. The algorithm regards regions of larger size and higher confidence as seed regions and gradually compare and merge with its neighbors. Finally, an optimized disparity plane is obtained and disparity map is constructed.

**KEY WORDS:** Genetic algorithm; Belief propagation algorithm; Adaptive belief propagation algorithm; Disparity rectification; Disparity optimization



## 目录

摘要.....	I
Abstract.....	II
目录.....	V
图目录.....	IX
表目录.....	XI
第1章 绪论.....	1
1.1 课题的背景与研究意义.....	1
1.2 国内外研究现状.....	2
1.3 论文的主要工作.....	6
1.4 论文的内容安排.....	7
第2章 立体视觉的基础理论与方法.....	9
2.1 摄像机标定理论与方法.....	9
2.1.1 摄像机模型.....	9
2.1.2 线性摄像机成像模型.....	12
2.1.3 极线几何与基础矩阵.....	14
2.1.3.1 极线几何.....	14
2.1.3.2 标准的双目立体几何.....	14
2.1.3.3 基础矩阵.....	15
2.1.3.4 基础矩阵的计算.....	17
2.2 立体匹配理论与方法.....	19
2.2.1 立体匹配理论.....	19
2.2.1.1 视差与深度的关系.....	19
2.2.1.2 立体匹配的难点.....	20
2.2.1.3 立体匹配的标准测试集以及评价标准.....	21
2.2.2 立体匹配方法.....	23
2.2.2.1 匹配测度.....	23
2.2.2.2 局部匹配方法.....	25
2.2.2.3 全局匹配方法.....	26
2.2.2.4 基于分割的全局匹配方法.....	28
2.3 本章小结.....	28
第3章 图像校正.....	31
3.1 特征点提取.....	31

3.1.1 Harris 角点 .....	32
3.1.2 SIFT 特征点 .....	33
3.1.2.1 SIFT 特征点提取 .....	34
3.1.2.2 SIFT 特征描述子的表示 .....	36
3.2 基于改进遗传算法的特征点匹配 .....	37
3.2.1 建立匹配问题的数学模型 .....	38
3.2.2 特征点粗匹配 .....	40
3.2.3 特征点精匹配 .....	41
3.2.3.1 编码 .....	41
3.2.3.2 种群的初始化 .....	42
3.2.3.3 选择算子 .....	42
3.2.3.4 交叉算子 .....	43
3.2.4 实验结果与比较 .....	43
3.3 双目图像对校正 .....	46
3.3.1 摄像机内参数自标定 .....	46
3.3.2 校正后的结果 .....	47
3.4 本章小结 .....	48
第 4 章 基于自适应置信传播算法的立体匹配 .....	51
4.1 置信传播算法原理 .....	51
4.2 置信传播算法目前存在的问题 .....	53
4.3 自适应置信传播算法 .....	55
4.3.1 图像分割 .....	55
4.3.2 低纹理区域检测 .....	55
4.3.3 自适应置信传播算法 .....	56
4.4 实验结果和比较 .....	58
4.4.1 精度比较 .....	58
4.4.2 性能比较 .....	59
4.5 本章小结 .....	61
第 5 章 基于分割的全局立体匹配 .....	63
5.1 初始视差计算 .....	63
5.2 视差校正 .....	64
5.2.1 左右一致性检查 .....	64
5.2.2 图像分割 .....	66
5.2.3 低纹理区域检测和误匹配像素检测 .....	66
5.2.4 像素分类 .....	67

5.2.5 视差校正.....	68
5.3 视差优化 .....	69
5.3.1 计算初始视差图.....	69
5.3.2 优化视差平面.....	70
5.4 实验结果与比较 .....	72
5.4.1 遮挡图的定量比较.....	72
5.4.2 实验结果.....	74
5.4.3 实验定量比较.....	75
5.5 本章小结 .....	76
总结与展望.....	77
参考文献.....	79
发表文章目录.....	87
致谢.....	89



## 图目录

图 1-1	双目立体视觉的应用	2
图 1-2	双目立体视觉框架图	6
图 2-1	图像像素坐标系与图像物理坐标系	10
图 2-2	摄像机成像模型	11
图 2-3	成像过程	12
图 2-4	对极几何	14
图 2-5	双目立体视觉设备 <sup>[44][45][46]</sup>	15
图 2-6	数据的规范化变换	18
图 2-7	标准双目结构下视差与深度之间的关系图	20
图 2-8	标准左右图像及标准左视差图	22
图 2-9	Teddy 的原图以及三类典型区域的图像	23
图 2-10	块匹配示意图	26
图 2-11	由动态规划算法得到的视差图	27
图 2-12	基于分割的立体匹配算法流程图	28
图 3-1	窗口在图像上移动示例 <sup>[58]</sup>	32
图 3-2	Harris 角点检测	33
图 3-3	高斯差分图	34
图 3-4	尺度空间局部极值点检测 <sup>[63]</sup>	34
图 3-5	利用梯度方向直方图确定主梯度方向 <sup>[63]</sup>	35
图 3-6	提取 SIFT 特征点图	35
图 3-7	由关键点邻域梯度信息生成特征向量 <sup>[63]</sup>	36
图 3-8	SIFT 特征描述子	37
图 3-9	遗传算法求解流程图	38
图 3-10	特征点粗匹配结果	41
图 3-11	三点交叉策略 <sup>[84]</sup>	43
图 3-12	实验结果比较	44
图 3-13	实验结果比较	45
图 3-14	实验结果比较	45
图 3-15	标定板图像	46
图 3-16	双目校正后的实验结果	47
图 3-17	校正后的实验结果	48

---

图 4-1	马尔可夫随机场模型.....	52
图 4-2	BP 算法中消息传播图.....	52
图 4-3	BP 算法迭代次数与运行时间变化图.....	54
图 4-4	图像分割.....	55
图 4-5	低纹理区域检测图.....	56
图 4-6	在 Middlebury 数据集上的结果.....	59
图 4-7	随着迭代次数增加不收敛像素所占比例的变化.....	60
图 4-8	ABP 算法和 BP 算法迭代时间的比较.....	61
图 5-1	基于分割的全局立体匹配算法流程图.....	63
图 5-2	初始视差图.....	64
图 5-3	初始可靠像素的视差图.....	65
图 5-4	图像分割.....	66
图 5-5	视差图.....	67
图 5-6	像素分类图.....	68
图 5-7	校正后的视差图.....	69
图 5-8	遮挡图结果比较.....	73
图 5-9	本文算法在 Middlebury 数据集上的结果.....	74
图 5-10	本文算法在其他数据集上的结果.....	75

## 表目录

表 3-1	匹配精度比较.....	46
表 4-1	ABP 算法和 BP 算法的精确度比较.....	59
表 4-2	ABP 算法和 BP 算法的运行时间比较.....	61
表 5-1	遮挡图的精度比较.....	72
表 5-2	遮挡图的定量结果比较.....	74
表 5-3	本文算法与其他算法定量比较结果.....	75



## 第1章 绪论

### 1.1 课题的背景与研究意义

视觉是人类感知外部环境、认知世界的主要途径。据统计，人类大约有百分之八十的外部信息是通过视觉途径获取的。这既说明了视觉信息量的巨大，也表明人类对视觉信息有较高的利用率。当人类的眼睛从周围环境获取信息并传入大脑后，大脑根据知识或经验，对信息进行加工、推理等处理工作，最后识别、理解周围环境，包括环境内的对象物，如运动物体，物体之间的相对位置、形状、大小、颜色、纹理等。因此，人类的视觉过程可看成是一个复杂的从感觉到知觉的过程。

随着计算机技术的发展，人类希望计算机能够替代他们与周围的环境交换信息，这就导致了计算机视觉这门学科的诞生<sup>[1]</sup>。

计算机视觉就是用计算机模拟人眼的视觉功能，即使得计算机具有通过二维图像认知三维环境信息的能力。这种能力将不仅使机器能感知三维环境中的物体的集合信息，包括形状、位置、姿态、运动等，而且能对其进行描述、存储、识别和理解，进而做出决断。

计算机视觉所具有的潜在应用十分广泛，所涉及的学科知识繁多，研究的问题又极有挑战性，因此它是计算机学科中最热门的研究课题之一，吸引了许多从事心理学、生物学、神经科学、生物物理学、数学与计算机等各种学科研究人员的关注，从而把图像处理，模式识别、人工智能、数学、认知科学、机器学习、计算机图像学等各方面的研究领域融汇进来。

双目立体视觉直接模拟人类双眼处理景物的方式，是计算机视觉中最常用的一种方法。它利用两台位置相对固定的相机，从不同角度同时获取统一景物的两幅图像，通过计算空间点在两幅图像中的视差来获得景物的三维坐标值。它可以非接触方式，自动、在线地获取场景三维信息，并且具有结构简单、灵活可靠、使用范围广等优点，因而具有十分广阔的应用前景，在三维重建、自主车导航、工业检测、基于图像的建模和绘制等领域具有较高的应用价值。

在三维重建和建模方面，双目立体视觉方法采用数字摄像机作为图像传感器，综合运用图像处理、视觉计算等技术进行非接触三维数据获取，用计算机程序自动获取物体的单位信息。该方法效率高，是三维重建和建模的一个重要发展方向，已广泛应用于交通事故现场重建、汽车建模、文物艺术品的重建、三维医学图像重建等领域。

在机器人、自主导航方面，利用双目立体视觉系统获取三维信息是目前最常用的技术，双目立体视觉系统是被动态测量方法，具有无辐射源，隐蔽性好，测量快速准确，尺寸小

等优点。在 100 米距离范围内，其测量精度与激光雷达测量精度相近。随着高性能处理器的出现，双目立体视觉算法的实时处理成为可能，将会在该领域得到更广泛的应用。

图 1-1 是双目立体视觉在实际应用中的一些图片<sup>[2]</sup>。其中最后一幅图片是作者所在实验室开发的自主导航爬壁小车系统实物图。

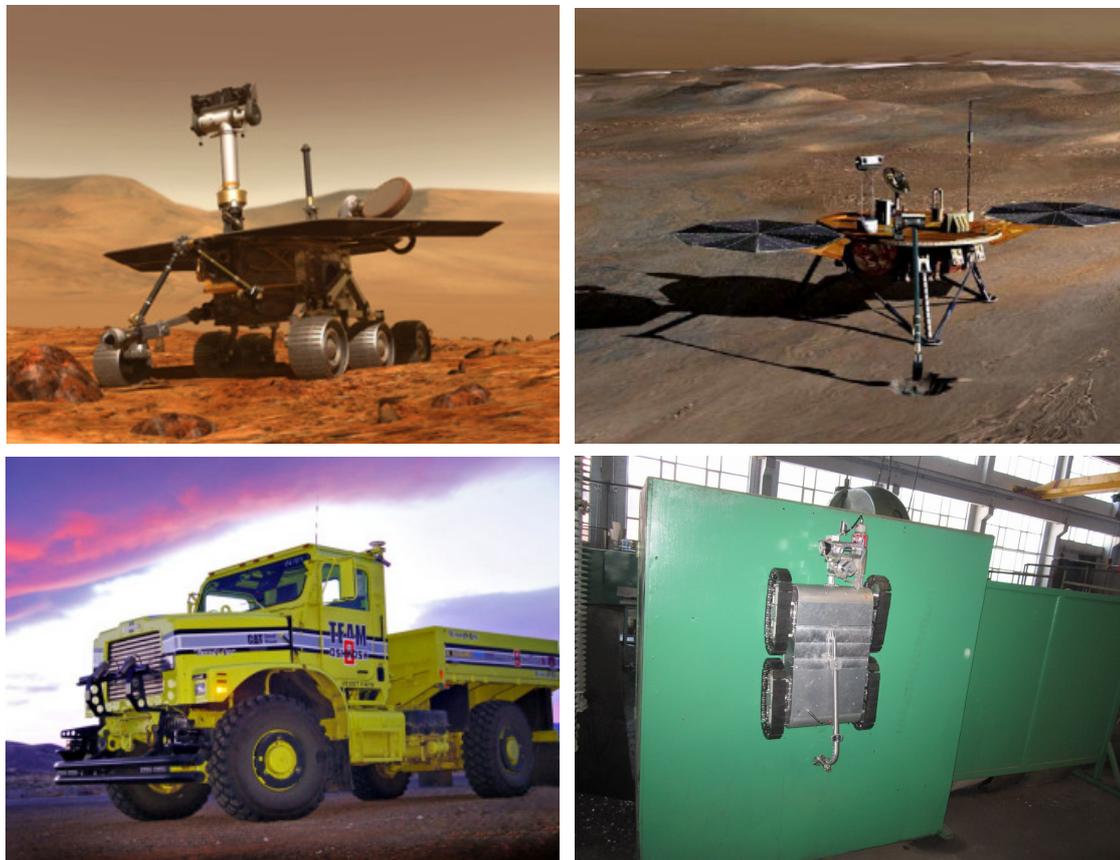


图 1-1 双目立体视觉的应用

本研究得到以下课题资助：

1. 基于多源异构信息的飞行器威胁目标检测跟踪及行为预测（国家自然科学基金重点项目）；
2. 风机塔筒焊缝自动检测系统（爬壁机器人）（航天科工风华机械厂横向课题）。

## 1.2 国内外研究现状

随着计算机视觉、数字图像处理技术的不断完善及计算机软硬件的快速发展，双目立体视觉系统在国内外得到了迅速发展。

立体匹配的发展始于二十世纪七十年代中期，以 Marr 和 Poggio<sup>[3]</sup>等人为代表的一些研究者提出了一整套视觉计算的理论来描述视觉过程，其核心是从图像恢复目标的三维形状。Barnard 和 Fischler<sup>[4]</sup>总结了 1981 年以前的立体匹配研究，主要集中在立体重建的基本原理、评估标准和当时比较流行的方法。

80 年代, 立体匹配成为计算机视觉的研究焦点之一。Dhond 和 Aggarwal<sup>[5]</sup>总结了 80 年代立体匹配的进步, 包括大量的新匹配方法、层次处理观念的引进以及利用三目约束来降低立体匹配的模糊性。到了九十年代早期, 立体匹配在各方面都逐渐成熟起来, 尽管通用的立体匹配算法的研究仍在继续, 但很多研究者这时把目光转向立体匹配在待定问题中的应用, 如摄影测量、三维重建、虚拟现实等。

Koschan<sup>[6]</sup>总结了 1989 年到 1993 年期间立体匹配技术的发展, 包括对遮挡的早期研究, 主动和动态立体匹配以及实现的实时性问题, 这些方面的研究在过去十年中都取得了很多成果。现在立体匹配又有了新的发展趋势并取得了重大发展, 包括基于区域和基于特征匹配的新方法、处理遮挡问题的方法、多相机立体匹配以及实时实现等。

Myron Z. Brown<sup>[7]</sup>等总结了 1993 到 2003 年期间立体匹配技术的发展, 着重于遮挡处理和实时系统。2002 年 Scharstein 和 Szeliski<sup>[8]</sup>将立体匹配过程分为四个模块: 相似性测度计算、支撑选择、视差计算及视差校正。他们把当前的各种匹配算法归结为模块相互变动的结果, 在充分了解变动因素的前提下, 通过对各种算法之间的性能比较, 得出变动因素对算法性能产生的具体影响。

国内外有很多政府机构、公司、大学、研究所致力于立体匹配基础算法的研究。国外主要有卡耐基梅隆大学的机器人技术研究所的视觉和自动化研究中心、英国哥伦比亚大学的智能计算实验室的 Point Grey 研究组<sup>[9]</sup>、德国波恩大学的计算机视觉研究组、日本大阪大学自适应机械系统研究院等。国内主要有清华大学自动化系、中科院自动化技术研究所、浙江大学、东南大学电子工程系、微软亚洲研究中心等。在立体匹配算法应用方面, 日本东京大学将实时双目立体视觉和机器人整体姿态信息集成, 开发了仿真机器人动态行走导航系统。该系统的实现分两个步骤: 首先, 利用平面分割算法分离所拍摄图像对中的地面与障碍物, 再结合机器人躯体姿态的信息, 将图像从摄像机的二维平面坐标系转换到描述躯体姿态的视觉坐标系, 建立机器人周围区域的地图。其次根据双目立体视觉系统进行障碍物检测, 并结合实时建立的地图确定机器人的行走方向。麻省理工学院计算机系提出了一种新的用于智能交通工具的传感器融合方法。该方法中雷达系统提供目标深度的大致范围, 利用双目立体视觉提供粗略的目标深度信息。结合改进的图像分割算法, 该方法能够在高速环境下对视频图像中的目标位置进行分割, 而传统的分割算法难以在高速实时环境中得到令人满意的结果。华盛顿大学与微软公司合作为火星卫星“探测者”号研制了宽基线立体视觉系统, 使“探测者”号能够在火星上对其即将跨越的几千米内的地形进行精确的定位和导航。系统使用同一个摄像机在“探测者”的不同位置上拍摄图像对, 拍摄间距越大, 基线越宽, 能观测到的地貌越远。系统采用非线性优化得到两次拍摄图像时摄像机的相对准确的位置, 利用鲁棒性强的最大似然概率法结合高效的立体搜索进行匹配, 得到亚像元精度的视差, 并根据此视差计算图像中各点的三维坐标。相比传统的体视系统, 该系统能够更精确地绘制“探测者”号周围的地貌和以更高的精度观测到更远的地形。

在国内, 通过广大科研工作者的共同努力, 双目立体视觉技术同样已具有广泛的应用。

哈尔滨工业大学采用异构双目活动视觉系统实现全自主足球机器人导航，将一个固定摄像机和一个可以水平旋转的摄像机，分别安装在机器人的顶部和中下部，可以同时监视不同方位视点，体现出比人类视觉优越的一面，通过合理的资源分配以及协调机制，使机器人在视野范围达到最佳匹配。双目协调技术可使机器人同时捕捉多个有效目标，观测相同目标时通过数据融合，也可以提高测量精度<sup>[17]</sup>。

浙江大学设计并实现了一种能够在道路和野外连续的实时的自主运动的智能移动机器人，该移动机器人通过双目立体视觉系统对各种道路场景进行快速识别和理解，并能通过智能控制技术使自主车沿规划的路径行走<sup>[18]</sup>。

近十年来，实时稠密立体匹配已经成为现实，但是在二十世纪九十年代期间，真正能够实现实时处理的系统都需要像数字信号处理器或者可编程门阵列这样的专用硬件支持。在本世纪初，随着计算机系统性能的不不断提高，一些仅依赖于软件的实时系统开始出现。如 Mulligan 等<sup>[10]</sup>在 2002 年提出的三目算法在单处理器的 PC 机上可以达到每秒处理 3-4 帧的速度；Hirschmuller 等<sup>[11]</sup>提出的变窗口立体匹配方法在 Pentium II 450Mhz 上可以达到每秒计算 12M 视差估计值的速度；Point Grey Research 的算法在 2.8G 的处理器上每秒可计算 80M 的视差估计值。

就目前的研究进展来看，立体实时系统的发展趋势有三个：一是处理速度越来越快，由早期的每秒 3.6 帧到后来的每秒 30 帧，这为实际应用系统提供了更好的基础；二是中心处理器从早期的专用处理器（如 DSP，FPGA）向通用的 CPU 芯片发展，这使得实时系统可以在普通电脑上实现，扩大了应用范围；三是匹配方法由简单的互相关或绝对差等方法向复杂的全局方法（如 DP）过渡，因此在保证效率的前提下得到的视差图质量在不断提高，以期满足实际需求。

立体匹配方法的有效性依赖于三个问题的解决：选择正确的匹配基元、寻找特征之间的本质属性及建立能正确匹配所选特征的稳定算法<sup>[19]</sup>。围绕这三个问题，目前研究者已提出了大量各具特色的匹配算法。从实现立体匹配的技术上考虑，这些算法可以分为局部算法和全局算法两大类。局部算法中，根据匹配基元的不同可分为基于区域的匹配（Area-based Matching）、基于梯度的优化方法（Gradient-Based Optimization）和特征匹配（Feature Matching）。其中基于区域的窗口匹配（为方便起见，以后简称区域匹配）和特征匹配是最常见的两类方法。

#### ● 区域匹配

区域匹配算法在局部小区域内寻找最佳视差使匹配误差最小，通常使用归一化互相关、平方差和、归一化平方差和、绝对差和等测度作为匹配的相似性测量准则。区域匹配方法一般是根据图像的灰度信息和窗口的区域相似度进行匹配的，因此它对于图像的旋转、光强和对比度的变化、仿射畸变和辐射畸变等变形非常敏感。区域匹配方法中窗口大小的选择也很重要，窗口过大会导致在视差不连续处出现误匹配，窗口过小会使得区域内的灰度分布特性得不到充分的体现。许多学者对此进行了研究，并提出了许多方法<sup>[12]-[16]</sup>，

其中比较经典的是 Kanade 和 Okutomi<sup>[12]</sup>于 1994 年提出的通过测算局部灰度和视差变化来选取适当窗口的方法及 2006 年 Yoon, K.-J. 与 I.-S. Kweon<sup>[14]</sup>提出的基于颜色和距离的自适应加权的匹配方法。

### ● 特征匹配

利用图像中较为明显的特征进行匹配，特征匹配基元包含了很好的统计特性以及算法编程上的灵活性。算法的许多约束条件均能清楚的应用于数据结构。目前常用的匹配基元主要有点特征、线状特征和区域特征等。点特征是最基本、最简单的特征基元；区域基元具有最好的全局属性；线状基元（包括直线和曲线）介于二者之间。一般来讲，点特征具有定位准确、检测和描述容易以及重建精度高的优点，但它所含图像信息较少，在图像中的数目较多，因而在匹配时需要较强的约束准则和匹配策略，以克服歧义匹配和提高运算效率。线状基元和区域基元则含有更丰富的图像信息，在图像中的数目较少，易于实现快速匹配。但它们在特征提取和描述时比较困难，需要进行复杂的预处理，而且定位精度较差。但自 1999 年 Tao<sup>[20][21]</sup>等利用线性方程来建模分割后所得区域内的视差与图像坐标之间的关系后，使得以分割后的区域为基元的方法逐渐得以流行。

全局算法使用全局约束解决由于遮挡和一致性纹理造成的局部匹配失败问题，全局算法的核心是正确定义场景模型。匹配问题通常可以被描述为能量最小化问题，能量函数一般满足形式  $E = E_{data} + E_{smooth}$ ，其中数据项  $E_{data}$  表示视差函数与输入图像的匹配程度，平滑项  $E_{smooth}$  体现了模型定义的场景约束，数据项和平滑项之间的相互平衡是保证获得准确匹配结果的关键。一旦定义了能量函数，就可以有多种算法来找出极值。如动态规划（DP, Dynamic Programming）<sup>[22]-[23]</sup>、图割（GC, Graph Cuts）<sup>[25]-[29]</sup>、置信传播算法（BP, Belief Propagation）<sup>[30]-[34]</sup>、半全局算法（Semi-global）<sup>[35][36]</sup>、模拟退火（SA, Simulated Annealing）<sup>[37]-[39]</sup>、协同优化（Cooperative Algorithm）<sup>[40]-[42]</sup>等优化算法。其中动态规划算法、图割法和置信传播算法是目前常用的方法。

#### - 动态规划

动态规划是一种用来降低全局优化问题复杂度的数学方法，它将整体的优化问题分解为分阶段决策进行。整体的代价函数分阶段进行计算，各阶段之间的转换通过一系列约束来进行。对于立体匹配而言，外极线单调顺序约束允许全局的匹配代价通过在视差空间图像上寻找最小代价路径来实现。最优路径的代价是所有子路径代价之和，这些子路径所经过点的匹配代价可以由区域相关独立算子来决定。

动态规划与其他优化方法相比较的优点在于它为那些缺乏纹理容易匹配错误的区域提供了全局支持约束，而这些区域由于不同视差下的局部能量值都很低而难以匹配，对于遮挡问题，动态规划中一般将遮挡部分的能量用一个固定值来代替，然后利用像素间的一致性约束来检测遮挡。动态规划方法的缺点是错误匹配可能沿核线方向扩展而导致其他正

确匹配的失败，因此利用动态规划方法得到的视差图经常有条纹出现。

#### - 图割 (Graph Cut)

近年来，基于图方法的优化算法被引入到图像重建、模式聚类等关键问题的研究中，2001年，康奈尔大学的 Boykov、Veksler 和 Zabih<sup>[29]</sup>发表在 PAMI 上的文章首先将图割优化算法引入到立体匹配中，此后将密集视差匹配转化为基于图割的能量函数最小化问题的研究倍受关注，各种基于图割的立体匹配方法都取得了不错的结果。Kolmogorov<sup>[28]</sup>指出能够利用图割来进行优化的能量函数必须具备正则性。因此目前此类方法的研究重点在如何将能量函数转换成适合用图割算法来求解的形式和构建更好更快捷的图结构上。

#### - 置信传播 (Belief Propagation)

图割算法更适合求解二进制变量的能量函数的最优解，而视差变量并非二进制的，另一方面，图割算法对函数的正则性要求也限制了其在立体匹配中的应用，使得人们在设计能量函数时必须考虑函数是否满足正则性，或者需要通过一定的约束条件和转换过程使得函数满足正则性。1982年由 Pearl<sup>[43]</sup>提出的置信传播算法尽管不能保证带有环的图结构收敛到最优值，但其在机器视觉和计算机视觉的应用说明了该方法的有效性。2003年，微软亚洲研究中心的 Sun<sup>[31]</sup>等在 PAMI 上发表的文章首先将置信传播算法引入到立体匹配中，其后很多学者对此进行了研究，目前此类方法的研究重点在如何提高算法的效率上。

经过三十多年的研究，立体匹配技术已经有了很大的发展，但由于立体匹配涉及到的问题太多，至今仍然没有一种方法可以完美的解决立体匹配问题，特别是在复杂场景中，如何提高算法的去歧义匹配和抗干扰能力，有效的解决遮挡问题，降低实现的复杂程度和计算量，都需要进行更为深入的探索和研究。

### 1.3 论文的主要工作

立体视觉是计算机视觉的核心研究领域，经过几十年努力，人们在立体视觉方面取得了较大的进展，但是仍然不能在速度与性能方面同时满足新兴应用领域的要求。针对以上问题，本论文主要研究了面向立体视觉的匹配算法，在图像校正和立体匹配方面做了一些探索。研究的框架如图 1-2:

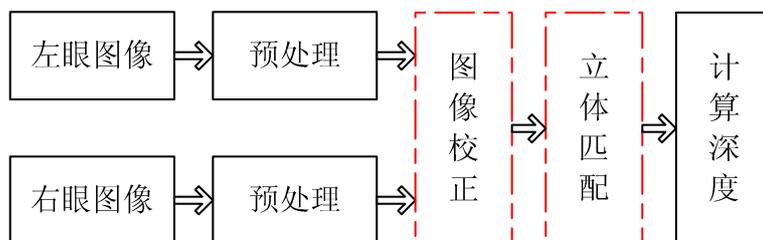


图 1-2 双目立体视觉框架图

论文主要有以下创新点:

- 1) 提出了一种由粗到精的特征点匹配算法。

在图像校正中，正确匹配的特征点对数和精度直接影响着基础矩阵的解，进而影响校正精度。为了使基础矩阵的解更加鲁棒，本论文提出了一种由粗到精的特征点匹配算法，该算法与 Lowe 的方法以及 KD-Tree+RANSAC 算法在三个标准测试集上进行了比较，实验结果表明本文算法对于不同视角、不同光照、不同尺度下的宽基线图像的匹配精度提高约 10% 左右。

2) 提出了一种自适应的置信传播算法。

针对置信传播算法在立体匹配中存在的一些问题，本文提出了一种自适应的置信传播算法。自适应置信传播算法对立体视觉问题有两个贡献，其一是该算法的计算时间比置信传播算法的计算时间减少了大约 50%；其二是该算法的计算精度比置信传播算法更高，这一点已经在 Middlebury 的数据集进行了验证。

3) 提出了一种基于视差校正和视差优化的立体匹配算法。

该算法可以比较准确的计算出遮挡图的位置，进而能更方便的处理图像中的遮挡像素，比如，可以在用全局最优化方法求解视差图的时候有针对性的加上遮挡约束；另外，本文提出了一个渐进的区域融合算法，该算法以大区域和置信系数高的区域为种子区域，然后逐步与其相邻的小区域进行比较与合并，最后形成一个大区域。在进行比较的过程中，本算法没有采用比较匹配代价的方法，而是用平面之间夹角和距离来衡量的，降低了算法复杂度。

## 1.4 论文的内容安排

全文共分六章。

第一章是绪论。介绍了计算机视觉中立体匹配问题的研究现状和本论文的主要工作。

第二章介绍了立体视觉的基础理论与方法，包括摄像机标定理论与方法和立体匹配理论与方法。详细介绍了现有的一些立体匹配算法，包括基于窗口的局部方法以及一些全局优化方法，比如动态规划、图割、置信传播算法等。

第三章是图像校正部分。提出了一种由粗到精的特征点匹配算法，该算法不仅能正确找到特征点对之间的对应关系，而且能获得比较多的匹配对，为计算两摄像机的相对位置关系奠定了良好的基础。

第四章提出了自适应的置信传播算法。自适应置信传播算法是建立在全局能量最小化框架下求最优解，在求解过程中，本文提出了一种低纹理区域的检测方法，同时还给出了判断一个点是否收敛的定义，结合这两个想法，置信传播算法就可以对低纹理区域中的点和不收敛的点进行尽可能多的光滑约束，而对于收敛的点，本文只需根据数据约束就能很准确的找到其正确解。实验证明，该算法可以有效缓减置信传播算法在立体匹配过程中出现的一些问题，比如：计算冗余、不同迭代次数下的实验效果不相同、精度不够高、解不

收敛等问题。

第五章提出了一种基于视差校正和视差优化的立体匹配算法。该算法将整个匹配过程分为三个阶段：初始匹配、视差校正和视差优化。在视差校正阶段，本文提出了一种区分遮挡像素和误匹配像素的分类方法。在视差优化阶段，本文提出了一种渐进的区域融合算法，该算法以大区域和置信系数高的区域为种子区域，然后一步一步与其相邻的小区域进行比较与合并，通过不断优化视差平面，得到了比较精确的视差图。

第六章是总结与展望。

## 第2章 立体视觉的基础理论与方法

客观世界的三维场景投影到相机的二维平面是一个损失深度信息的透视投影过程，因此，深度获取技术作为计算机视觉的关键研究问题之一受到了众多学者的关注，从二维场景中恢复出三维几何结构，也成为计算机视觉中发展最迅速的领域。立体视觉是最关键的技术。

立体视觉的开创性工作是从 20 世纪 50 年代中期开始的，美国麻省理工学院的 Robert 把二维图像分析推广到三维景物分析，标志着计算机立体视觉技术的诞生。在随后的二十几年中，该技术迅速发展，成为一门新的学科。特别是 20 世纪 70 年代末，Marr 和 Poggio<sup>[3]</sup> 等创立的视觉计算理论对立体视觉的发展产生了巨大的影响，现已形成了图像获取到最终的景物可视表面重建的比较完整的体系。1982 年，Barnard 和 Fischler<sup>[4]</sup> 将立体视觉分为图像获取、相机建模、特征提取、图像匹配、深度恢复和深度插值六个步骤。这过程可粗略分为三个阶段：第一个阶段是相机建模，这个阶段是为第二个阶段做准备的，该过程可以将第二个阶段匹配过程中的搜索范围从二维降到一维，大大减少了算法的计算复杂度；第二个阶段是在两幅图像之间建立匹配点；第三个阶段是根据相机参数和对应的匹配点重构物体的三维信息，如果已知两部相机之间的几何关系和点与点的对应关系，第三阶段的任务实质上是一个三角测量的过程，相对比较简单。而重构景物点三维位置信息的准确度依赖于匹配点的准确度，因此解决整个立体视觉问题的关键在第一和第二个阶段，即两幅图像的校正和匹配。下面给出图像校正的相关理论（摄像机标定理论）和方法以及立体匹配的相关理论与方法。

### 2.1 摄像机标定理论与方法

借助射影几何、齐次坐标以及矩阵等数学工具，我们可以描述三维空间到二维图像的成像原理、两幅图像之间的极几何关系。本节主要介绍摄像机模型、线性摄像机成像模型、同一三维场景在两个不同视点处得到的两幅二维图像之间的几何关系：极几何以及极几何的代数表示基础矩阵。两幅图像可以是由两个摄像机在不同位置同时采集的，也可以是同一摄像机顺序采集的，例如摄像机相对场景移动。对于这两种情况，几何上认为是相等的，最后介绍估计基本矩阵的一些主要算法。

#### 2.1.1 摄像机模型

计算机视觉系统从摄像机获取的图像出发，计算三维环境物体的位置、形状等几何信

息，并由此识别环境中的物体。图像上每一点的亮度反映了空间物体表面上某一点放射光的强度，而该点在图像上的位置则与空间物体表面响应点的几何位置有关，这些位置的相互关系由摄像机成像几何模型决定，该几何模型的参数称为摄像机参数，这些参数必须由实验和计算来确定，实验和计算的过程称为摄像机标定。摄像机模型是光学成像几何关系的简化，最简单的模型为线性模型，或称为针孔模型。

成像变换涉及到不同坐标系之间的变换。为了定量的描述摄像机成像过程，首先定义以下四个坐标系：图像像素坐标系、图像物理坐标系、摄像机坐标系与世界坐标系。

### (一) 图像像素坐标系

图像像素坐标系是计算机内部数字图像所用的坐标系。摄像机采集的图像是以标准电视信号的形式输入到计算机，经过计算机中的专用数模转换板变换成数字图像，每幅数字图像在计算机内为  $M \times N$  数组， $M$  行  $N$  列的图像中的每一个元素的数值即为图像点的亮度。如图 2-1，图像坐标系是在图像上定义直角坐标系  $uv$ ，坐标系原点一般位于图像的左上角，每一个像素的坐标  $(u, v)$  分别是该像素在数组中的列数和行数，所以  $(u, v)$  是以像素为单位的图像坐标系的坐标。

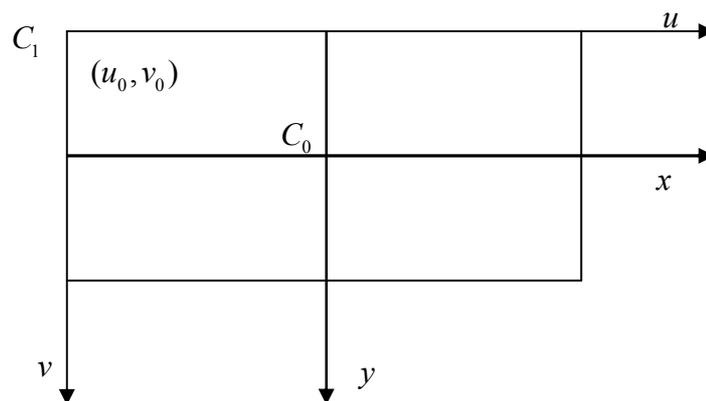


图 2-1 图像像素坐标系与图像物理坐标系

### (二) 图像物理坐标系

图像物理坐标系是在摄像机内形成的平面坐标系，由于图像坐标系中的坐标  $(u, v)$  只表示像素位于数组中的列数和行数，并没有用物理单位表示出该像素在图像中的位置，因此需要在建立以物理单位表示的图像物理坐标系  $xy$ ，该坐标系以图像内某一点  $C_0$  为原点， $x$  轴和  $y$  轴分别与  $u$ ， $v$  轴平行，如图所示，在  $xy$  坐标系中，原点  $C_0$  定义为摄像机光轴与图像平面的交点，该点一般位于图像中心处，图像坐标系与像平面坐标系的关系如图，若  $C_1$  在  $uv$  坐标系下的坐标是  $(u_0, v_0)$ ，每个像素在  $x$  轴和  $y$  轴方向上的物理尺寸为  $dx$  和  $dy$ ，则图

像中任意一个像素在两个坐标系下的坐标有如下关系：

$$\begin{aligned} u &= \frac{x}{dx} + u_0 \\ v &= \frac{y}{dy} + v_0 \end{aligned} \quad (2.1)$$

采用齐次坐标和矩阵形式，上式表示为：

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} 1/dx & 0 & u_0 \\ 0 & 1/dy & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (2.2)$$

### (三) 摄像机坐标系

摄像机坐标系是以摄像机为中心制定的坐标系，摄像机成像几何如图 2-2 表示，其中  $O$  点称为摄像机光心，也为坐标系原点， $X_c$  轴和  $Y_c$  轴与图像平面的  $x$  轴和  $y$  轴分别平行， $Z_c$  轴为摄像机的光轴，它与图像平面垂直。光轴与图像平面的焦点即为像平面坐标系的原点  $O_1$ ，由点  $O$  与  $X_c$ 、 $Y_c$ 、 $Z_c$  轴组成的直角坐标系称为摄像机坐标系， $OO_1$  为摄像机焦距  $f$ 。

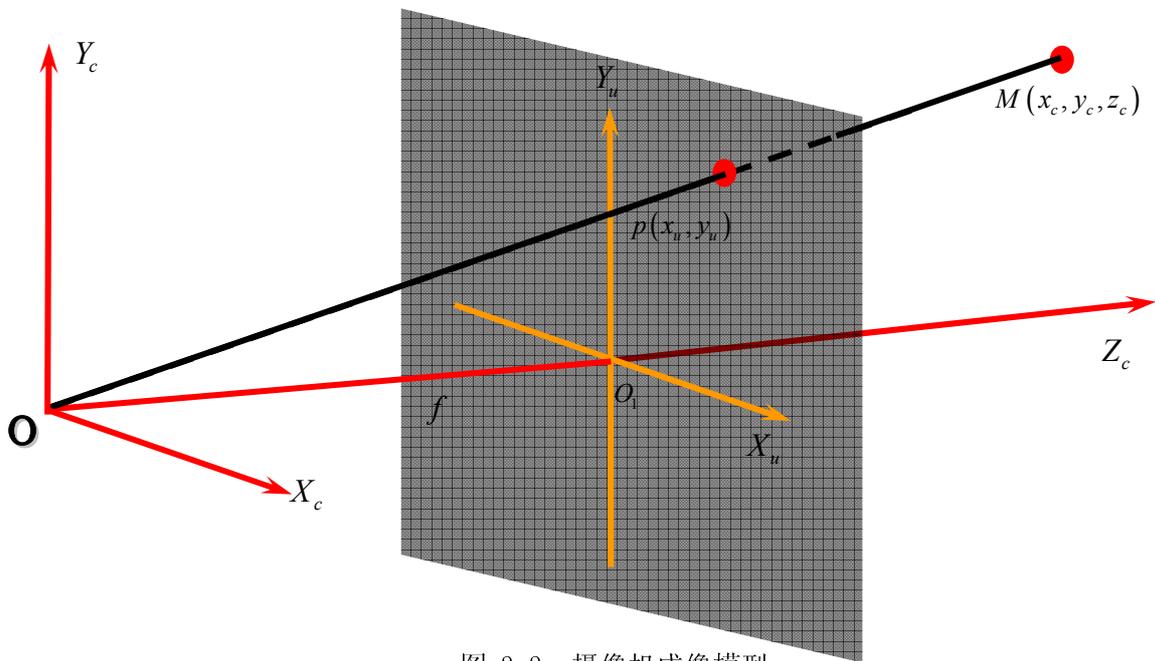


图 2-2 摄像机成像模型

### (四) 世界坐标系

世界坐标系是客观三维世界的绝对坐标系，也称客观坐标系，一般的三维场景都是用这个坐标系来表示的。由于摄像机可以安放到环境中的任何位置，所以我们在环境中还需要选择一个基准坐标系来描述摄像机的位置，并用它来描述环境中任何物体的位置，该坐

标系称为世界坐标系，由  $X_w, Y_w, Z_w$  三个轴组成，摄像机坐标系与世界坐标系之间的关系可以用旋转矩阵  $R$  和平移向量  $t$  来藐视，设三维空间中任意一点  $M$  在世界坐标系与摄像机坐标系下的齐次坐标分别为  $(X_w, Y_w, Z_w, 1)^T$  和  $(X_c, Y_c, Z_c, 1)^T$ ，则存在如下关系：

$$\begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} = \begin{bmatrix} R & t \\ 0_{1 \times 3}^T & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \quad (2.3)$$

其中  $R$  为  $3 \times 3$  单位正交矩阵， $t$  为三维平移向量。

### 2.1.2 线性摄像机成像模型

针孔模型是适合于许多计算机视觉应用的最简单的模型，该模型是 15 世纪初由 Brunelleschi 首先提出的，这个模型尽管简单，但是它对于成像过程的近似程度是可以接受的。

三维空间任意一点  $M$  在图像上的成像位置可以用针孔模型近似表示，即任意点  $M$  在图像上的投影位置  $m$  为光心  $C$  与  $M$  的连线  $CM$  与图像平面的交点，这种关系也称为中心射影或透视投影。由比例关系有如下关系式：

$$\begin{aligned} x_u &= f \frac{x_c}{z_c} \\ y_u &= f \frac{y_c}{z_c} \end{aligned} \quad (2.4)$$

其中， $(x_u, y_u)$  为  $m$  点的像平面坐标。 $(x_c, y_c, z_c)$  为空间点  $M$  在摄像机坐标系下的坐标，用齐次坐标与矩阵表示上述透视投影关系：

$$z_c \begin{bmatrix} x_u \\ y_u \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} \quad (2.5)$$

从世界坐标系（三维场景）到数字图像（图像坐标系）的成像变换可以通过以下 3 个步骤来实现，如图 2-3，本文不考虑摄像机的畸变。

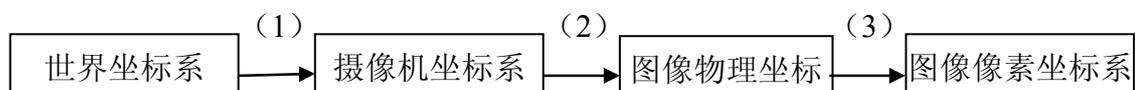


图 2-3 成像过程

(1) 从世界坐标系到摄像机坐标系的变换可以表示为:

$$\begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} = \begin{bmatrix} R & t \\ \mathbf{0}_{1 \times 3}^T & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \quad (2.6)$$

(2) 从摄像机坐标系到图像物理坐标系的变换可以表示为:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \frac{1}{z_c} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} \quad (2.7)$$

(3) 从图像物理坐标系到图像像素坐标系的变换可以表示为:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_u & -f_u \cot \theta & u_0 \\ 0 & f_v / \sin \theta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (2.8)$$

将上述三个步骤结合起来, 我们就可以得到世界坐标系表示的点与其投影点之间的坐标变换关系:

$$z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_u & -f_u \cot \theta & u_0 \\ 0 & f_v / \sin \theta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & t \\ \mathbf{o}_3^T & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \quad (2.9)$$

令  $P_1 = \begin{bmatrix} f_u & -f_u \cot \theta & u_0 \\ 0 & f_v / \sin \theta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$ ,  $P_2 = \begin{bmatrix} R & t \\ \mathbf{o}_3^T & 1 \end{bmatrix}$ , 则上式转化为:

$$z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = P_1 P_2 \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} = P \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \quad (2.10)$$

其中  $P_1$  只与摄像机内部结构有关, 因此称为摄像机内部参数,  $P_2$  完全由摄像机相对于世界坐标系的位置决定, 称为摄像机外部参数, 确定某一摄像机的内外参数的过程称为摄像机标定。

### 2.1.3 极线几何与基础矩阵

极线几何是同一场景的两幅投影图之间的内在的投影几何关系，它只与摄像机内外部参数有关，基础矩阵是极线几何的代数表示，它是一个秩为 2 的  $3 \times 3$  矩阵，基础矩阵可以利用图像的对应点计算出来。

#### 2.1.3.1 极线几何

在一个立体视觉系统中，有两个摄像机，如图 2-4，设  $C$  和  $C'$  分别是两个摄像机的光心，两个摄像机获得的图像分别记为  $I, I'$ ， $M$  为三维空间中的任意一点， $m, m'$  是点  $M$  在两幅图像上的投影点，我们称  $m, m'$  是一对对应点。链接光心  $C$  和  $C'$  的直线称为基线，空间点  $M$  和两个光心  $C$  和  $C'$  共面，设它们所在的平面为  $\pi$ ，该面称为极平面。极平面与图像平面的交线  $l$  和  $l'$  称为极线。因为  $m$  在平面  $\pi$  和像平面  $I$  上，因此， $I$  必然过  $m$  点，也就是说  $m'$  的对应点  $m$  在直线  $l$  上。于是，在图像  $I'$  上搜索点  $m$  的对应点时，不必在整幅图像上寻找，我们只需在其对应的极线上搜索就可以了，这样的话，对应点的搜索空间从二维降到了一维<sup>[48]</sup>。

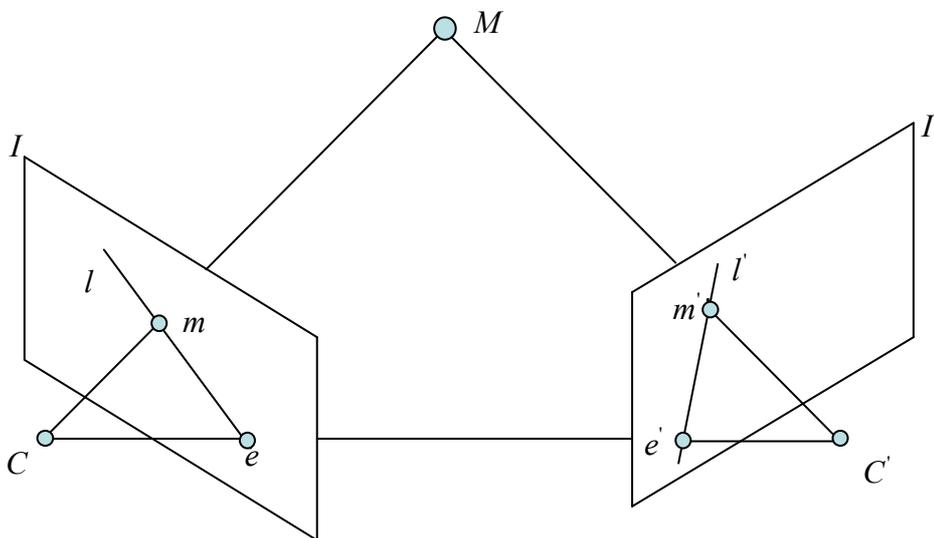


图 2-4 对极几何

#### 2.1.3.2 标准的双目立体几何

在双目视觉中，当采用理想的平行光模型，也就是说两个相机视线平行，基线与水平轴重合且两相机的光轴平行，这种情况称之为标准的双目立体几何。当两相机的焦距相等，各内部参数也相等，这样做匹配搜索时，只需要在另一幅图像的水平扫描线进行搜索，从

而大大减少了匹配搜索的难度，提高了立体匹配的速度。但实际上，两个相机的成像平面通常不在同一个平面上，且它们的内外参数也不完全相同，所以相应的基线也不平行。为了减少匹配搜索的难度，可以通过极线校正过程使两幅图像的极线相互平行。

目前，很多公司都基于标准的双目立体视觉系统开发了自己的产品，如图 2-5，加拿大 point grey 公司生产的 Bumblebee2 立体视觉系统；英国生产的 Hydra stereo webcam 等都是为了将其能合理的应用在稠密匹配、跟踪、人机交互等领域中<sup>[44][45][46]</sup>。下图给出了现有的一些产品图：



图 2-5 双目立体视觉设备<sup>[44][45][46]</sup>

### 2.1.3.3 基础矩阵

设两个摄像机的投影矩阵分别为 $[P_1 \ p_1]$ 和 $[P'_1 \ p'_1]$ ，则两个摄像机的投影方程如下：

$$\begin{aligned} Z_e m &= P_1 M + p_1 \\ Z_e m' &= P'_1 M + p'_1 \end{aligned} \quad (2.11)$$

其中 $M = (X_w, Y_w, Z_w)$ ， $(M \ 1)^T$ 为三维空间点在世界坐标系下的齐次坐标， $m, m'$ 分别是投影点在图像坐标系下的坐标。

将上式消去 $M$ 得：

$$Z_e m' - Z_e P'_1 P_1^{-1} m = p'_1 - P'_1 P_1^{-1} p_1 \quad (2.12)$$

若将上式的右端的向量记为 $P$ ，则上式变为： $p = p'_1 - P'_1 P_1^{-1} p_1$

定义 1：如果 $t$ 为三维向量， $t = (t_x, t_y, t_z)^T$ ，称下列矩阵为由 $t$ 定义的反对称矩阵，记为 $[t]_x$ ：

$$t_x = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix}$$

设由 $p$ 定义的反对称矩阵记为 $[p]_x$ ，由 $[p]_x p = 0$ 可知

$$[p]_x(Z_e m' - Z_e P_1' P_1^{-1} m) = 0 \quad (2.13)$$

将上式两边除以  $Z_e$ ，并记  $Z = Z_e / Z_e'$ ，于是：

$$[p]_x Z P_1' P_1^{-1} m = [p]_x m' \quad (2.14)$$

将上式变形得：

$$m'^T [p]_x P_1' P_1^{-1} m = 0 \quad (2.15)$$

上式给出了  $m$  和  $m'$  之间必须满足的关系，可以看出，在给定  $m$  的情况下，上式是一个关于  $m'$  的线性方程，即图像  $I'$  上的极线方程。

令  $F = [p]_x P_1' P_1^{-1}$ ，则  $m'^T F m = 0$ 。矩阵  $F$  称为基础矩阵。等式  $m'^T F m = 0$  的重要性在于它给出了一个描述基本矩阵的方法，不需要摄像机的投影矩阵，而仅仅根据两幅图像间的对应点，这就使得我们可以利用对应点的关系恢复出基础矩阵  $F$ ，这一过程被称为立体摄像机的弱标定。基础矩阵具有以下性质：

- (1) 基础矩阵  $F$  是一个秩为 2 的  $3 \times 3$  矩阵，其自由度为 7；
- (2) 基础矩阵  $F$  在相差一个非零常数因子下是唯一的；
- (3) 基础矩阵  $F$  可以由摄像机的投影矩阵求出，或者由图像间的对应点对求出；
- (4) 给定图像  $I$  中的一点  $m$ ，其相应的图像  $I'$  中的极线  $l'$  可以表示为  $l' = F m$ 。类似地，

$l = F^T m'$  表示图像  $I$  中的点  $m'$  的极线；

- (5) 图像  $I$  上的所有极线交于极点  $e$ ，因此， $F e = 0$ ，同样  $F^T e' = 0$ 。

在双目立体视觉系统中，两个摄像机的位置是固定的，我们以第一个摄像机坐标系建立世界坐标系，则两个摄像机投影矩阵分别为：

$$P_1 = [K_1 | 0] \quad P_2 = [K_2 R | K_2 t] \quad (2.16)$$

其中  $K_1, K_2$  分别为左右摄像机的内部参数， $R$  和  $t$  为第二个摄像机坐标系相对于世界坐标系的旋转矩阵和平移向量。

本质矩阵是一种特殊的基础矩阵，本质矩阵与基础矩阵具有如下关系：

$$E = K^T F K \quad (2.17)$$

且基础矩阵与摄像机内外参数具有下面的关系：

$$F = K_2^{-T} [t]_x R K_1^{-1} \quad (2.18)$$

于是我们可以推出  $EE^T = [t]_x [t]_x^T$ ，从而可以得到平移向量  $t$ ，又由本质矩阵  $E = [t]_x R$ ，求出旋转矩阵  $R$ 。

#### 2.1.3.4 基础矩阵的计算

由于基础矩阵在计算机视觉领域中的重要地位，精确地估计基础矩阵已经成为人们研究的一个重要方向。所有的估计基础矩阵的方法都是基于求解一个齐次方程组，假设给定一个对应点的集合  $\{(m_i, m'_i) | i=1, 2, \dots, n\}$ ，其中  $m_i = (u_i, v_i, 1)^T$ ， $m'_i = (u'_i, v'_i, 1)^T$ 。当使用  $n$  对对应点时，方程可以写成如下形式：

$$U_n F = 0 \quad (2.19)$$

其中：

$$F = (F_{11}, F_{21}, F_{31}, F_{12}, F_{22}, F_{32}, F_{13}, F_{23}, F_{33})^T$$

$$U_n = \begin{bmatrix} u_1 u'_1 & u_1 v'_1 & u_1 & v_1 u'_1 & v_1 v'_1 & v_1 & u'_1 & v'_1 & 1 \\ u_2 u'_2 & u_2 v'_2 & u_2 & v_2 u'_2 & v_2 v'_2 & v_2 & u'_2 & v'_2 & 1 \\ \vdots & \vdots \\ u_n u'_n & u_n v'_n & u_n & v_n u'_n & v_n v'_n & v_n & u'_n & v'_n & 1 \end{bmatrix}$$

上述方程是一个齐次方程组， $F$  在相差一个比例因子的情况下，有 8 个独立变量，但是因为  $F$  还满足  $\det(F) = 0$ ，因此，在  $F$  的 9 个未知变量中只有 7 个独立变量。

计算基础矩阵一般有三种方法：线性方法、迭代方法和鲁棒方法。当匹配点对的个数大于或等于 8 时，就可以用最小二乘法求解基础矩阵  $F$ ，于是问题转化为求解目标函数：

$$\begin{aligned} \min & \|U_n f\|^2 \\ \text{s.t.} & \|f\| = 1 \end{aligned} \quad (2.20)$$

在本论文中，我们采用改进 8 点法求解上述目标函数。改进的 8 点算法是建立在 8 点算法的基础上的，Hartley<sup>[47]</sup>分析了 8 点算法不稳定的原因，认为主要是直接使用原始图像像素坐标时，系数矩阵的条件数不好造成的，而引起方程组的条件数恶化的原因是点的齐次坐标表示式的各个分量的数量级相差太大，基于上面的认识，Hartley 提出了改进 8 点算法，在应用 8 点算法估计基础矩阵之前，先对数据  $\{(m_i, m'_i) | i=1, 2, \dots, n\}$  进行一个规范化处

理，即对原始数据做各向同性的变换，就可以减少噪声的干扰，大大提高 8 点算法的精度。规范化变换如下：

- 1) 对图像点做位移变换，使图像的原点位于图像点集的质心；
- 2) 对图像点做缩放变换，使图像点分布在以质心为圆心，半径为 $\sqrt{2}$ 的圆内。

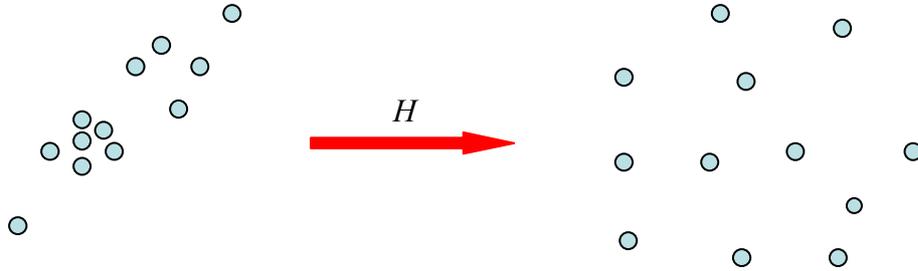


图 2-6 数据的规范化变换

在上面的数据规范化过程图中， $H$  为变换矩阵，写成如下形式：

$$H = \begin{bmatrix} k_m & 0 & -k_m u_0 \\ 0 & k_m & -k_m v_0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$u_0 = \frac{1}{N} \sum_{i=1}^N u_i, \quad v_0 = \frac{1}{N} \sum_{i=1}^N v_i$$

$$k_m = \frac{\sqrt{2}}{\sqrt{\frac{1}{N} \sum_{i=1}^N [(u_i - u_0)^2 + (v_i - v_0)^2]}}$$

下面给出改进的 8 点算法步骤：

---

算法 2.1：改进的 8 点算法

---

- 1、对图像坐标规范化变换：

$$\hat{m}_i = H m_i, \quad \hat{m}'_i = H' m'_i \quad (2.21)$$

其中  $H, H'$  是  $3 \times 3$  的变换矩阵，使得点集的质心在原点，并且点到原点的平均距离为 $\sqrt{2}$ ，得到新的对应点集。

- 2、由新的对应点集构造系数矩阵  $U_n$ ；

- 3、对矩阵  $U_n$  进行奇异值分解  $U_n = U D V^T$ ，由向量  $f = v_0$  构造基础矩阵  $F$ ；

- 4、对矩阵  $F$  进行秩 2 约束，即对  $F$  进行奇异值分解得  $F = U \text{diag}(s_1 \ s_2 \ s_3) V^T$ ，然后令  $s_3 = 0$

得到基础矩阵的估计  $\hat{F} = U \text{diag}(s_1 \ s_2 \ 0) V^T$ ；

5、根据变换  $F = H^T \hat{F} H$ ，得到基础矩阵  $F$ 。

在本文中，我们利用改进 8 点法对双目立体图像对进行校正。

## 2.2 立体匹配理论与方法

随着三维重建技术、虚拟现实等领域的兴起，人们对视差图的精度和密度的要求越来越高，力图对图像中的每一个像素点得到可靠的深度信息，这也就迫切的需要我们提高立体匹配算法的精度和速度。在本节中，我们先给出立体匹配的相关理论，然后介绍立体匹配的一些常用的方法。

### 2.2.1 立体匹配理论

立体匹配的目的就是对同一三维场景的两幅成像平面中的每一个像素点找到与其相匹配的点，于是可以计算出每个坐标点的视差值  $d$ ，进而可以准确的描述场景的三维空间信息。在本节中，我们首先给出了视差与深度之间的几何关系，然后给出了立体匹配中常用的约束条件以及匹配测度。

#### 2.2.1.1 视差与深度的关系

双目视差简称视差。人的双眼从不同的角度去观察三维世界的景物，由于几何学的投影，离观察者不同距离的像点在左右两眼视网膜上的成像就处在不同的位置上，这种两眼视网膜上的位置差就称之为双目视差，它反映了客观事物的深度，立体视觉是模拟人眼深度感知功能的系统，因此，在立体视觉中，同一物体在左右两个成像平面上的位置差就称之为视差， $X$  方向的视差称之为水平视差， $Y$  方向的视差称之为垂直视差，经过极线校正后的图像的垂直视差为 0。两幅图像上所有点的视差组成的图称为视差图，在不同的坐标系下，视差和深度之间的关系是不一样的，为了更好的说明视差和深度的关系，图 2-7 给出标准双目结构下两者之间的几何关系：

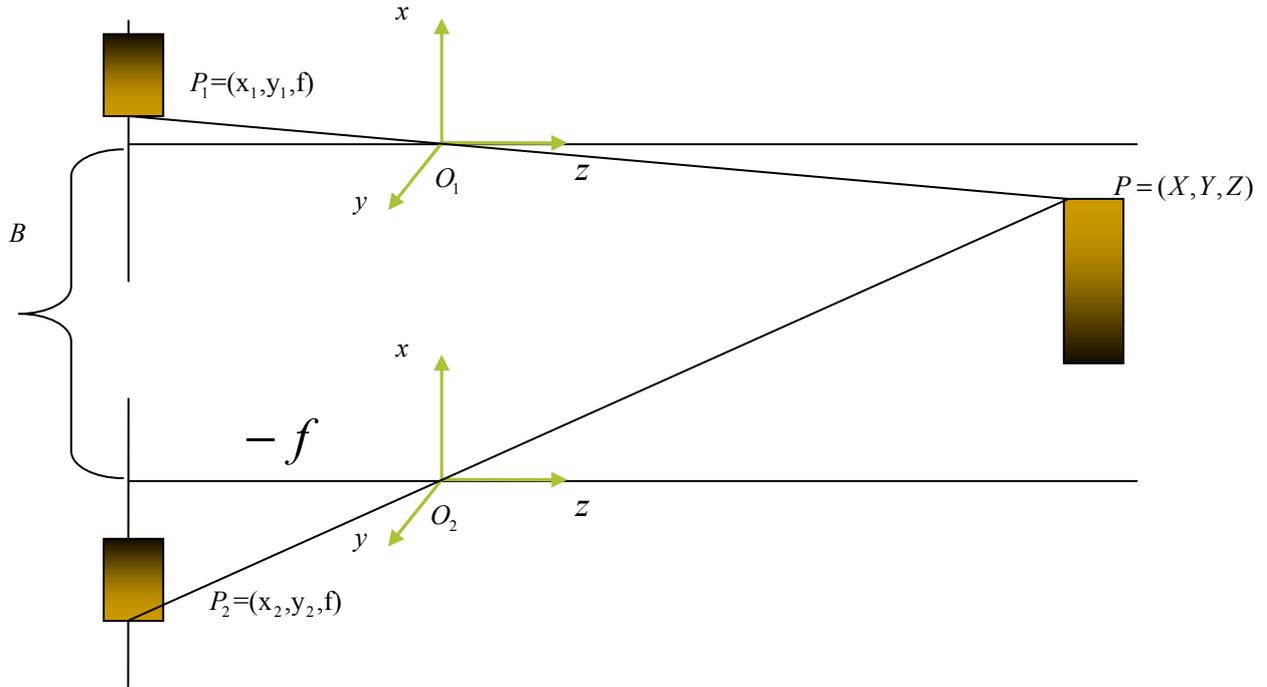


图 2-7 标准双目结构下视差与深度之间的关系图

在图 2-7 中,  $O_1$  和  $O_2$  分别表示两个水平放置的摄像机的光心,  $P = (X, Y, Z)$  是三维世界坐标系中的一点,  $P_1 = (x_1, y_1, f)$  和  $P_2 = (x_2, y_2, f)$  分别表示点  $P$  的成像点。根据相似三角形的性质, 我们得到下面的等式:

$$\begin{aligned} x_1 &= -f \frac{X}{Z} \\ x_2 &= -f \frac{X+B}{Z} \end{aligned} \quad (2.22)$$

于是:

$$Z = \frac{fB}{x_1 - x_2} = \frac{fB}{d} \quad (2.23)$$

由上面的等式可以看出, 在标准的双目立体视觉系统中, 视差  $d$  与深度  $Z$  成反比, 只要我们能知道图像中每一点的视差, 我们就可以计算出该点在三维场景中的深度信息。

### 2.2.1.2 立体匹配的难点

在双目立体视觉系统中, 主要有以下几方面的因素影响立体匹配算法的精度:

- 1) 光照变化: 成像过程中不可避免的存在光线变化、噪声以及一些非线性因素的影响, 因此, 左右图像中对应点的亮度会有明显差异。
- 2) 遮挡: 遮挡是指对于世界坐标系中的一点, 该点在其中一幅图像是可见的, 在另

一幅图像中是不可见的。一般来说，两个相机位置相隔越远，左右图像中存在遮挡像素的可能性就越大。

3) 无纹理区域：在真实场景中，总存在一些无纹理的区域，使得匹配难以进行，对于这些区域，灰度一致性约束是没有用的。通常我们需要通过光滑约束将纹理较为明显的区域的信息传播到纹理不明显的区域。

4) 重复纹理：重复纹理是指实际场景中有很多相似的场景，比如窗户，墙等，对于重复纹理区域中的像素点，将可能出现多个匹配点与其对应。

5) 视差不连续区域：视差不连续区域一般位于区域的边界处，这些区域也是容易发生遮挡的区域。

### 2.2.1.3 立体匹配的标准测试集以及评价标准

Middlebury 大学<sup>[49]</sup>提供了 Tsukuba、Venus、Teddy 和 Cones 立体像对的视差标准图，同时可以上传自己的实验结果到该网站进行测试。图 2-8 给出这几幅图像的左右图像以及左图的标准视差图。

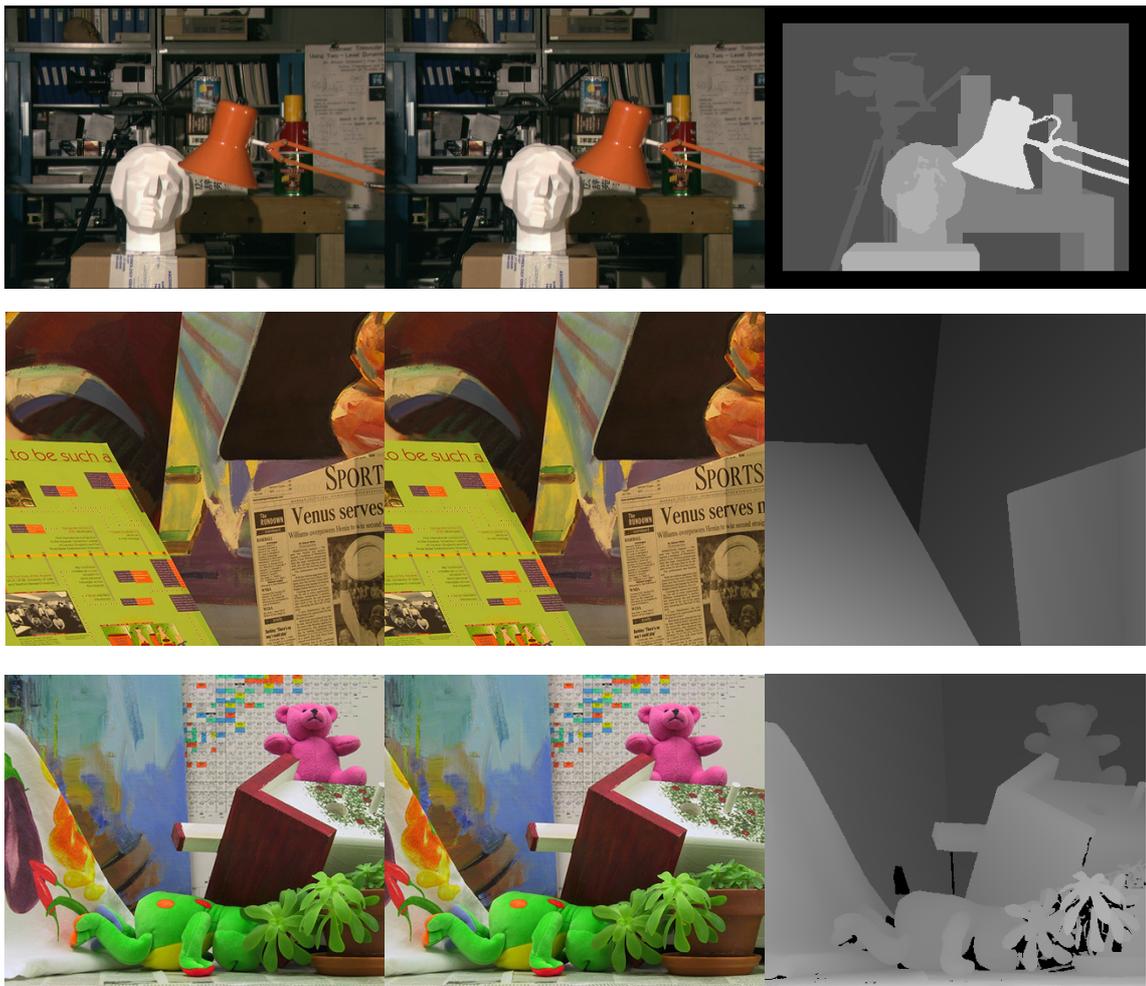




图 2-8 标准左右图像及标准左视差图

为了对各种立体匹配算法进行定量比较，Scharstein<sup>[8]</sup>等又提出了均方根误差和错误匹配百分比两类定量指标。均方根误差  $R$  (Root-Mean-Square) 和错误匹配百分比  $B$  利用下式来计算：

$$R = \left( \frac{1}{N} \sum_{(x,y)} |d_c(x,y) - d_r(x,y)|^2 \right)^{\frac{1}{2}} \quad (2.24)$$

$$B = \frac{1}{N} \sum_{(x,y)} |d_c(x,y) - d_r(x,y)| > \delta_d \quad (2.25)$$

其中  $N$  为图像像素的个数， $d_c(x,y)$  和  $d_r(x,y)$  分别为计算所得视差图和标准视差图， $\delta_d$  是视差误差容差，在实验中  $\delta_d=1$ 。

为了更好的衡量算法在不同区域下的效果，Scharstein 从图像中提取出三类典型区域进行测试：非遮挡区域 nonocc (non-occlusion regions)、所有区域 all (all region) 和视差不连续区域 disc (depth discontinuity region)。平坦区域是指参考图像固定大小的窗口内灰度水平方向的梯度平均值低于某个阈值的区域；遮挡区域是指匹配图像中被遮挡区域；视差不连续区域是指邻域像素视差之差超过一定阈值的相邻像素组成的区域。图 2-9 给出了测试图像 Teddy 的原图以及三类典型区域的图像。

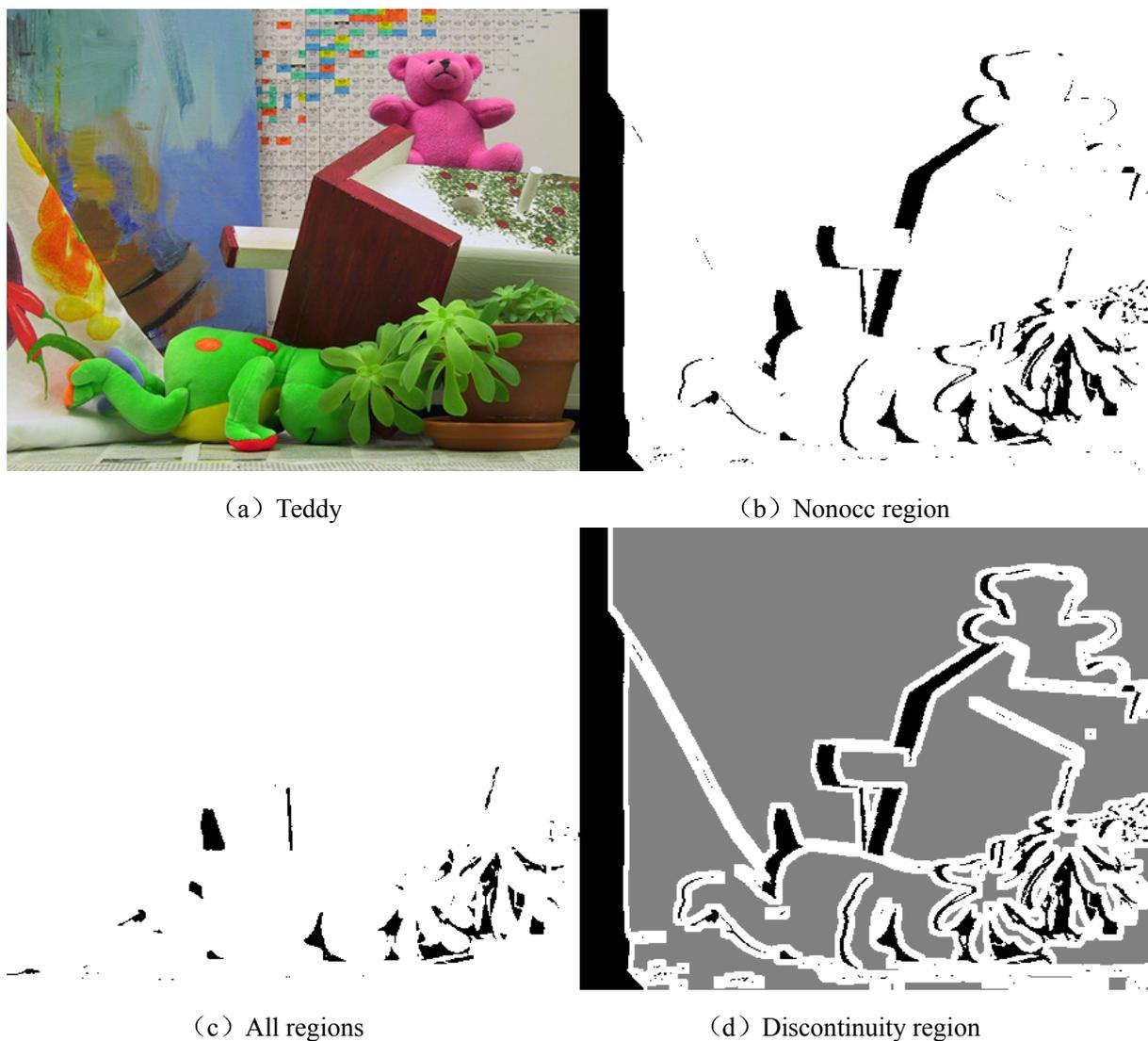


图 2-9 Teddy 的原图以及三类典型区域的图像

为了与其他算法的结果做定性和定量的比较，我们在后续章节中提出的立体匹配算法都是以 Middlebury 网站的标准图像为测试图像进行了测试，并计算了用来比较的各个量化指标，每个指标都是在  $\delta_i=1$  的情况下得到的。

## 2.2.2 立体匹配方法

立体匹配方法可分为两大类：基于窗口的局部方法和基于能量函数的全局方法。局部方法的效率较高，但对噪声敏感，鲁棒性较差；全局方法对图像噪声不敏感，可以得到较为精确的视差图，但其效率较低，无法应用于实时系统。

### 2.2.2.1 匹配测度

匹配过程中经常用到的匹配测度包括归一化相关（Normalized Cross-Correlation）、平方差和（Sum of Squared Differences）、绝对差和（Sum of Absolute Differences）等，除此之外，Birchfield 和 Tomasi<sup>[50]</sup>提出利用线性插值来减小测度对图像采样效应的敏感度，下面

我们分别进行介绍

### 1) 归一化互相关

对于二维数字图像  $I_L(x, y)$  和  $I_R(x, y)$ ，其归一化互相关的定义为：

$$\rho(i, j) = \frac{\sum_{x=1}^m \sum_{y=1}^n (I_L(x+i, y+j) - \hat{I}_L) \cdot (I_R(x+i+d, y+j) - \hat{I}_R)}{\sqrt{\sum_{x=1}^m \sum_{y=1}^n (I_L(x+i, y+j) - \hat{I}_L)^2 \cdot (I_R(x+i+d, y+j) - \hat{I}_R)^2}} \quad (2.26)$$

其中  $d \in [d_{\min}, d_{\max}]$ ， $d_{\max}$  和  $d_{\min}$  分别为可能的最大视差和最小视差， $m$  和  $n$  分别是搜索窗口的长度和宽度， $\hat{I}_L$  和  $\hat{I}_R$  分别为左右图像搜索窗口内的灰度平均值。

$$\hat{I}_L = \frac{1}{mn} \sum_{x=1}^m \sum_{y=1}^n I_L(x+i, y+j) \quad (2.27)$$

$$\hat{I}_R = \frac{1}{mn} \sum_{x=1}^m \sum_{y=1}^n I_R(x+i+d, y+j) \quad (2.28)$$

### 2) 差平方和

差平方和的思想很简单，假如两个窗口内包含相同的图像特征，那么这两个窗口内的对应像素的灰度信息应该是一致的，所以它们之间的差值应该趋近于 0，于是差的平方和也应该趋近于 0。因此可用差平方和的大小来表示两个窗口内图像的相似程度，其计算公式为：

$$E(i, j) = \sum_{x=1}^m \sum_{y=1}^n (I_L(x+i, y+j) - I_R(x+i+d, y+j))^2 \quad (2.29)$$

其中  $d \in [d_{\min}, d_{\max}]$ ， $d_{\max}$  和  $d_{\min}$  分别为可能的最大视差和最小视差， $m$  和  $n$  分别是搜索窗口的长度和宽度。差平方和的思想很简单，但是该计算方法对图像的噪声、光照变化等过于敏感。

### 3) 差绝对值和

差绝对值和与差平方和的思想较为类似，其计算公式为：

$$E(i, j) = \sum_{x=1}^m \sum_{y=1}^n |I_L(x+i, y+j) - I_R(x+i+d, y+j)| \quad (2.30)$$

其中  $d \in [d_{\min}, d_{\max}]$ ,  $d_{\max}$  和  $d_{\min}$  分别为可能的最大视差和最小视差,  $m$  和  $n$  分别是搜索窗口的长度和宽度。该方法与差平方和相比, 一个是 1-范数, 一个是 2-范数, 换句话说, 一个受某一点的影响大一些, 一个小一些, 没有什么其他本质的区别。

#### 4) Birchfield 测度<sup>[50]</sup>

在匹配过程中, 除了两个相机的进光量、增益的不同会影响匹配结果外, 相机的采样个也是影响匹配的因素之一。在图像校正过程中, 我们不可避免的要对图像进行采样, Birchfield 提出采用线性插值来减小图像采样效应的敏感程度, 下面给出相关定义:

$$\bar{e}(x_L, x_R, I_L, I_R) = \min_{x_R - \frac{1}{2} \leq x \leq x_R + \frac{1}{2}} |I_L(x_L) - \hat{I}_R(x)| \quad (2.31)$$

其中,  $x_R = x_L + d$ ,  $\hat{I}_R$  是图像  $I_R$  的插值函数, 同样可以定义:

$$\bar{e}(x_R, x_L, I_R, I_L) = \min_{x_L - \frac{1}{2} \leq x \leq x_L + \frac{1}{2}} |\hat{I}_L(x) - I_R(x_R)| \quad (2.32)$$

由此我们可以得到 Birchfield 匹配测度为:

$$e(x_L, x_R) = \min \{ \bar{e}(x_L, x_R, I_L, I_R), \bar{e}(x_R, x_L, I_R, I_L) \} \quad (2.33)$$

因为一个分段线性单调函数极值点位于它的断点处, 因此我们可以这样计算  $\bar{d}$ , 令:

$$\begin{aligned} I_R^- &= \hat{I}_R(x_R - 0.5) = 0.5(I_R(x_R) + I_R(x_R - 1)) \\ I_R^+ &= \hat{I}_R(x_R + 0.5) = 0.5(I_R(x_R) + I_R(x_R + 1)) \end{aligned} \quad (2.34)$$

$$\begin{aligned} I_{\min} &= \min \{ I_R^-, I_R^+, I_R(x_R) \} \\ I_{\max} &= \max \{ I_R^-, I_R^+, I_R(x_R) \} \end{aligned} \quad (2.35)$$

综合上面的等式, 我们可以得到:

$$\bar{e}(x_L, x_R, I_L, I_R) = \max \{ 0, I_L(x_L) - I_{\max}, I_{\min} - I_L(x_L) \} \quad (2.36)$$

同样我们可以得到  $\bar{e}(x_R, x_L, I_R, I_L)$ , 于是 Birchfield 匹配测度  $e(x_L, x_R)$  和  $e(x_R, x_L)$  就可以计算出来了。

#### 2.2.2.2 局部匹配方法

局部匹配方法的基本思想是根据一定的相似准则, 比较两幅图像中局部图像块的相似

度来确定其是否相同，从而确定是否匹配。基于窗口的局部匹配方法通常是一个二维搜索的过程，设有两幅待匹配的灰度图像，如图 2-10，在左图像上选择目标点  $(x, y)$  并以该点为中心确定大小为  $m \times n$  的模板窗口，为了在右图像上确定目标点的匹配点，我们在右图像上以  $(x, y)$  为中心，在  $(m+c) \times (n+d)$  大小的邻域内寻找一个  $m \times n$  大小的窗口，进一步来确定其匹配点的位置。

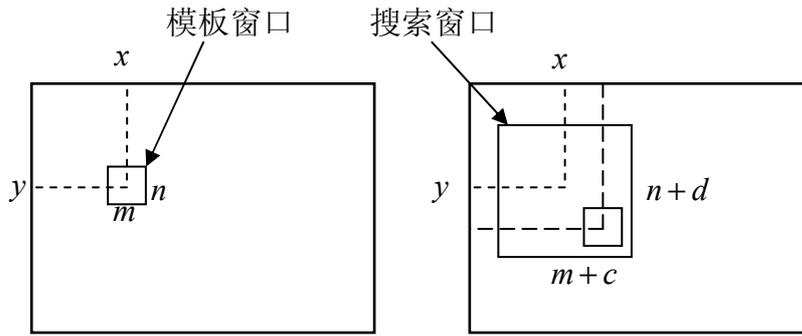


图 2-10 块匹配示意图

基于窗口的局部匹配方法的计算量是非常大的，假设图像有  $N$  个像素，模板有  $n$  个像素，视差最大值为  $D$  个像素，那么区域匹配的计算复杂度为  $O(NDn)$ 。尽管人们提出了各种改进方法，但是基于窗口的匹配方法仍存在着以下不足：首先局部匹配是直接利用图像的像素值进行匹配的，因此对于图像的旋转以及光照和对比度的变化非常敏感；其次，计算窗口之间的相似性测度的计算量大；最后，窗口的大小难以选择，窗口过大，在视差跳跃处会出现误匹配，窗口过小，信噪比小一些。

### 2.2.2.3 全局匹配方法

大部分全局算法都是基于能量函数最小化的思想，目标是找一个视差函数，使得全局能量最小化，动态规划法作为一种比较常用的全局算法，本文简单的介绍一下。

动态规划是用来处理多阶段决策问题的一种方法，是一种用来降低全局优化问题复杂度的数学方法，它通过将整体的优化问题分解为分阶段决策进行，整体的代价函数通过分阶段进行计算，各阶段之间的转换加入一系列约束来进行，对于立体匹配来说，外极线单调顺序约束允许全局的匹配代价通过在视差空间图像上寻找最小代价路径来实现，最优路径的代价是所有子路径的和，这些子路径所经过点的匹配代价可以由点的相似性度量因子来决定。

在 Birchfield 的动态规划算法中，作者定义全局能量函数为：

$$E(M) = N_{occ} k_{occ} - N_m K_r + \sum_{i=1}^{N_m} DSI(x_i, y_i) \quad (2.37)$$

此算法中的全局能量优化是针对每一行而言的，目的就是针对每一行找到匹配序列  $M$  使得  $E(M)$  最小。上式中  $E(M)$  表示当匹配序列为  $M$  时的匹配代价， $N_{occ}$  为序列  $M$  中遮挡点的个数， $k_{occ}$  为遮挡因子。 $N_m$  为序列  $M$  中匹配成功的对应点个数， $K_r$  为匹配因子。 $DSI(x_i, y_i)$  为匹配点的不相似度。本算法在标准数据集下进行了测试，图 2-11 是其生成的视差图，如下：

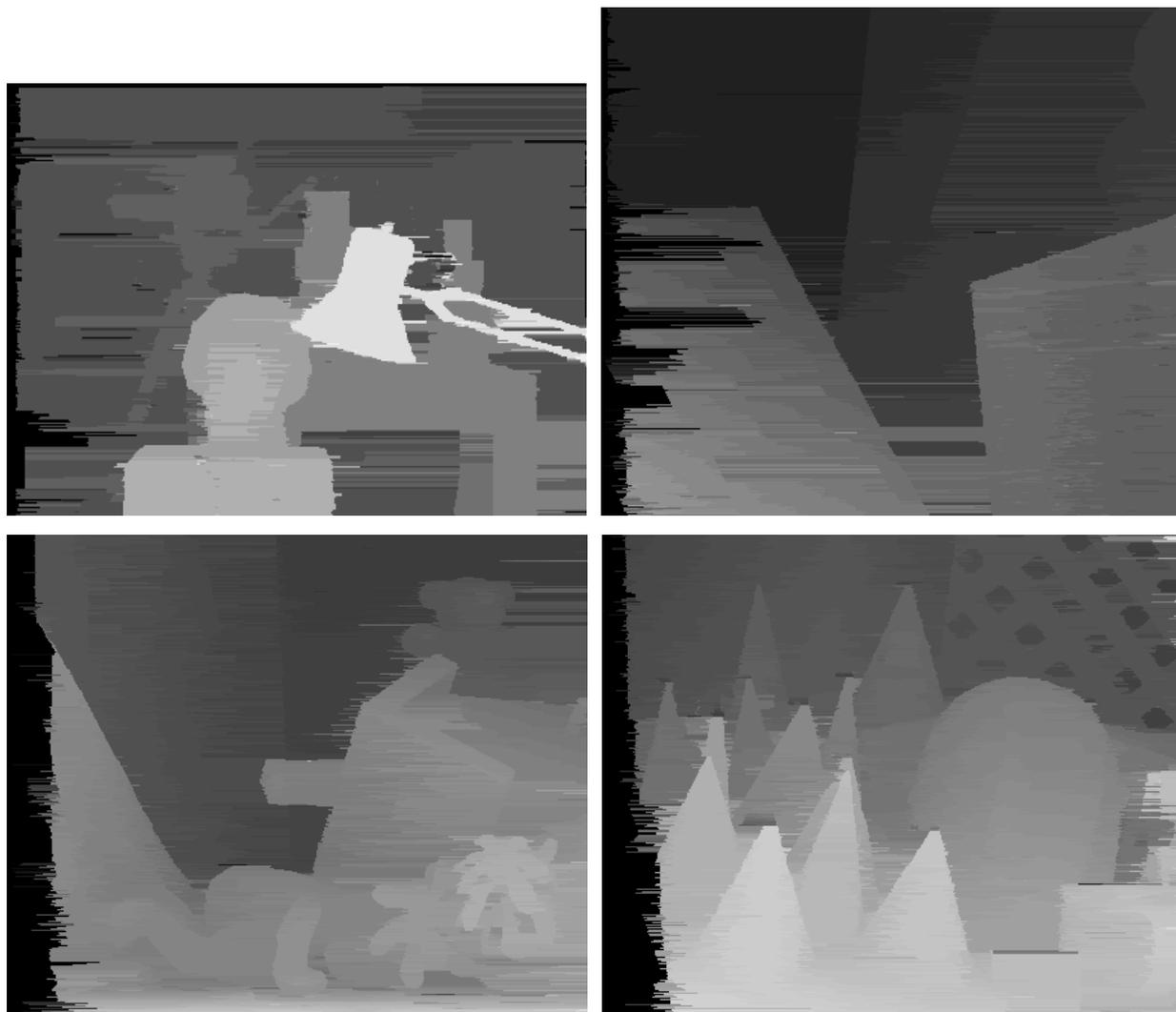


图 2-11 由动态规划算法得到的视差图

实验结果表明，动态规划是对一条扫描线计算全局匹配代价，提高了匹配的准确性，但是动态规划有很大的局限性，它缺乏对水平连续性约束和垂直连续性约束的融合，当单点像素产生误匹配时往往会影响到同一条扫描线上后续像素的匹配，产生明显的条纹效应。

除了常用的动态规划算法外，还有两种比较常用的全局匹配算法：图割法和置信传播算法，这两种算法也是构造一个全局能量函数，通过能量函数取最小实现最优标号。在第四章将详细的给出置信传播算法的具体实现过程。

### 2.2.2.4 基于分割的全局匹配方法

基于分割的立体匹配算法是近年来比较流行的全局匹配方法，与传统的方法相比，该方法可以得到稠密的视差图，因此很多学者<sup>[51]-[57]</sup>对其进行了研究，取得了不错的效果。

下面我们给出基于分割的匹配算法的基本流程。整个匹配过程包括四个主要步骤：

- 1) 利用颜色分割算法对输入图像进行分割，这样可以提取图像中的一致性区域；
- 2) 利用局部或全局匹配算法得到可信度高的初始视差点；
- 3) 利用分割区域和初始视差点得到区域视差的分布；
- 4) 以区域视差分布为变量设计全局能量代价函数得到视差的全局最优解。

图 2-12 以图像 Tsutuba 为例，对基于分割的立体匹配算法每一个步骤所得到的中间图像进行表示如下：

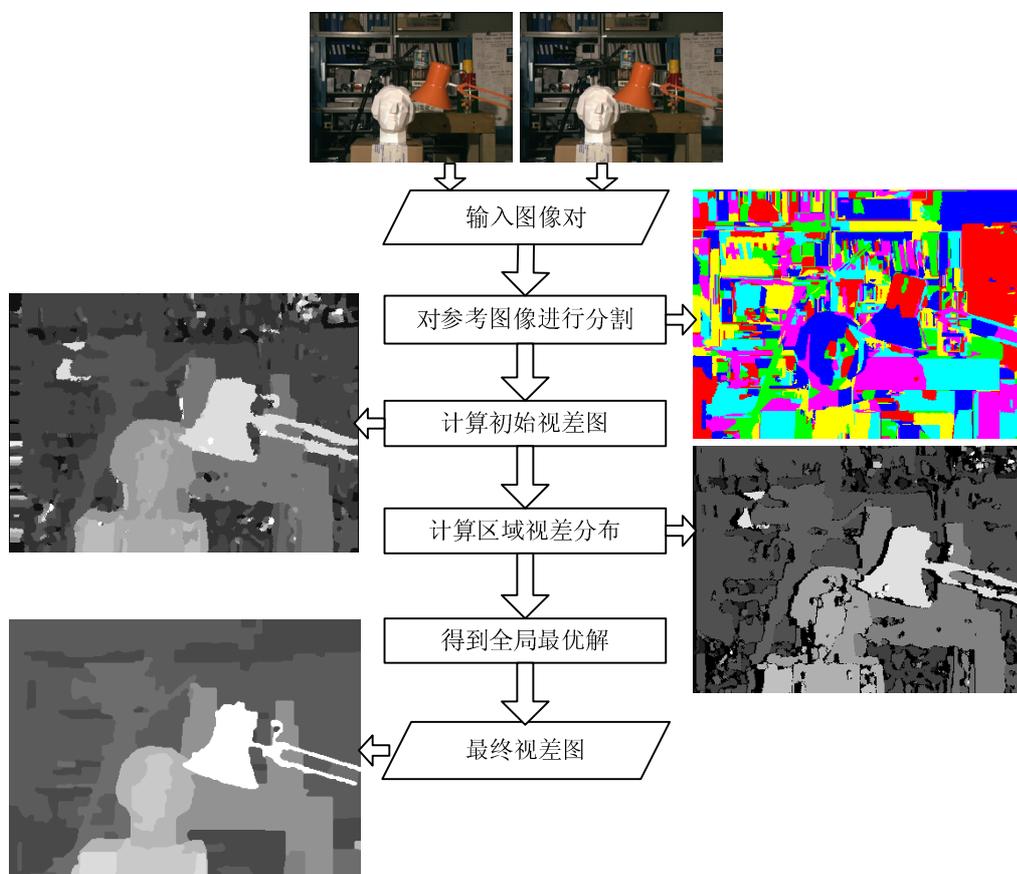


图 2-12 基于分割的立体匹配算法流程图

## 2.3 本章小结

摄像机标定和立体匹配是立体视觉中最关键的两个技术问题。许多学者都提出了相关的算法解决各种情况下的实际问题。

在摄像机标定的理论与方法中，我们介绍了摄像机模型、线性摄像机的成像模型以及

各种坐标系下图像之间的相互转化关系，同时，给出了双目摄像机下图像之间的几何关系：极线几何与基础矩阵，最后给出了几种常用的求基础矩阵的方法。在双目立体视觉系统中，为了降低搜索的维数，需要对拍摄的双目图像进行校正，因此，准确计算基础矩阵是非常必要的。

在立体匹配的相关理论与方法中，给出了在标准双目立体视觉系统下，视差与深度之间的关系、立体视觉问题目前存在的难点、立体匹配的标准测试集以及评价标准。同时我们也给出了几种常用的计算匹配测度的方法以及目前常用的计算立体匹配的方法。目前还没有任何一种方法可以很完美的解决立体匹配问题，这也是在此研究领域中还有很多研究人员在不断努力的原因。就目前的研究成果来看，立体匹配算法的研究主要沿着两个方向前进，一是采用全局约束策略来实现全局优化的立体匹配算法；二是采用局部方法实现局部最优的立体匹配算法。这两种算法各有优缺点。对于全局的立体匹配算法来说，它的主要优点是获取的视差图准确度比较高，但是该问题往往是一个 NP 问题，所以一般只能得到问题的近似最优解且算法的计算复杂度高。对于局部算法而言，它的主要优点是算法的时间复杂度不高，比较容易实现，但是局部算法得到的视差图的准确度不高。



## 第3章 图像校正

在这一章中，我们对未校正的双目图像对进行校正。在校正之前，需要标定摄像机的内外参数。在标定过程中，需要计算基础矩阵，进而得到旋转矩阵和平移向量。为了能精确获取摄像机外参数，必须得到尽可能多且正确的匹配点对。因此，本文提出了一种由粗到精的特征点匹配算法。

在该算法中，我们首先根据特征点之间的相似性约束和空间约束建立了数学模型。由于相似度约束和空间约束是相互独立的，于是我们对该模型分开求解。先用互相相似度算法求解相似度约束建立的模型，得到初始解，然后用改进遗传算法求解空间约束的数学模型，进而得到模型的精确解。在改进遗传算法中，我们提出了整数编码、优化初始种群、多点交叉等策略的改进策略，特别地，初始解将作为改进遗传算法初始种群的一个解。在得到正确的匹配点对后，本文利用改进 8 点法求解基础矩阵，进一步校正双目图像对。

本章是这样安排的：3.1 节给出了常见的几种特征点提取方法；3.2 节介绍了利用改进遗传算法求解特征点匹配的具体过程；3.3 节给出了双目图像对校正过程以及校正后的结果；3.4 给出本章小结部分。

### 3.1 特征点提取

特征提取可分为提取点特征、线特征和面特征。特征匹配也可在这三种方式下采用不同的方法匹配，但多数情况下，无论是线特征还是面特征最终也要转化为点特征。因此，本节将简单介绍点特征以及其在图像配准中的发展。

点特征是指在一幅图像内灰度在水平和垂直方向都有显著变化的一类特殊点，具有信息量少、信息含量高的特点，已经被广泛应用于计算机视觉的诸多领域。特征是用来建立两幅图像之间的匹配对应关系。提取“好”的特征点是图像匹配非常关键的一步。特征提取的难点在于自动、稳定、一致性的特征提取。

最简单的配准方法即人工选取图像上一系列同名特征点对，带入多项式以得到图像的转换参数。人工选点的方法具有错误率低，灵活性高，适应性好的特点，但在大量数据处理的应用中要耗费巨大的人力，一般的自动点匹配算法利用图像的固有性质如角点、边缘、形状、封闭区域的重心等获得控制点。

国内外学者针对不同的图像配准应用问题进行了大量的研究工作，其中基于特征的图像配准的技术起源于 Moravec<sup>[57]</sup>在 1981 年利用角点对立体图像进行配准。Harris 和 Stephens<sup>[58]</sup>在 1988 年对其进行了改进，在 1992 年提出了 Harris 角点检测，并将其应用于运动物体跟踪和三维重建，从此 Harris 角点检测在图像配准的其它方面得到了广泛应用。

Tomasi 和 Kanade<sup>[59][60]</sup>提出了 KLT 特征点用于视频中的目标跟踪和稳像，取得了不错的效果。剑桥大学的 Lisa Gottesfeld Brown<sup>[64]</sup>总结了图像配准的主要理论及其在各个领域的应用。当时他讨论的图像配准技术应用主要还是着眼于医学图像处理、遥感图像处理等传统应用领域。Harris 角点检测对于图像的尺度变化比较敏感，因此不同尺度的图像难以产生好的匹配效果。于是 Lowe<sup>[61]-[63]</sup>提出了基于尺度空间和不变量理论的图像特征描述算子——SIFT(Scale Invariant Feature Transform)的概念，即用 SIFT 技术得到的图像特征向量对尺度的变化保持不变性，将其应用于图像检索。随着该算法的发展，SIFT 算子已不仅仅对尺度变化保持不变性，而且对旋转、尺度缩放、部分的三维视角变化和部分光照变化保持不变，从而成为稳定的、适应性较强的局部特征匹配算子。M Brown 在 2003 将 SIFT 技术应用于全景图像自动拼接，实验中所采用的图片是用同一架相机在等焦距上围绕固定轴旋转一圈所拍摄的图片的配准和拼接。David G Lowe 在 2004 年将 SIFT 技术应用于目标识别。SIFT 以其独特的优越性被应用到各个方面，据不完全统计，到目前为止围绕着 SIFT 算法所撰写的论文超过了 3500 篇。下面我们给出两种常用的特征点：Harris 角点和 SIFT 特征点的生成方法。

### 3.1.1 Harris 角点

Harris 角点是一种常用的特征点提取方法，其原理是从图像局部的小窗口观察图像特征，当一个窗口在图像上移动时，若沿任意方向的移动都会导致窗口内的灰度值发生明显变化，我们认为该点为角点。

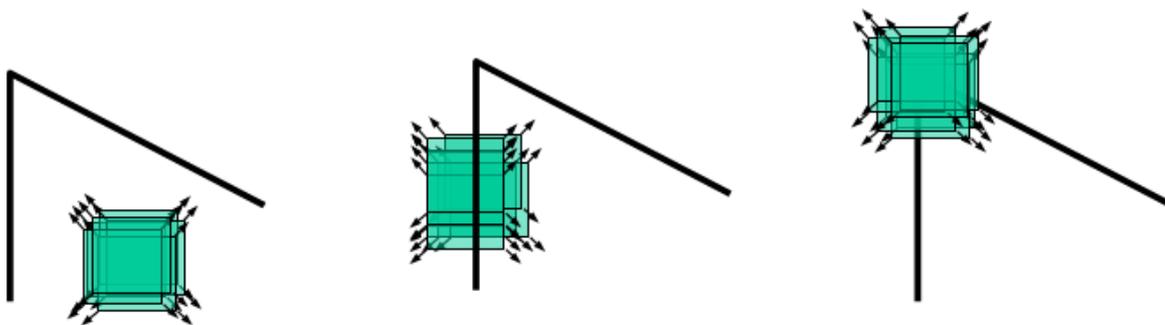
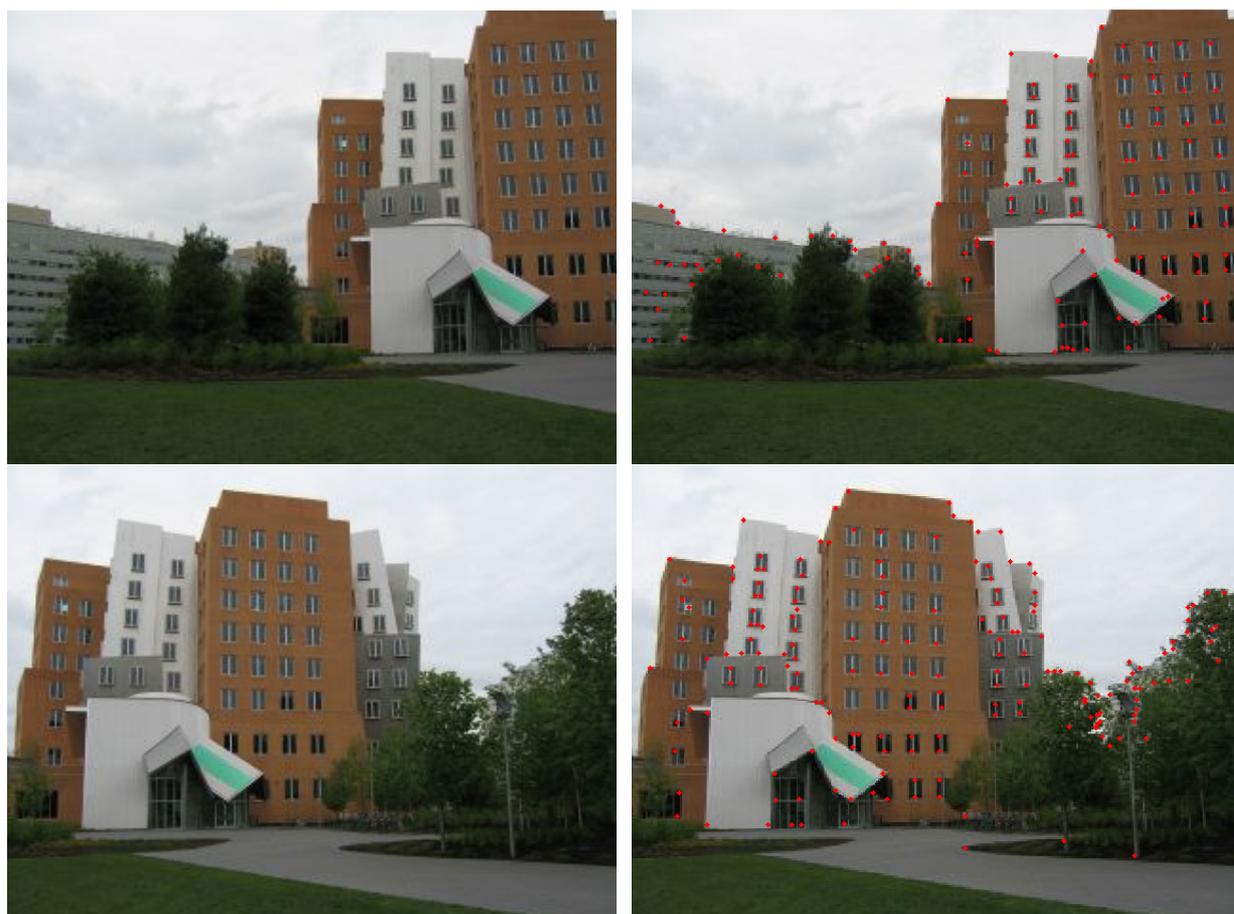


图 3-1 窗口在图像上移动示例<sup>[58]</sup>

从图 3-1 可以看出，当窗口在平滑区域沿各个方向上移动时，窗口内的灰度值没有变化，当窗口沿着边缘的方向上移动时，窗口内的灰度值也没有变化，但是窗口在角点处的沿着各个方向移动时，灰度值发生了变化，Harris 正是利用了特征来检测图像中的角点。

Harris 角点检测可以很有效的提取出目标物体的特征点，并且计算简单，只用到了一阶差分及滤波，它对图像中的每个角点都计算其兴趣值，然后选择最优点，实验证明，Harris 角点探测在纹理信息丰富的区域中可以提取出大量有用的特征点，而在纹理信息少的区域提取的特征点较少。

图 3-2 是 Harris 角点在实验图像上提取的结果。



(a) 原图

(b) 提取 Harris 角点后的图

图 3-2 Harris 角点检测

### 3.1.2 SIFT 特征点

David Lowe 在 2004 年总结了现有的基于不变量技术的特征检测方法，并正式提出了一种基于尺度空间的、对图像缩放、旋转甚至仿射变换保持不变性的图像局部特征描述算子 SIFT 算子。SIFT 特征描述子是目前比较流行的特征描述算法，它能够在发生仿射变换和局部畸变的两幅图像上确定可靠的特征匹配点对。它受到大家推崇的原因主要是因为相同类型的特征描述子中，它能得到最好的效果。

利用 SIFT 算法提取的 SIFT 特征向量具有如下特性：

- 1) SIFT 特征是图像的局部特征，其对旋转、尺度缩放、亮度变化保持不变性，对视角变化、仿射变换、噪声也保持一定程度的稳定性。
- 2) 独特性好，信息量丰富，适用于在海量特征数据库中进行快速、准确的匹配。
- 3) 多量性，小图片也可以产生大量 SIFT 特征向量。
- 4) 高速性，经优化的 SIFT 匹配算法甚至可以达到实时的要求。
- 5) 可扩展性，可以很方便的与其他形式的特征向量进行联合。

下面我们给出 SIFT 特征点提取的方法和 SIFT 描述子的表示方法。

### 3.1.2.1 SIFT 特征点提取

SIFT 特征点提取算法首先在尺度空间进行特征检测，并确定关键点的位置和关键点所处的尺度，然后使用关键点邻域梯度的主方向作为该点的方向特征，以实现算子对尺度和方向的无关性。下面将给出 SIFT 特征点提取算法的步骤：

- 1) 尺度空间极值检测，初步确定关键点位置和所在尺度，图 3-3 是高斯差分图。



图 3-3 高斯差分图

图 3-4 给出了相邻尺度下局部极值点的检测图。

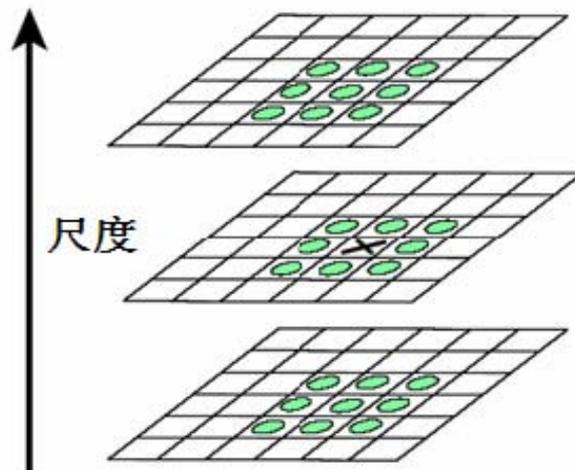


图 3-4 尺度空间局部极值点检测<sup>[63]</sup>

- 2) 精确确定关键点的位置

初步确定了关键点的位置和尺度后，这一步我们要精确定位关键点的位置。Brown 提出通过拟合一个三元二次函数来确定极值点的精确位置，得到关键点的位置信息之后，去除低对比度的关键点和不稳定的边缘响应点(因为高斯差分算子会产生较强的边缘响应)是很有必要的。

- 3) 确定关键点的方向

利用关键点邻域像素的梯度方向分布特性为每个关键点指定方向参数，使算子具备旋

转不变性。

在以关键点为中心的邻域窗口内采样，用直方图统计邻域像素的梯度方向。图 3-5 是采用 7 个柱时，使用梯度直方图为关键点确定主方向的示例。

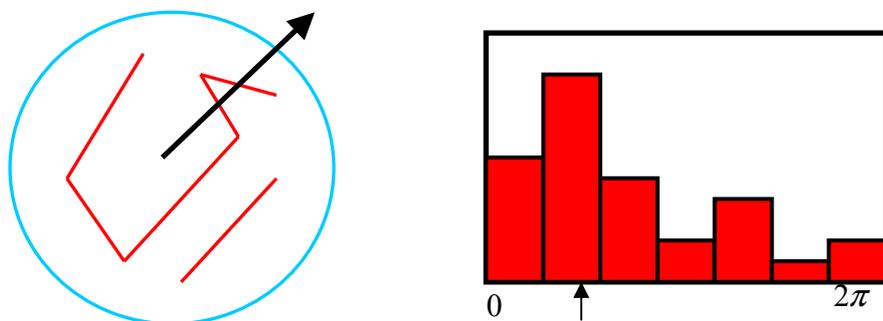
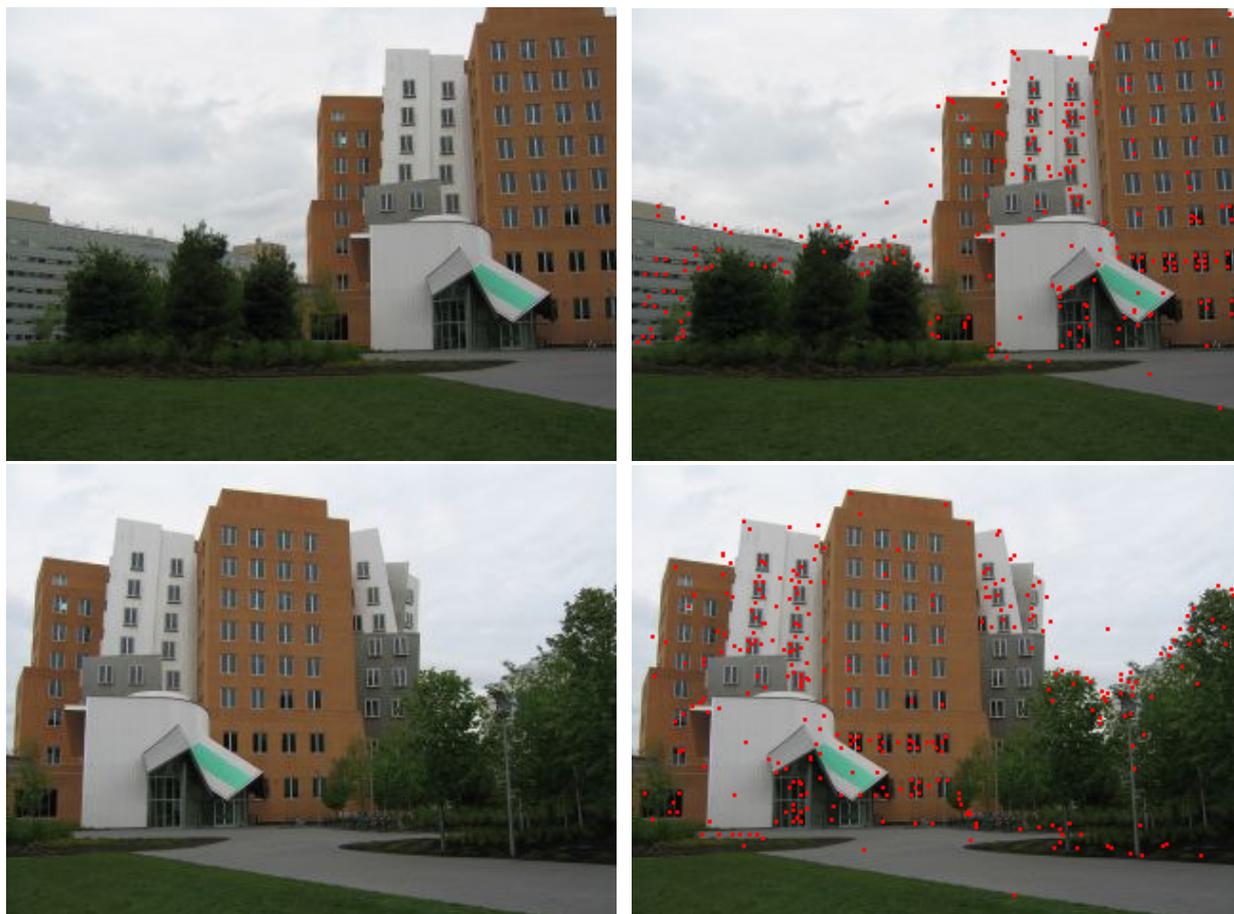


图 3-5 利用梯度方向直方图确定主梯度方向<sup>[63]</sup>

至此，对于每幅图像，我们可以根据上述算法提取其 SIFT 特征点，每个特征点有三个信息：位置、尺度、方向。图 3-6 给出了按照以上步骤在实验图片上提取 SIFT 特征点（关键点）的结果图，采用红点标注特征点的位置，两幅图像上提取的 SIFT 特征点的数量分别为 333 和 376 个，和 Harris 角点相比，SIFT 特征点比较多。



(a) 原图

(b) SIFT 特征点图

图 3-6 提取 SIFT 特征点图

### 3.1.2.2 SIFT 特征描述子的表示

在确定了 SIFT 特征点的位置、尺度和方向后，图 3-7 给出了 SIFT 特征点的向量表示方法。首先将坐标轴旋转为关键点的主方向，以确保旋转不变性。

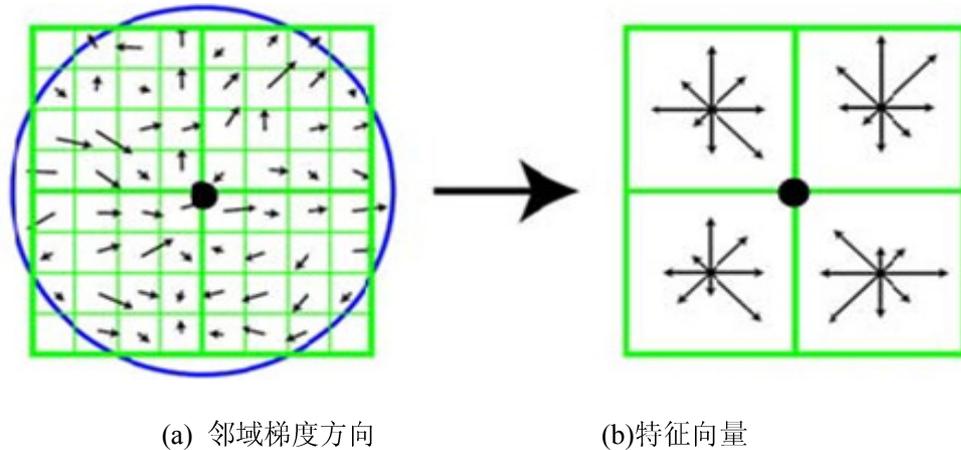


图 3-7 由关键点邻域梯度信息生成特征向量<sup>[63]</sup>

以关键点为中心取  $8 \times 8$  的邻域窗口。上图左部分的中央黑点为当前关键点的位置，每个小格代表关键点所在尺度空间的一个像素，箭头方向代表该像素的梯度方向，箭头长度代表梯度模值，图中蓝色的圈代表高斯加权的范围（越靠近关键点的像素梯度方向信息贡献越大）。然后在每  $4 \times 4$  的小块上计算 8 个方向的梯度方向直方图，绘制每个梯度方向的累加值，即可形成一个种子点，如图 3-7 右部分所示。该图中一个关键点由  $2 \times 2$  共 4 个种子点组成，每个种子点有 8 个方向向量信息。

为了增强匹配的稳健性，Lowe 建议对每个关键点使用  $4 \times 4$  共 16 个种子点来描述，一个关键点就可以产生 128 个数据，最终形成 128 维的 SIFT 特征向量。此时 SIFT 特征向量已经去除了尺度变化、旋转等几何变形因素的影响，再继续将特征向量的长度归一化，则可以进一步去除光照变化的影响。

下面我们在图 3-7 的基础上给出了 SIFT 特征描述子，采用红色箭头的尾端标注特征点在图像上的坐标位置，箭头的顶端朝向代表特征点的主方向，若存在另一个方向的峰值超过主峰值的 80%，则箭头还有一个辅方向。箭头的长度与特征点所在的尺度成正比，尺度越大，箭头主体部分越长。

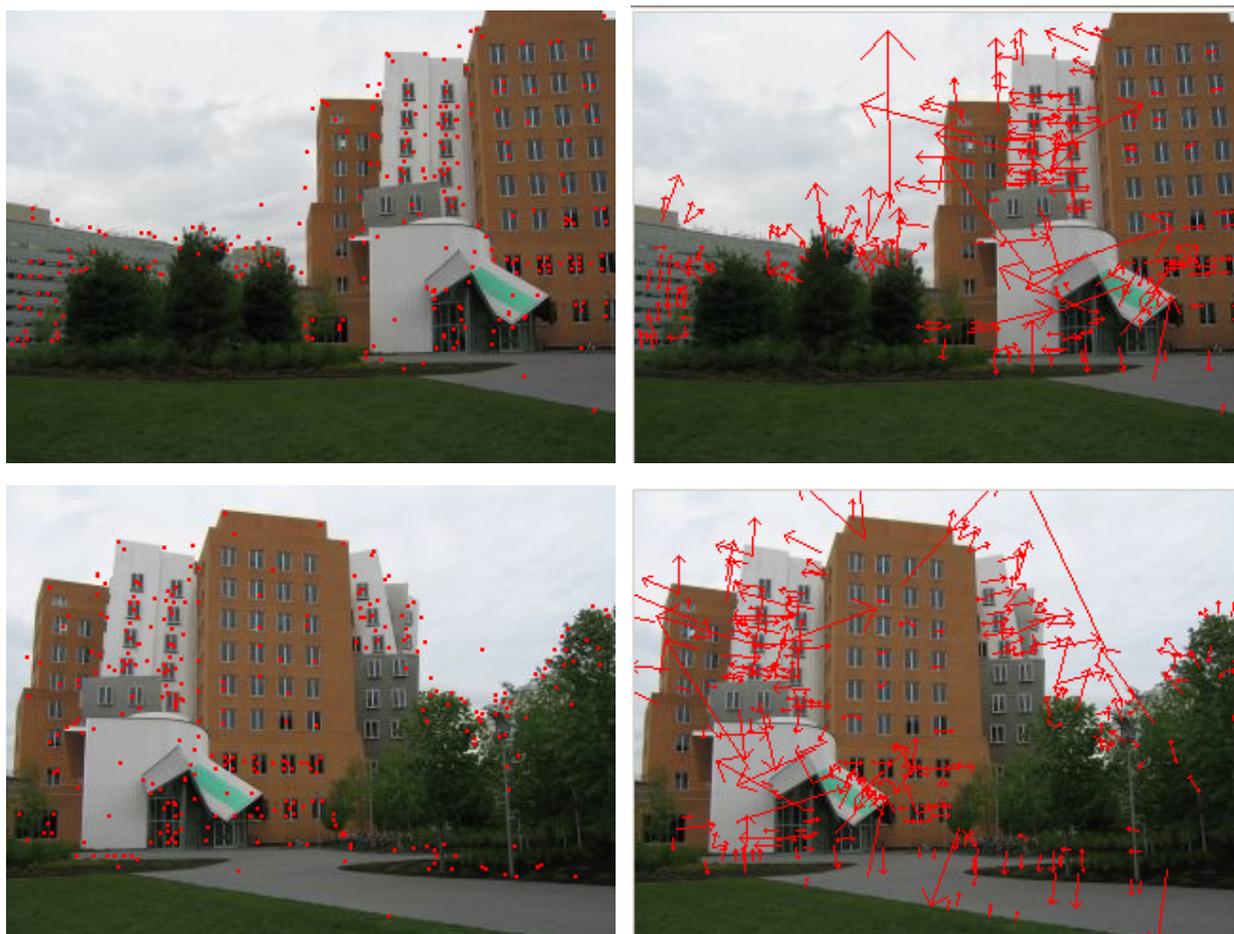


图 3-8 SIFT 特征描述子

6) 由于 SIFT 特征点有很多的优势, 比如对旋转、尺度缩放、亮度变化保持不变性, 对视角变化、仿射变换、噪声也保持一定程度的稳定性等。

因此, 在本文中, 我们提取图像的特征点并对其进行匹配。

### 3.2 基于改进遗传算法的特征点匹配

在现有的特征点匹配算法中, KD-tree<sup>[65]-[68]</sup>是运用比较广泛的算法之一。它通过为图像特征描述子查找最相似的配对从而建立匹配的数据结构, 该算法计算速度比较快, 但是准确度不是很高。为此, Chanop Silpa-Anan<sup>[69]</sup>改进了KD-tree的搜索性能, 最初通过在原始的数据集合中建立多个KD-tree并同时已在建立的多个KD-tree上查找, 缩短了搜索时间; 后来使用主成分分析法对坐标轴里的数据进行主轴排列, 提高主要数据的检索概率。

Xie<sup>[70]</sup>提出了在建立特征点匹配之前先建立块匹配, 在块对应的基础上建立特征点匹配, 这样可以减小搜索的范围, 提高匹配准确度。Myronenko<sup>[71]</sup>提出了在场景发生一定变形的情形下, 将运动一致性运用到匹配求解中。Alexander<sup>[72]</sup>将特征匹配算法和块匹配算法相结合, 并且两者起到互相约束的作用。

Thomas<sup>[73]</sup>提出了一种对于非刚体变换图像模型达到像素级匹配的匹配算法。在图像的

四维图空间中不断的迭代优化目标函数，其中这一思想的代表算法是ICP(Iterative Closest Point)算法<sup>[74]</sup>，假定给出一组初始值，ICP算法以欧氏距离作为匹配度量函数，计算变换系数矩阵，将得到的系数值带入，不断的迭代优化目标函数，直至达到目标要求。该算法的不足之处就是容易陷入局部最优。Fitzgibbon<sup>[75]</sup>引入了一种鲁棒因子，通过Levenberg-Marquardt 算法来优化代价函数，虽然这种方法在一定程度上改善了收敛的窄带，但该算法大大依赖初始值，噪声、聚类和外点的影响可能会导致该方法失效。还有一些启发式算法，如遗传算法<sup>[77]-[84]</sup>，模拟退火算法等，通过模拟生物进化来求解目标函数的最优解。图3-9给出了利用遗传算法求解目标函数的流程图。

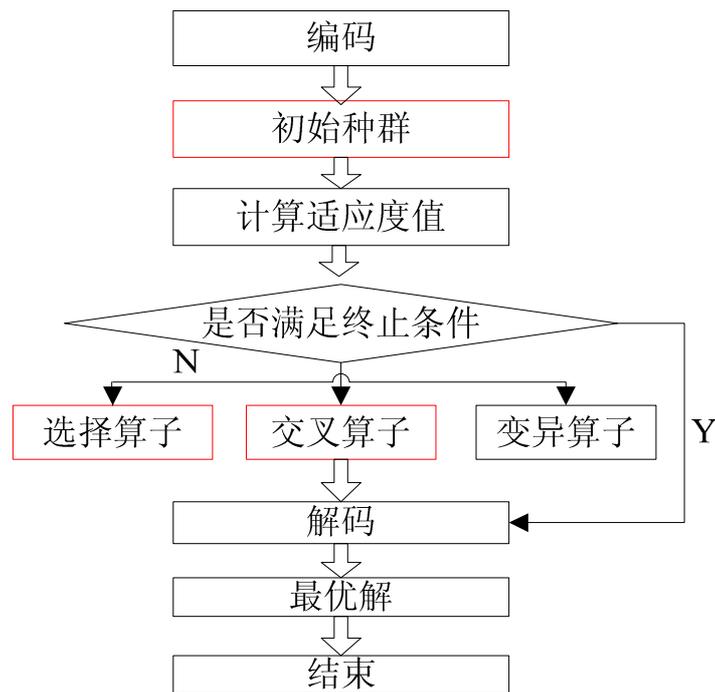


图 3-9 遗传算法求解流程图

在本节中，我们提出了一种由粗到精的特征点匹配算法，首先建立了匹配问题的数学模型，然后利用互相相似度算法和改进遗传算法对该模型求解。

### 3.2.1 建立匹配问题的数学模型

假设  $G_1$  和  $G_2$  是两幅图像，这两幅图像可能取自不同时间，不同视角下的同一个场景， $P^1$  ,  $P^2$  分别对应图像  $G_1$  和  $G_2$  的 SIFT 特征点集合，分别被表示为：  
 $P^1 = \{(x_i^1, y_i^1), i = 1, 2, \dots, m\}$  ,  $P^2 = \{(x_s^2, y_s^2), s = 1, 2, \dots, n\}$  。由于匹配的特征点对是相似的且相互匹配的特征点对的对应边是兼容的<sup>[76]</sup>，根据这两个特征，我们建立了下面的数学模型：  
 模型 (I):

$$\begin{aligned}
 E(M) = \max & \left[ \sum_{i=1}^m \sum_{s=1}^n \sum_{j=1}^m \sum_{t=1}^n M_{p_i^1 p_s^2} M_{p_j^1 p_t^2} C(w_1(p_i^1, p_j^1), w_2(p_s^2, p_t^2)) + \alpha \sum_{i=1}^m \sum_{s=1}^n M_{p_i^1 p_s^2} S(p_i^1, p_s^2) \right] \\
 \text{subject to: } & \sum_{i=1}^m M_{p_i^1 p_s^2} \leq 1, \sum_{s=1}^n M_{p_i^1 p_s^2} \leq 1, M_{p_i^1 p_s^2} \in \{0, 1\}
 \end{aligned} \tag{3.1}$$

在上面的能量函数  $E(M)$  中，第一部分是空间约束，第二部分是相似度约束。其中  $w_1(p_i^1, p_j^1)$  表示图  $G_1$  中特征点  $p_i^1$  和  $p_j^1$  之间的空间距离， $w_2(p_s^2, p_t^2)$  表示图  $G_2$  中特征点  $p_s^2$  和  $p_t^2$  之间的空间距离， $C$  表示  $w_1$  和  $w_2$  的距离函数， $m$  和  $n$  分别表示图  $G_1$  和图  $G_2$  中特征点的个数， $S(p_i^1, p_s^2)$  表示特征点  $p_i^1$  和  $p_s^2$  之间的相似性度量函数， $\alpha$  表示能量函数  $E(M)$  中空间约束与相似度约束之间的权重系数，矩阵  $M$  是一个 0-1 矩阵，由下面的式子定义：

$$M_{p_i^1 p_s^2} = \begin{cases} 1 & \text{若 } p_i^1, p_s^2 \text{ 匹配} \\ 0 & \text{其他} \end{cases} \tag{3.2}$$

在本节中，我们定义  $C(m, n) = mn$ ，显然，特征点对之间的相似度越大且兼容性越好，那么特征点对的匹配效果就越好，这时得到的能量函数  $E(M)$  的值也就越大。

由于模型 (I) 中目标函数的两个组成部分是相互独立的，于是我们将这个模型等价的分解成两个子模型分别求解。

模型(II):

$$\begin{aligned}
 E_1(M) = \max & \sum_{i=1}^m \sum_{s=1}^n M_{p_i^1 p_s^2} S(p_i^1, p_s^2) \\
 \text{subject to: } & \sum_{i=1}^m M_{p_i^1 p_s^2} \leq 1, \sum_{s=1}^n M_{p_i^1 p_s^2} \leq 1, M_{p_i^1 p_s^2} \in \{0, 1\}
 \end{aligned} \tag{3.3}$$

模型(III):

$$\begin{aligned}
 E_2(M) = \max & \sum_{i=1}^m \sum_{s=1}^n \sum_{j=1}^m \sum_{t=1}^n M_{p_i^1 p_s^2} M_{p_j^1 p_t^2} C(w_1(p_i^1, p_j^1), w_2(p_s^2, p_t^2)) \\
 \text{subject to: } & \sum_{i=1}^m M_{p_i^1 p_s^2} \leq 1, \sum_{s=1}^n M_{p_i^1 p_s^2} \leq 1, M_{p_i^1 p_s^2} \in \{0, 1\}
 \end{aligned} \tag{3.4}$$

针对上述两个模型的求解，我们提出了一种由粗到精的求解策略，先求模型 (II) 的初始近似解，然后在这个近似解的基础上利用改进遗传算法求解模型 (III) 的解。

### 3.2.2 特征点粗匹配

在本节中，我们先利用角度相似性度量函数来衡量特征点之间的相似度，定义如下的相似性度量函数  $S(p_i^1, p_s^2)$ ：

$$S(p_i^1, p_s^2) = \frac{A(p_i^1) \cdot A(p_s^2)^T}{|A(p_i^1)| |A(p_s^2)|} \quad (3.5)$$

其中  $p_i^1$  和  $p_s^2$  分别是图像  $G_1$  和  $G_2$  两个特征点， $A(p_i^1)$  和  $A(p_s^2)$  是其相对应的 SIFT 向量描述子。显然，这两个描述子越相似，那么其对应的相似度函数值越接近于 1，也就是说这两个特征点匹配的可能性也就越大。

于是模型 (II) 转变为：

$$\begin{aligned} E_1(M) = \max \sum_{i=1}^m \sum_{s=1}^n M_{p_i^1 p_s^2} \cdot \frac{A(p_i^1) \cdot A(p_s^2)^T}{|A(p_i^1)| |A(p_s^2)|} \\ \text{subject to: } \sum_{i=1}^m M_{p_i^1 p_s^2} \leq 1, \sum_{s=1}^n M_{p_i^1 p_s^2} \leq 1, M_{p_i^1 p_s^2} \in \{0, 1\} \end{aligned} \quad (3.6)$$

模型 (II) 一般利用匈牙利算法求解，然而，由于 SIFT 特征点个数相对比较多，因此，利用匈牙利算法求解模型 (II) 计算速度很慢，为了应对这一情况，我们提出了一种基于贪婪算法的互相相似度算法（简称 MS）来求解模型 (II) 的局部最优解。

---

#### 算法 3.1: MS 算法

---

- (1) 对图  $G_1$  中的特征点  $p_i^1$ ，与图  $G_2$  中的每个特征点计算相似度并利用贪婪算法找到最相似的特征点  $p_j^2$ ，若  $p_i^1$  与  $p_j^2$  的相似度大于一个给定的阈值，则认为  $p_i^1$  与  $p_j^2$  对应。
  - (2) 对图  $G_2$  中的特征点  $p_j^2$ ，与图  $G_1$  中的每个特征点计算相似度并利用贪婪算法找到最相似的特征点  $p_k^1$ ，若  $p_j^2$  与  $p_k^1$  的相似度大于一个给定的阈值，则认为  $p_j^2$  与  $p_k^1$  对应。
  - (3) 若  $p_i^1$  与  $p_k^1$  是相同的特征点，则  $(p_i^1, p_j^2)$  是一对匹配点。
- 

从上面的计算过程中我们可以看出，由 MS 算法计算得到的解是模型 (II) 的一个局部最优解，然而，这些解已经能够满足我们的应用需求，这一点从图 3-10 可以看出。

利用 MS 算法求解模型 (II) 有以下几点好处：

- (i) 计算速度快；
- (ii) 几乎没有去掉任何满足模型 (II) 的正确解，但是去掉了很多错误解，降低了解空间的维数；

(iii)将多对多的特征点匹配问题转化为一对一的特征点匹配问题，降低了算法的计算复杂度；



图 3-10 特征点粗匹配结果

### 3.2.3 特征点精匹配

从上面的特征点粗匹配结果我们可以看出，由 MS 算法得到的解不仅仅包含大量的正确解，同时还包含一些错误解，这些可能是由于图像的拍摄角度不同，噪声引起的，另外我们不得不考虑仅仅根据相似度约束并不能保证特征点之间的正确对应，我们还需要考虑特征点相互之间的位置信息，也就是模型 (III)。

下面在模型 (II) 的解的基础上求解模型(III)，模型 (III) 是一个优化问题，我们采用启发式算法-遗传算法 (简称 GA) 来求解。为了提高 GA 的速度和效率，我们对标准的 GA 算法做了一些改进：

- 1) 整数编码；
- 2) 对初始种群进行了优化并将粗匹配的结果加入到初始种群中；
- 3) 采用了稳态选择策略；
- 4) 采用了多点交叉方法。

#### 3.2.3.1 编码

GA 一般采用二进制的编码策略，然而当遇到维数比较高的时候，码会很长而且容易出

错。为了避免这一情况，我们采用了整数编码的策略，定义编码函数  $K: K(i) = j$ ，该函数表示图  $G_1$  中第  $i$  个特征点与图  $G_2$  中第  $j$  个特征点对应，每一个映射  $K$  是一个可行解，也称染色体。在图 3-11 中， $K_1$ ， $K_2$ ， $K_3$  和  $K_4$  分别代表一个染色体。

### 3.2.3.2 种群的初始化

在这一部分，我们用贪婪搜索方法来优化初始种群，优化初始种群的具体过程由下面的算法给出：

---

#### 算法 3.2 种群初始化算法

---

1、随机产生初始种群  $P$ ，假设种群的大小表示为  $pop\_size$ ；

2、设  $P^* = \emptyset$ ，对初始种群  $P$  中的每一个染色体  $\pi \in P$ ：

$$\pi^* = GS(\pi);$$

$$P^* = P^* \cup \{\pi^*\};$$

则： $P^*$  是优化后的初始种群，其中函数  $GS$  的定义如下：

对每一个  $\pi \in P$ ，当  $Q = 1:\tau$  时

(i) 随机生成整数  $m$ ， $1 \leq m \leq n$ ；

(ii) 交换  $N(m)$  中的基因，生成新染色体  $\pi'$ ，若  $E_2(\pi') > E_2(\pi)$ ，则用  $\pi'$  代替  $\pi$ ，否则保留  $\pi$ ；

其中  $N(i)$  定义为： $N(i) = \{j \mid |i - j| \leq 2\}$ ， $\tau$  表示搜索的次数， $n$  表示染色体的长度， $E_2$  对应着模型 (III) 的目标函数。

根据算法 3.2 可以得到优化后的初始种群，特别地，粗匹配的结果也作为一个可行解并加入到该种群中。

### 3.2.3.3 选择算子

在改进的 GA 中，我们采用了稳态选择策略，设  $f_k$  ( $k = 1, 2, \dots, pop\_size$ ) 表示每个染色体的自适应函数值，适应度函数定义为模型 (III) 的目标函数  $E_2(M)$ ，选择算子的阈值  $\bar{f}$  定义如下：

$$f = \sum_{k=1}^{pop\_size} f_k \quad (3.7)$$

$$\bar{f} = f / pop\_size \quad (3.8)$$

如果一个染色体的自适应函数值大于阈值  $\bar{f}$ ，那么这个染色体就被保留，否则，删除该染色体，然后重新生成一个新的染色体。在这个选择过程中，我们扩大了 GA 中可行解的搜索范围。

### 3.2.3.4 交叉算子

在改进的 GA 中，我们采用了多点交叉策略，下图 3-11 给出了多点交叉<sup>[84]</sup>的具体过程，在该过程中，当前的两个染色体  $K_1$  和  $K_2$  交叉生成了下一代的新的染色体  $K_3$  和  $K_4$ ，我们比较这四个染色体  $K_1$ ， $K_2$ ， $K_3$ ， $K_4$  的自适应度值，适应度值最大的两个将被保存到下一代中。在本算法中，交叉概率的设置是 0.8。

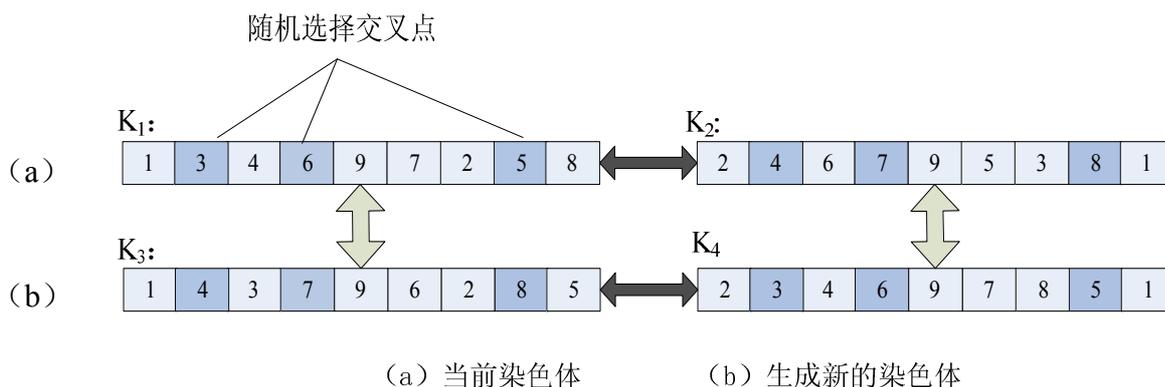


图 3-11 三点交叉策略<sup>[84]</sup>

在染色体进行完交叉后，为了保持种群的多样性，我们对种群进行变异，变异概率  $p_m=0.1$ 。

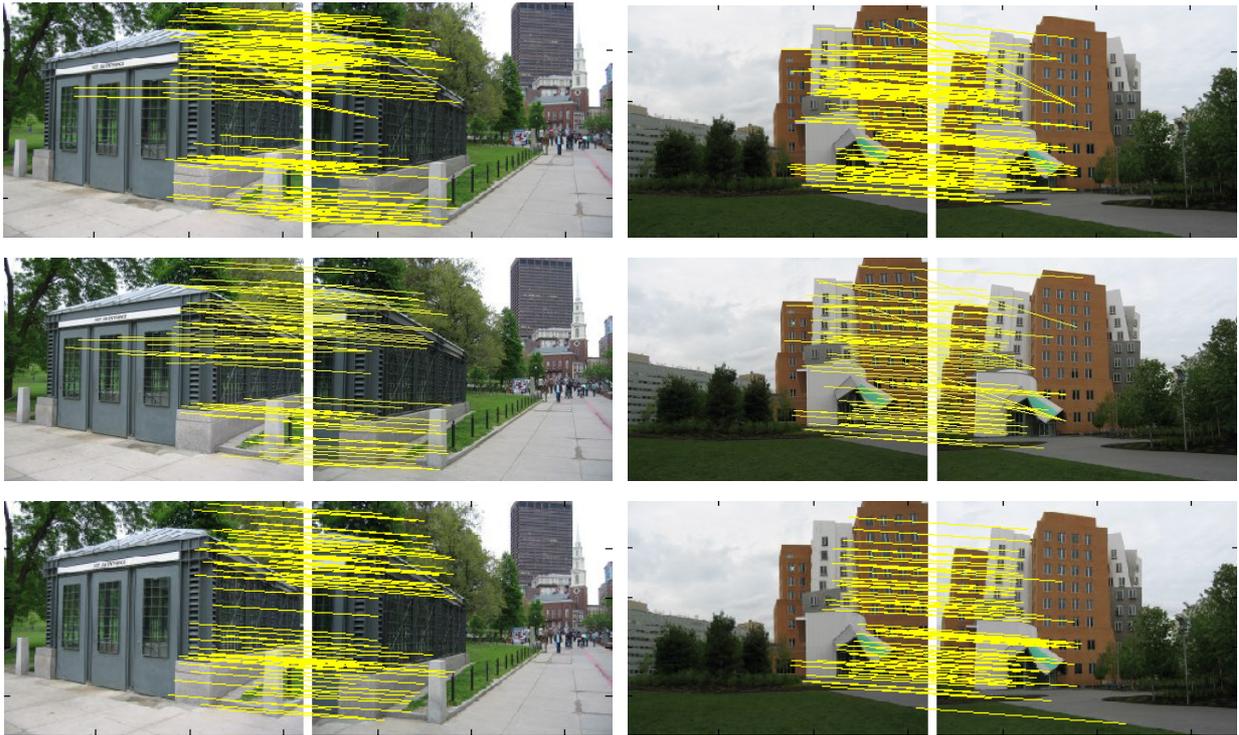
在实验中，我们发现当 GA 的迭代次数接近于 1000 时，此时可以得到的稳定的最优解。

### 3.2.4 实验结果与比较

在这一节中，我们给出本文提出的算法与 Lowe 的方法，KD-Tree+RANSAC 算法在数据样本集合上的结果，这些样本集分别来自 ICCV2005 上的两个测试集<sup>[85]</sup>以及我们的测试集<sup>[86]</sup>，分别包含 29，38 以及 30 幅图像。这些图像大都是不同视角、不同光照、不同尺度下的宽基线图像。为了计算方便，我们将每组实验中图像的大小缩小到 320\*240。

图 3-12 中，我们选择不同视角、不同尺度下具有很多相似特征点的两对图像来进行实验。每幅图像对中都有很多相似的特征点，比如：建筑物的窗户，铁栅栏之间的空隙等。下图中第一行是 Lowe's 的实验结果，第二行是由 KD-Tree + RANSAC 算法得到的结果，

第三行是本文算法结果。



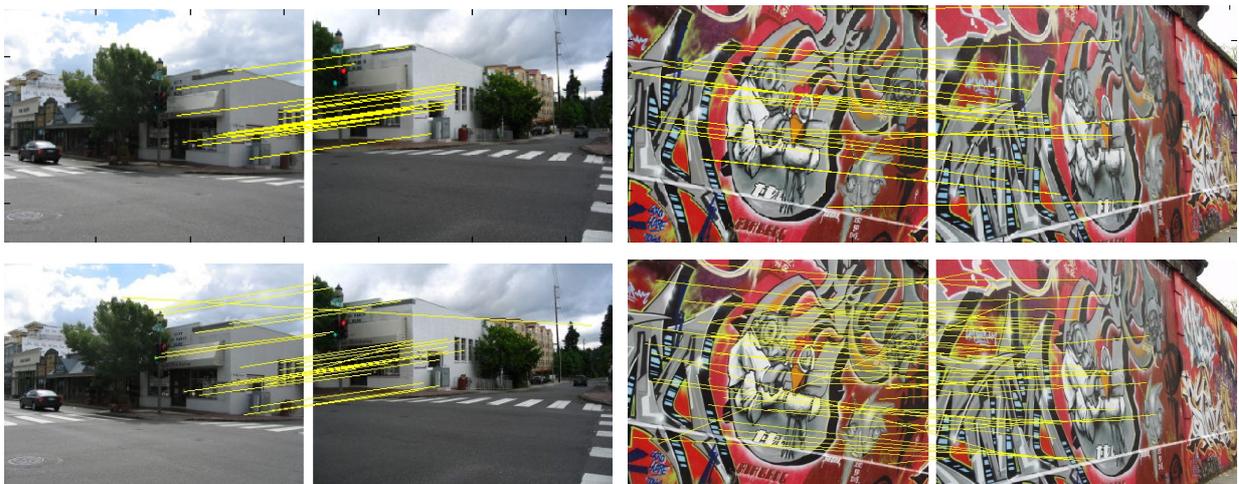
(a) 第一组实验结果比较

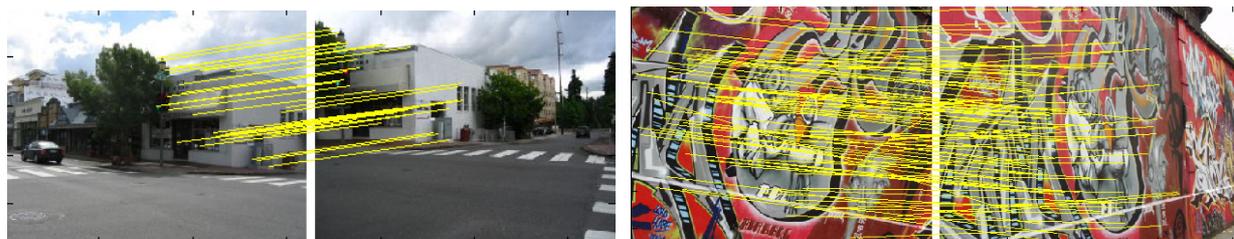
(b) 第二组实验结果比较

图 3-12 实验结果比较

可以看出，在 Lowe's 的算法中，这些相似点由于缺乏可辨别性所以都被舍弃了；而在 KD-Tree + RANSAC 算法中，只利用了相似性约束，因此还是很难区别比较相似的特征点对的；我们提出的算法取得了不错的实验结果，不仅匹配对数多，而且没有误匹配。

图 3-13 中，我们选择了尺度变化和光照变化都比较大的两组图像，从下面实验的结果可以看出，本文与其他另外两种算法的结果相比，不仅没有有误匹配，而且匹配对数也多一些。



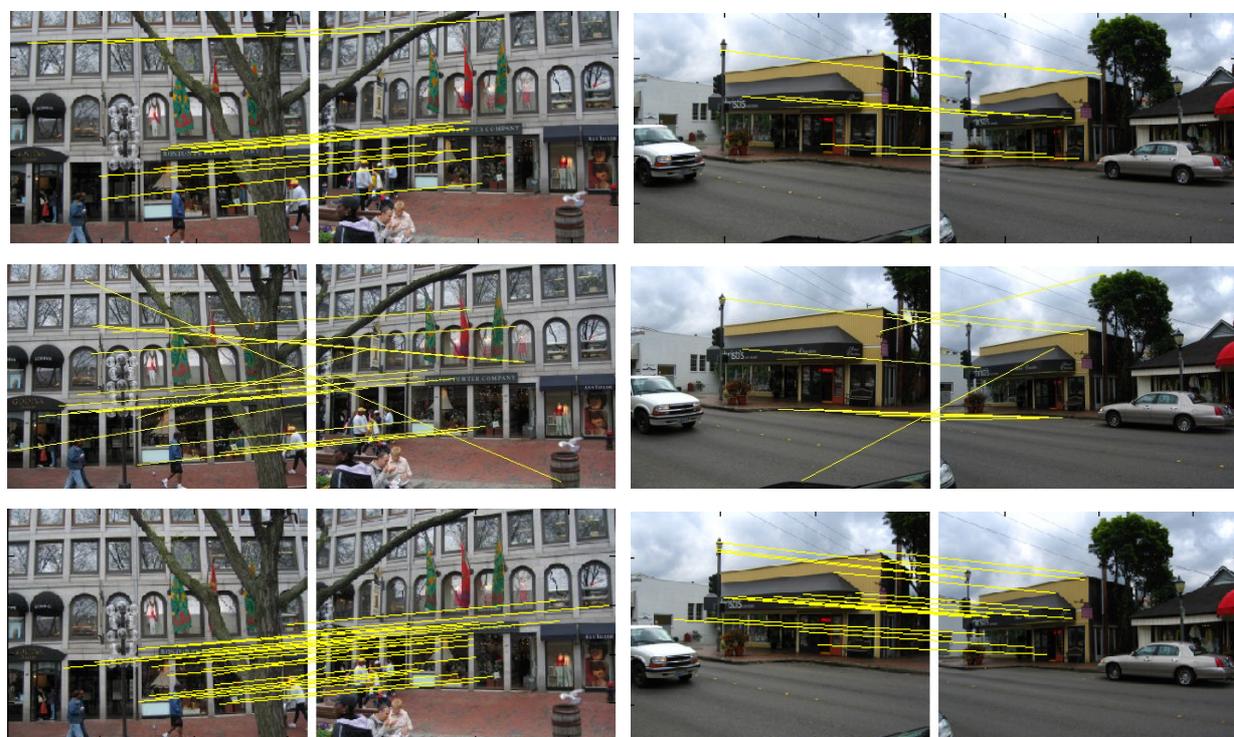


(c) 第三组实验结果

(d) 第四组实验结果

图 3-13 实验结果比较

图 3-14 中，我们选择了尺度变化比较大且宽基线下的图像对。本文与其他另外两种算法的结果相比，不仅没有有误匹配，而且匹配对数也多一些。尤其是 KD-Tree + RANSAC 算法得到匹配点对，不仅少，而且误匹配多，在这种情况下，根本不能准确的计算其基础矩阵，无法完成图像校正。



(e) 第五组实验结果

(f) 第六组实验结果

图 3-14 实验结果比较

为了进一步说明本文算法在特征点匹配方面的优势，我们在测试图像上定量的比较了匹配的精度，其中图像特征点匹配精度的定义如下：

$$P = \frac{\text{匹配正确的特征点对数目}}{\text{匹配正确的特征点对数目} + \text{匹配错误的特征点对数目}}$$

表 3-1 给出了三种算法在测试集上的平均精度值的定量比较。从表中可以看出，我们算法的匹配精度明显优于其他两种算法的精度，尤其是在我们的测试图片上更加明显一些。总的来说，我们提出的算法比其他两种算法在测试集上的匹配精度提高了大约 10% 左右。

表 3-1 匹配精度比较

图像集合	图像对数	P of KD-Tree + RANSAC	P of David Lowe's	P of Proposed method
Test4	29	0.74	0.70	0.80
Final5	38	0.70	0.71	0.87
Our image sets	30	0.84	0.78	0.92
Total	97	0.76	0.73	0.86

### 3.3 双目图像对校正

在得到正确匹配对之后，我们将对采集的图像对进行校正，首先标定摄像机内参数，然后利用改进 8 点法计算基础矩阵，进一步计算两个摄像机之间的旋转矩阵和平移向量，最后我们对实验图片进行校正。

#### 3.3.1 摄像机内参数自标定

我们采用了基于标定板的摄像机内参数标定方法<sup>[87]-[93]</sup>。实验中摄像头固定，通过旋转标定板得到不同角度下的图像。图 3-15 是其中的几幅标定图像。

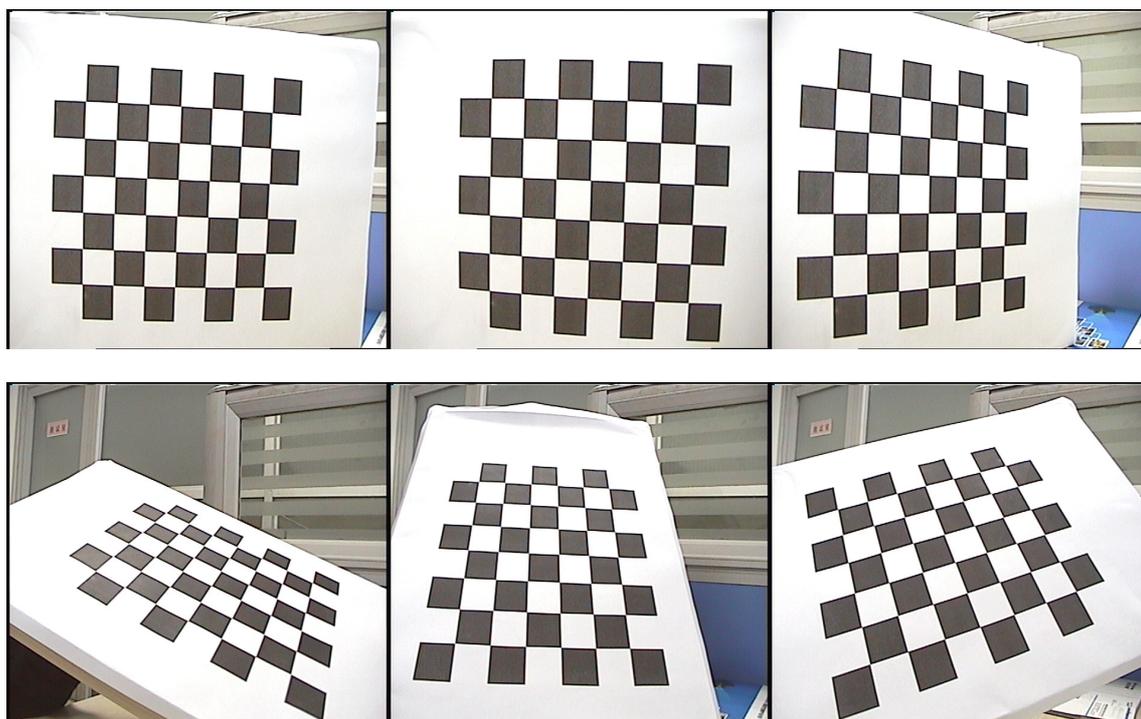


图 3-15 标定板图像

实验得到两个摄像头的总体平均误差为：0.368988 像素，0.313311 像素。

两个摄像机的内参数矩阵分别为：

$$K_1 = \begin{Bmatrix} 694.7 & 0 & 317.533 \\ 0 & 821.339 & 286.939 \\ 0 & 0 & 1 \end{Bmatrix}; K_2 = \begin{Bmatrix} 698.064 & 0.302 & 302.484 \\ 0 & 824.259 & 262.738 \\ 0 & 0 & 1 \end{Bmatrix}$$

畸变系数分别为：{-0.238058, 0.177433, 0.00141298, 0.00161284}

{-0.272147, 0.328438, 0.00203419, 0.00160238}

### 3.3.2 校正后的结果

在得到摄像机内参数后，我们用改进8点法计算基础矩阵，进而得到旋转矩阵和平移向量，下面给校正后的实验结果。

图3-16和图3-17中第一行和第三行是双目摄像机拍的原始图片对，第二行和第四行分别是校正后的实验结果。



图 3-16 双目校正后的实验结果

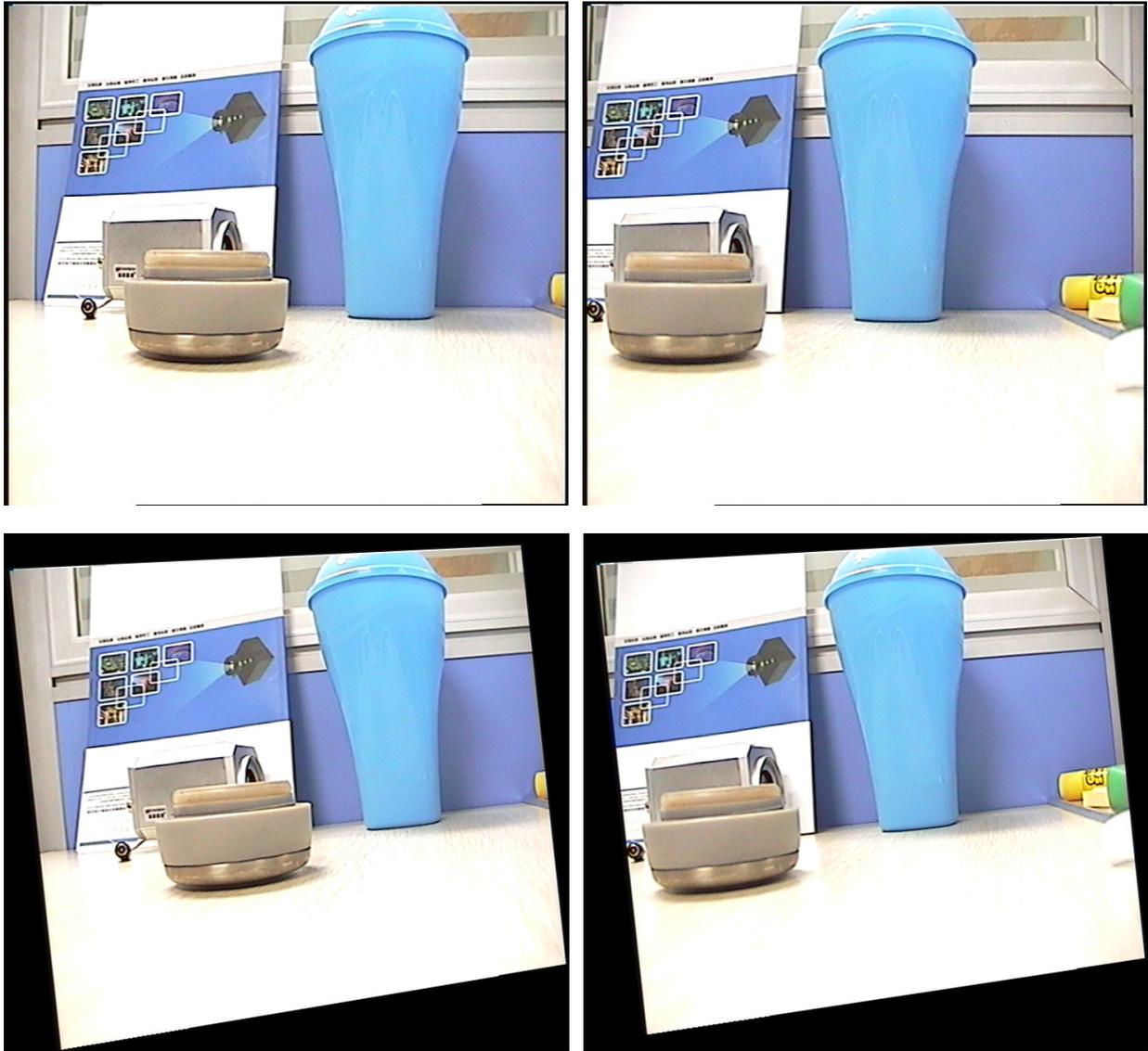


图 3-17 校正后的实验结果

图像对经过校正后，匹配点对搜索的维数从二维降低到了一维空间，大大降低了搜索的范围，减少了算法的计算时间，为后面的双目立体匹配打下了基础。

### 3.4 本章小结

在这一章中，我们对由双目立体视觉系统成像后的图像对进行校正。为了使得校正后的结果更加鲁棒、准确，本文提出了一种由粗到精的特征点匹配算法，首先我们提取两幅图像的 SIFT 特征点，根据特征点之间的相似性约束以及同一幅图像中特征点对之间边的兼容性建立了相应的数学模型。为了求解该模型，本文分两步对其求解，先用互相似度算法求相似性约束的局部最优解，然后利用改进遗传算法求空间约束的最优解，在求解过程中，本文将粗匹配的结果作为遗传算法的一个初始解，同时对遗传算法做了一些改进。

我们将本文提出的算法、Lowe's 方法以及 KD-Tree + RANSAC 算法在三个图片测试

集上进行了定量的比较，实验结果发现我们算法对于不同视角、不同光照、不同尺度下的宽基线图像的匹配精度更高一些。在得到更多、更准确的特征点对之后，我们用这些特征点对计算两个摄像机之间的相对位置关系：旋转矩阵和平移向量，最后给出了校正后的实验结果。



## 第4章 基于自适应置信传播算法的立体匹配

随着图像在三维重建、虚拟现实等领域的应用，人们对视差图的精度要求越来越高。在第二章我们已经提到，视差图的计算方法有局部和全局两种。利用局部的方法得到的视差图质量不高，而全局方法可以得到比较好的视差图，但是计算复杂度高一些，基本可以满足实际应用，因此受到了越来越广泛的重视。

本章的内容安排如下：第一节回顾了置信传播算法原理；第二节给出了置信传播算法目前存在的问题；第三节提出了自适应的置信传播算法；第四节给出了自适应置信传播算法与置信传播算法在标准测试集上的比较；第五节是本章小结。

### 4.1 置信传播算法原理

置信传播算法最早是由 Pearl<sup>[94][95][96]</sup>提出并广泛应用于机器视觉的各个领域，虽然在理论上置信传播算法对含环的图问题上不能收敛到问题的全局最优解，但是却取得了不错的实验效果。

置信传播算法是目前常用的求解立体匹配问题的方法之一。本节先给出由立体匹配问题构造出的能量函数模型。设  $I = \{I_L, I_R\}$  是校正后的双目立体图像对，立体匹配问题就是对左图  $I_L$  中的每个像素点  $p$ ，在右图  $I_R$  中找到与其相对应的像素点，这两个像素横坐标差的绝对值就是视差  $f = |x_L - x_R|$ 。也就是说，对图像  $I_L$  中的每个像素点，我们需要寻找一个标号函数  $f: P \rightarrow L$  使得下面的能量函数  $E(f, I)$  最小化：

$$E(f, I) = \sum_{p \in P} E_d(f_p) + \sum_{(p, q) \in N} E_s(f_p, f_q) \quad (4.1)$$

其中  $P$  是左图  $I_L$  中所有像素的集合，数据约束  $E_d(f_p)$  表示将标号  $f_p$  赋予像素点  $p$  的代价，光滑约束  $E_s(f_p, f_q)$  表示将标号  $f_p$  和  $f_q$  同时赋予两个相邻像素点的代价。数据约束模型和光滑约束模型有很多种类型，比如我们可以利用绝对值差和，平方差和等方法来计算数据约束，而对于光滑约束模型，有点模型，线性截断模型或者二次模型等。在本实验中，我们采用了 Birchfield 和 Tomasi 匹配代价  $C(f_p, I)$  来计算数据约束以及线性截断模型计算光滑约束。

$$E_d(f_p) = -\ln(1-e) \exp((-|C(f_p, I)|/\sigma) + e) \quad (4.2)$$

$$E_s(f_p, f_q) = \min(c \cdot |f_p - f_q|, d) \quad (4.3)$$

其中参数  $e$  和  $\sigma$  用来控制数据约束的形状， $f_p$  和  $f_q$  分别表示两个相邻像素点的视差。

为了使得能量函数 (4.1) 最小化，我们构造了马尔可夫随机场，立体匹配问题就是根据图像中的观察顶点值来推断图像中每个顶点的视差值。如图 4-1，黑色圆圈代表观察点值，白色圆圈表示要求的视差值。

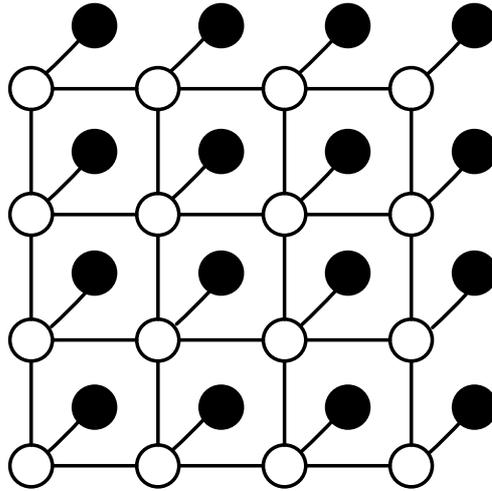


图 4-1 马尔可夫随机场模型

BP 算法是一个不断迭代的推理算法，主要通过消息的传递来求能量函数的最优解。图 4-2 定义了一个四连接的网格，变量  $m_{p \rightarrow q}(f_q)$  表示从隐藏顶点  $p$  到其相邻顶点  $q$  传播的消息，每个消息都是一个向量，其维数等于图像中每个像素点可能的标号，一般情况下，标号的范围是已知的，也就是说消息的维数是固定的。

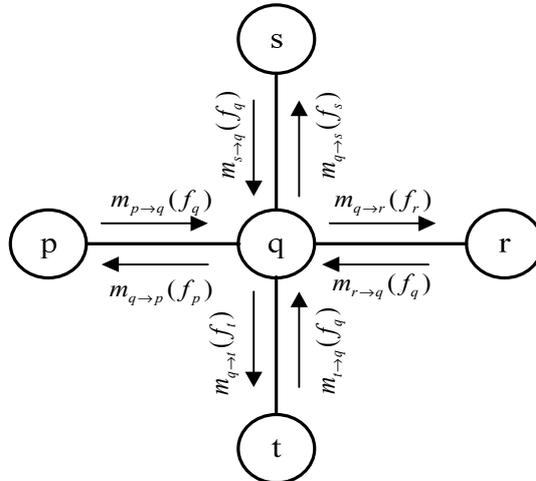


图 4-2 BP 算法中消息传播图

$p, q, r, s, t$  是相邻的顶点， $m_{p \rightarrow q}(f_q)$  表示从顶点  $p$  到顶点  $q$  的消息

假设  $b_q(f_q)$  表示 BP 算法中像素点  $q$  的置信度，则标准的“和-积”BP 算法描述为：

---

#### 算法 4.1 置信传播算法

---

- 1、初始化所有像素点的消息  $m_{p \rightarrow q}^0(f_q)$  为归一化分布；
- 2、当  $i=1:N$  时，迭代更新消息  $m_{p \rightarrow q}^i(f_q)$ ，其中：

$$m_{p \rightarrow q}^i(f_q) \leftarrow \kappa_1 \min_{f_p} (E_s(f_p, f_q) + E_d(f_p) + \sum_{s \in N(x_s) \setminus q} m_{s \rightarrow p}^{i-1}(f_p)) \quad (4.4)$$

- 3、计算置信度

$$b_q(f_q) \leftarrow \kappa_2 (E_d(f_q) + \sum_{p \in N(q)} m_{p \rightarrow q}^i(f_q)) \quad (4.5)$$

$$f_q = \arg \min_{f_q} b_q(f_q)$$

其中  $\kappa_1, \kappa_2$  是归一化常数

---

在上面的置信传播算法中，顶点  $q$  的消息是沿着上、下、右、左四个方向传播的，于是随着迭代次数  $i$  的增加，在点  $q$  将形成了四个方向的消息序列，分别表示为： $up\_message^i(q)$ 、 $down\_message^i(q)$ 、 $right\_message^i(q)$  和  $left\_message^i(q)$ ，且每个消息序列的长度都相同。

#### 4.2 置信传播算法目前存在的问题

虽然 BP 算法已经被广泛使用，然而还是存在着一些问题，该算法对含环图不能取得问题的全局最优解<sup>[97][98][99]</sup>。从算法 4.1 我们可以看出，置信传播算法的计算复杂度为  $O(N \cdot M^N \cdot S \cdot I)$ ，其中  $M$  是图像中每个顶点可能的状态或者标号的个数， $N$  是每个顶点的相邻顶点的个数， $S$  是图像的大小， $I$  表示置信传播算法迭代的次数。下面我们给出置信传播算法在标准数据集‘Tsukuba’，‘Venus’，‘Teddy’，‘Cones’上的运行时间情况，为了使得结果更具有代表性，我们统计了置信传播算法在每对图像上运行 5 次的结果并取均值，图 4-3 给出了随着迭代次数的增加，运行时间的变化图。

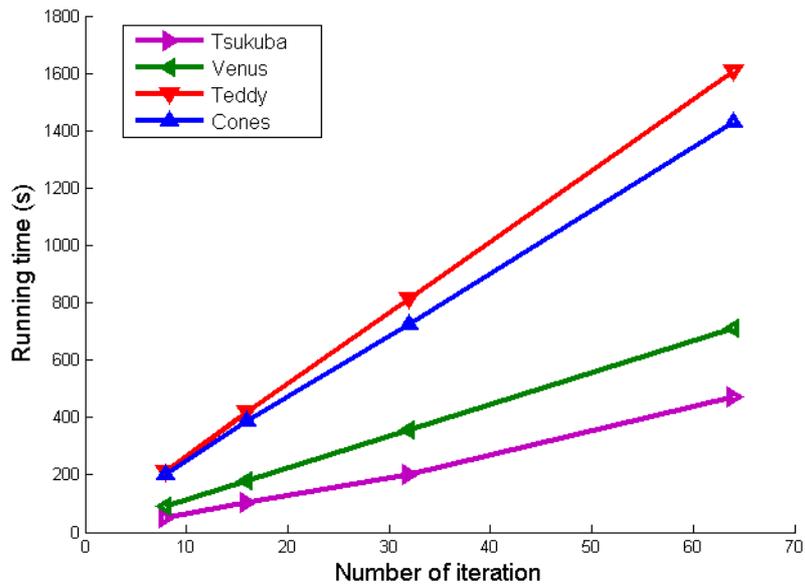


图 4-3 BP 算法迭代次数与运行时间变化图

从图 4-3 可以看出，随着 BP 算法迭代次数的增加，算法的运行时间呈线性增长的趋势。事实上，在 BP 算法运行的过程中，有计算冗余的现象。比如，随着迭代次数  $i$  的增加，图像中绝大多数顶点消息的改变量越来越小，甚至趋近于 0。另外，图像中每个顶点具有不同的纹理特征，它们需要不同的迭代次数收敛到问题的近似最优解。

对于第一个问题，若某个像素点的消息序列趋于收敛，那么用收敛时的消息值来代替当前迭代的消息值，就会省下重复计算消息值的时间。当然，若所有点的消息序列都收敛到一个固定的值，那么 BP 算法迭代就自动终止了

然而，仅仅考虑消息序列的收敛性是远远不够的。判断消息序列的收敛性可以加快算法的运行速度，但它可能使得算法的运行结果的正确率降低。主要原因在于图像中存在一部分点，它们的消息序列可能满足本文定义的收敛条件，但事实上他们并不收敛。通过实验发现这部分点位于图像的低纹理区域中。

基于这种情况，本文提出了图像的区域分类策略，先将图像进行分割并检测图像中的低纹理区域，对于低纹理区域的点，由于它们没有明显的灰度变化或者亮度变化，它们更多的依靠 BP 算法中光滑约束来找到最优解，而对于其他分割区域中的点，它们的灰度值或者亮度值变化比较明显，仅仅根据数据约束就可以找到最优解。这样的话，不仅仅可以解决收敛性判断带来的负面影响，而且可以提高视差图的质量。

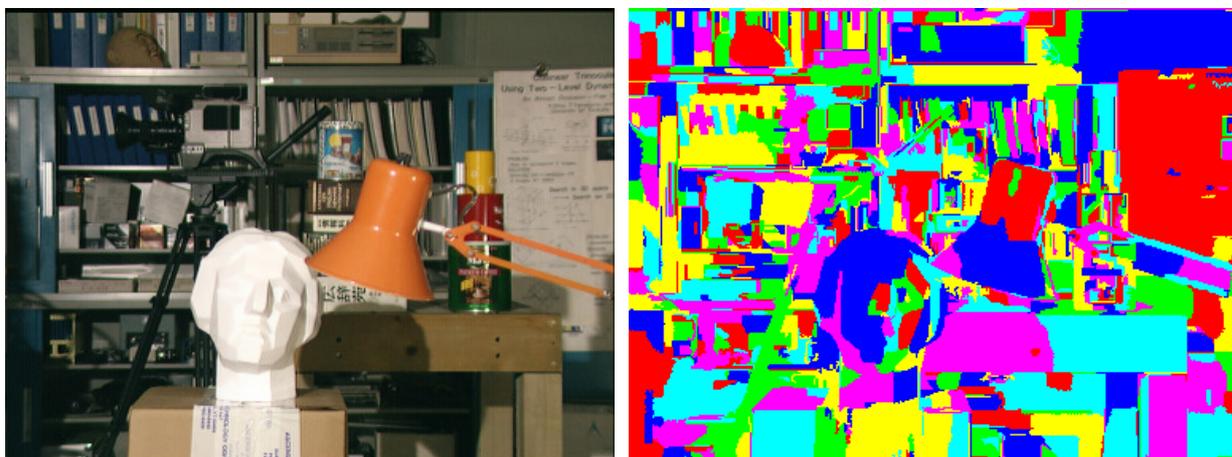
因此，本文给出了消息收敛的定义，同时提出了一种检测图像中低纹理区域的方法并将该方法运用到 BP 算法中，改进后的置信传播算法可以使得图像中的每个像素点自适应的收敛到问题的最优解，因此，本文称其为自适应置信传播算法（简称 ABP）。

### 4.3 自适应置信传播算法

在这一节中，先对图像进行彩色图像分割，然后利用灰度的方差以及视差的方差这两个约束条件来检测图像中的低纹理区域，最后提出了自适应置信传播算法。

#### 4.3.1 图像分割

设图像对中的左图为参考图像，利用均值漂移算法<sup>[99]</sup>对该图像进行图像分割。分割后，颜色相同的在同一个区域，为了确保在每个分割区域中视差具有连续性，本文采用了过分割的方法。图 4-4 给出了图像 Tsukuba 及其分割后的结果。



(a) 图像 Tsukuba

(b) 图像 Tsukuba 的分割结果

图 4-4 图像分割

#### 4.3.2 低纹理区域检测

一般来说，低纹理区域中的像素点没有明显的颜色或者灰度变化，换句话说，低纹理区域中像素的灰度值方差趋近于 0，因此仅仅根据颜色或者灰度值是很难在另一幅图像中找到其匹配点的。另外，由于低纹理区域中像素颜色或者灰度变化不明显，因此，这些像素点视差很不稳定，也就是说，低纹理区域中像素的视差方差比高纹理区域中像素的视差方差大。根据这两个特性，本文提出了一种低纹理区域的检测方法。定义检测函数  $F$ ：

$$F(R_s) = \frac{\text{Var}(R_{\text{grayscale}})}{\text{Var}(R_{\text{disparity}}) + \delta} \quad (4.6)$$

其中  $R_s$  表示图像中第  $s$  个分割区域， $\text{Var}(R_{\text{grayscale}})$  表示灰度值的方差， $\text{Var}(R_{\text{disparity}})$  表示视差值的方差， $\delta$  是一个常数。为了防止检测函数  $F$  的分母为 0，在实验中，取  $\delta=0.01$ 。根据上面的函数，我们计算每个区域对应的  $F$  值， $F$  值越小，该区域的低纹理程度就越强。图

4-5 给出了图‘teddy’的低纹理区域检测结果（阈值分别取 2, 4, 6, 8），白色区域是低纹理区域。

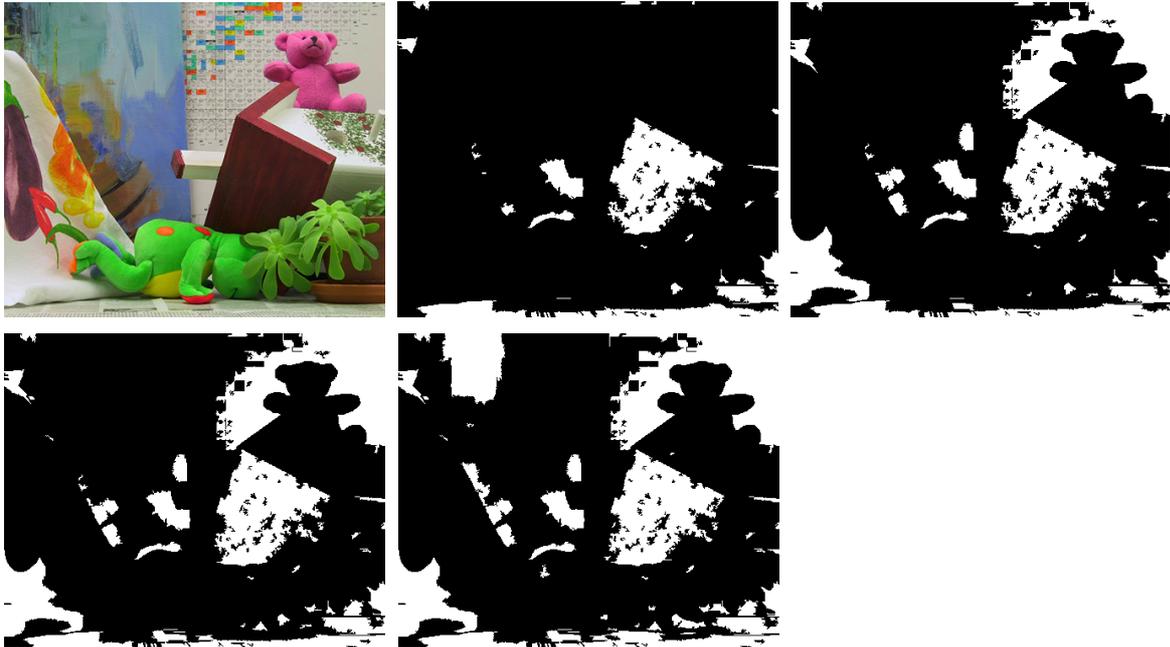


图 4-5 低纹理区域检测图

### 4.3.3 自适应置信传播算法

对于 BP 算法而言，大量的迭代次数是很必要的，在 Sun 论文<sup>[31]</sup>中指出，当迭代次数等于 64 时，实验结果比较稳定而且效果好。然而，我们发现，随着迭代次数的增加，图像中的很多顶点出现了过收敛的现象。为了减少 BP 算法的运行时间的同时提高算法的精度，我们希望图像中的每个顶点都能尽可能的找到一个最优的迭代次数。

为了保证低纹理区域中每个像素点的迭代次数，同时提高高纹理区域中像素点视差的精度，本文给出了 BP 算法中消息收敛的定义：

假定图像中顶点  $p$  的消息序列为  $\{m_p^i\}$ ,  $i=1,2,\dots,k$ ，如果该序列  $\{m_p^i\}$  是单调的且随着迭代次数  $i$  的增加，消息的变化量  $\{|m_p^i - m_p^{i-1}|\}$  趋近于 0，那么我们称消息序列  $\{m_p^i\}$  是收敛的。

在置信传播算法中，消息序列是有限序列。在上面的收敛定义中，如果每个有限序列能最大程度上满足数学上的无限序列的收敛条件，那么我们就认为这个消息序列是收敛的。

显然，如果图像中某一顶点沿着某一方向的消息序列是收敛的，那么这个顶点沿着这个方向在下一次迭代中就不需要再计算消息了。换句话说，在后面的消息传播过程中，该消息用收敛后的消息来代替就可以了，这样就可以节省很多的计算时间。下面给出了自适应置信传播算法：

## 算法 4.2 自适应置信传播算法

- 1、利用均值漂移算法对参考图像进行分割；
- 2、初始化所有的消息  $m_{p \rightarrow q}^0(f_q)$  为归一化分布；
- 3、当  $i=1:k$  时，迭代更新消息  $m_{p \rightarrow q}^i(f_q)$ ，其中：

$$m_{p \rightarrow q}^i(f_q) \leftarrow \kappa_1 \min_{f_p} (E_s(f_p, f_q) + E_d(f_p) + \sum_{s \in N(x_s) \setminus q} m_{s \rightarrow p}^{i-1}(f_p)) \quad (4.7)$$

计算置信度：

$$b_q(f_q) \leftarrow \kappa_2 (E_d(f_q) + \sum_{p \in N(q)} m_{p \rightarrow q}^i(f_q)) \quad (4.8)$$

$$f_q = \arg \min_{f_q} b_q(f_q)$$

- 4、检测参考图像中的低纹理区域  $R$ ；
- 5、当  $i=k+1:N$  时，若消息序列  $\{m_{p \rightarrow q}^i(f_q)\}$  是不收敛的或者顶点  $q \notin R$

迭代更新消息  $m_{p \rightarrow q}^i(f_q)$

$$m_{p \rightarrow q}^i(f_q) \leftarrow \kappa_1 \min_{f_p} (E_s(f_p, f_q) + E_d(f_p) + \sum_{s \in N(x_s) \setminus q} m_{s \rightarrow p}^{i-1}(f_p)) \quad (4.9)$$

否则

$$m_{p \rightarrow q}^i(f_q) \leftarrow m_{p \rightarrow q}^{i-1}(f_q) \quad (4.10)$$

- 6、计算置信度：

$$b_q(f_q) \leftarrow \kappa_2 (E_d(f_q) + \sum_{p \in N(q)} m_{p \rightarrow q}^i(f_q)) \quad (4.11)$$

$$f_q = \arg \min_{f_q} b_q(f_q)$$

其中  $\kappa_1, \kappa_2$  是归一化常数

在上面的自适应置信传播算法中，对图像中的每个像素点  $q$ ，若其处于低纹理区域，那么在每次的迭代过程中都要用等式(4.9)计算新消息，换句话说，在每次的迭代过程中都要进行光滑约束。同样，若点  $q$  沿着某一个方向的消息序列不收敛，那么同样需要在点  $q$  重新计算消息。而对于图像中的收敛像素点，本文都是利用收敛点的消息来代替下一次迭代的消息，这样就可以减少运行时间。在实验中，收敛点的消息是用最后一次迭代的消息来

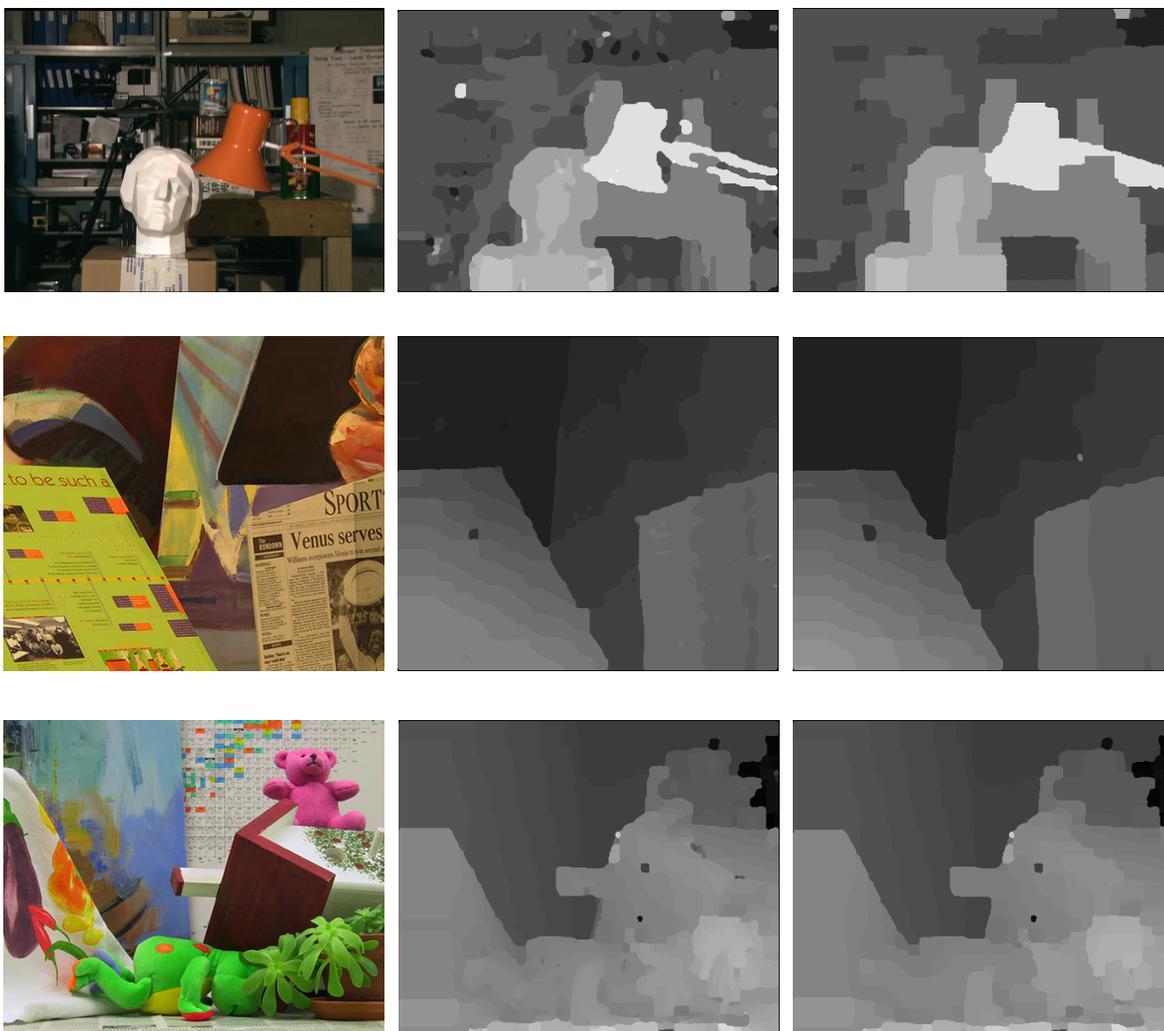
代替的。

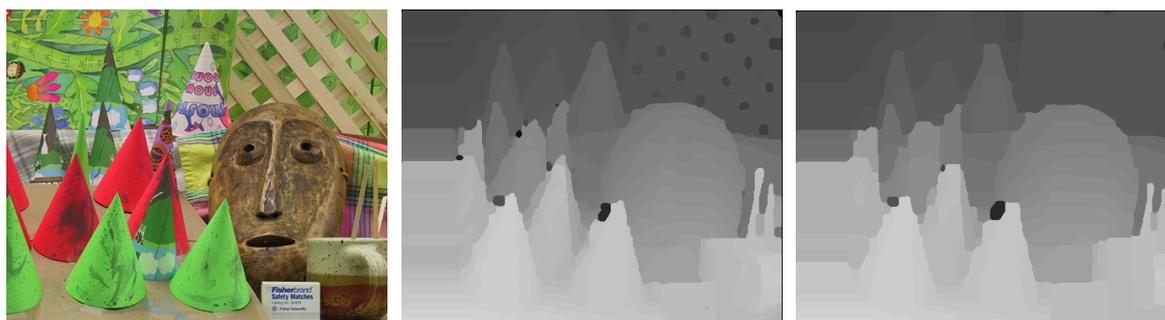
#### 4.4 实验结果和比较

在这一节中，我们分别从精度和计算速度两个方面分别给出了 ABP 算法和 BP 算法在数据集 ‘Tsukuba’， ‘Venus’， ‘Teddy’ 和 ‘Cones’ 上的比较。这些数据集来自 <http://vision.middlebury.edu/stereo/data/>。

##### 4.4.1 精度比较

我们先给出了 ABP 算法和 BP 算法在精度方面的比较，如图 4-6，第一列是参考图像，第二列是 ABP 算法的实验结果，第三列是 BP 算法的实验结果。





(a) 参考图像

(b) ABP 算法的结果

(c) BP 算法的结果

图 4-6 在 Middlebury 数据集上的结果

从上面的实验结果我们可以看出，对这四副图像对，ABP 算法的结果明显优于 BP 算法的结果。尤其是对于图像中的边缘区域和图像中的高纹理区域，ABP 算法的结果明显好于 BP 算法的结果。

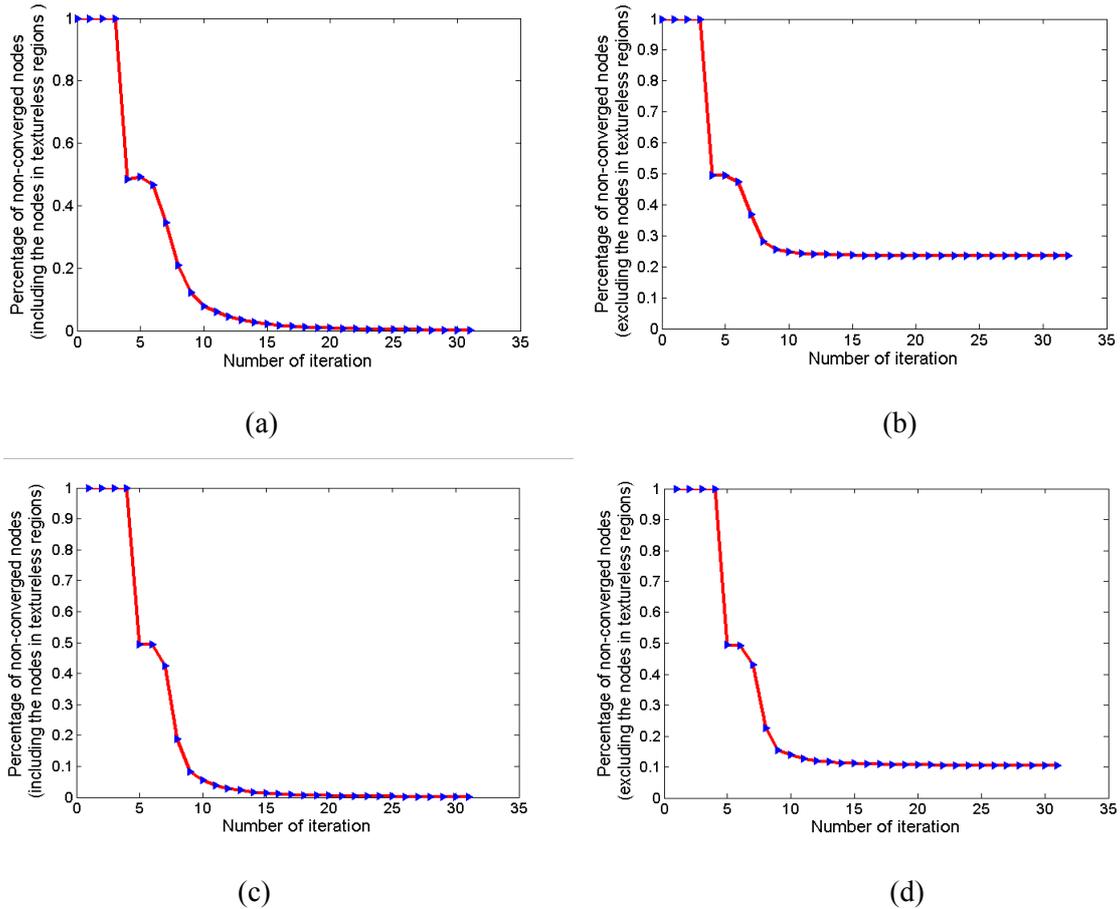
为了更精确的比较 ABP 算法与 BP 算法的结果，本文对上面得到的视差图与标准的视差图进行了比较，主要是衡量两种算法条件下错误匹配像素的百分比（这里的错误像素是指比真实结果的视差值大于 1）。主要从以下三个方面来衡量：（1）非遮挡（非遮挡区域中的像素点）；（2）全部（非遮挡区域和半遮挡区域中的像素点）；（3）不连续（接近遮挡区域的可视像素点）。表 4-1 给出了 ABP 算法和 BP 算法实验结果的定量比较。从中可以看出，ABP 算法的实验结果中不连续区域像素点的错误远远小于 BP 算法的实验结果，而且 ABP 算法在另外两个因素的衡量结果也比 BP 算法结果的精度高。

表 4-1 ABP 算法和 BP 算法的精确度比较

图像对 算法	Tsukuba			Venus			Teddy			Cones		
	nonocc	all	disc	Nonocc	all	disc	nonocc	all	disc	Nonocc	all	disc
BP 算法	5.22	7.24	22.1	3.14	4.34	20.4	16.1	24.4	34.0	8.55	16.6	21.2
ABP 算法	4.40	6.52	12.7	1.66	2.58	17.6	15.4	23.8	31.1	6.54	14.6	16.6

#### 4.4.2 性能比较

本文先给出了 ABP 算法的收敛图。图 4.7 给出了图像‘Venus’和‘Cones’图像中非收敛像素变化的百分比。从图像中可以看出，随着迭代次数的增加，不收敛的像素点在减少。当迭代次数等于 5 的时候，大约有 50%的像素已经收敛，当迭代次数等于 20 的时候，不收敛的像素点已经趋于稳定。



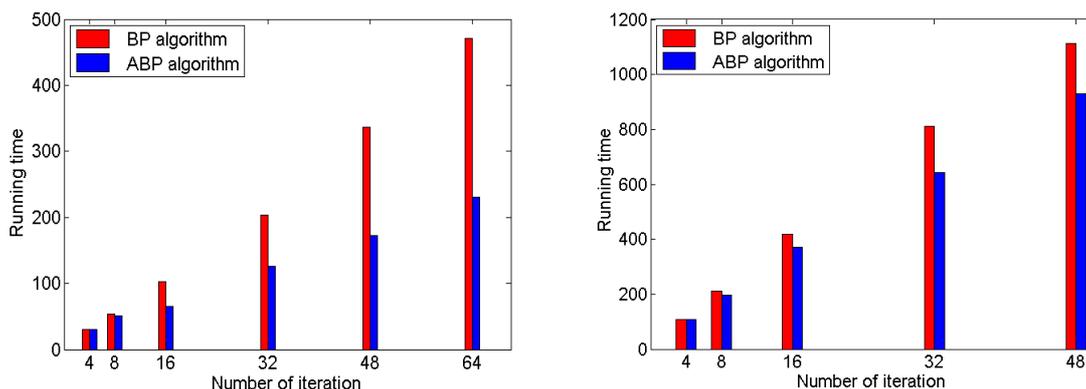
(a) 图‘Venus’中，随着迭代次数增加不收敛像素（包含低纹理区域中顶点）的变化百分比；(b) 图‘Venus’中，随着迭代次数增加不收敛像素（不包含低纹理区域中顶点）的变化百分比；(c) 图‘Cones’中，随着迭代次数增加不收敛像素（包含低纹理区域中顶点）的变化百分比；(d) 图‘Cones’中，随着迭代次数增加不收敛像素（不包含低纹理区域中顶点）的变化百分比

图 4-7 随着迭代次数增加不收敛像素所占比例的变化

图 4-8 给出了 BP 算法和 ABP 算法在标准图像‘Tsukuba’、‘Teddy’运行时间的比较，其迭代次数分别为 4, 8, 16, 32, 64。为了使得计算结果更准确、鲁棒，我们统计了每幅图像在相同情况下运行 5 次的结果并取平均值。

图 4-8 (a) 是在图像‘Tsukuba’上的运行结果，可以看出，当迭代次数等于 32 的时候，ABP 算法的运行时间大约是 BP 算法运行时间的一半。

图 4-8 (b) 是在图像‘Teddy’上的运行结果，可以看出，随着迭代次数的增加，ABP 算法的运行时间和 BP 算法的运行时间相比只减少将近 20%。这是因为图像‘Teddy’中低纹理区域比较多，在每次迭代中，我们需要利用等式 (4.9) 去计算消息而不是直接用收敛后的结果代替。



(a) ABP 算法和 BP 算法在图像‘Tsukuba’上的运行时间比较；(b) ABP 算法和 BP 算法在图像‘Teddy’上的运行时间比较

图 4-8 ABP 算法和 BP 算法迭代时间的比较

表 4-2 给出了 ABP 算法和 BP 算法在四个图像对‘Tsukuba’，‘Venus’，‘Teddy’和‘Cones’上的运行时间比较。为了更能代表算法的统计独立性，本文对其运行了 5 次然后取平均。从下表可以看出，和 BP 算法相比，本文提出的 ABP 算法在标准测试图像上大约节省了将近 50%的运行时间。当然，运行时间的多少与图像中低纹理区域的多少有直接的关系，若图像中的低纹理区域越多，运行时间就越少，反之运行时间就越长。

表 4-2 ABP 算法和 BP 算法的运行时间比较

时间(s) 算法	图像	Tsukuba	Venus	Teddy	Cones
	BP 算法		471	1606	1113
ABP 算法		231	525	931	657

## 4.5 本章小结

置信传播算法是立体匹配算法中常用的一种算法，由于该算法存在计算冗余以及计算复杂度高问题，我们在本章中提出了一种自适应置信传播算法。

自适应置信传播算法是置信传播算法的一个改进，也是建立在全局能量最小化框架下求最优解的一种方法。在这个算法中，为了使得图像中的每个点都能尽可能的收敛到其最优解，本文给出了有限消息序列收敛的定义，同时提出了一种低纹理区域检测的方法并将这两个思想加入到置信传播算法中。ABP 算法不仅可以减少算法的计算冗余，而且能使得图像中各种类型的点都近似收敛到问题的全局最优解

我们将自适应置信传播算法与置信传播算法在标准数据集上进行了比较。该算法对立体视觉问题有两个比较大的贡献，其一是自适应置信传播算法的计算时间比置信传播算法的计算时间减少了大约 50%；另一个贡献是自适应置信传播算法的计算精度高于传统的置

信传播算法的计算精度。

## 第5章 基于分割的全局立体匹配

本章给出基于分割的全局立体匹配算法。整个算法的计算过程分为三个阶段：(1) 初始匹配；(2) 视差校正；(3) 视差优化。在视差校正阶段，本文将初始视差图中由低纹理引起的误匹配像素和遮挡像素进行分类并分别对其校正；在视差优化阶段，本文提出了一个渐进的区域融合算法，该算法先融合大区域和置信度高的区域，增加了平面拟合的可靠度。图 5-1 给出了基于分割的全局立体匹配算法的流程图。

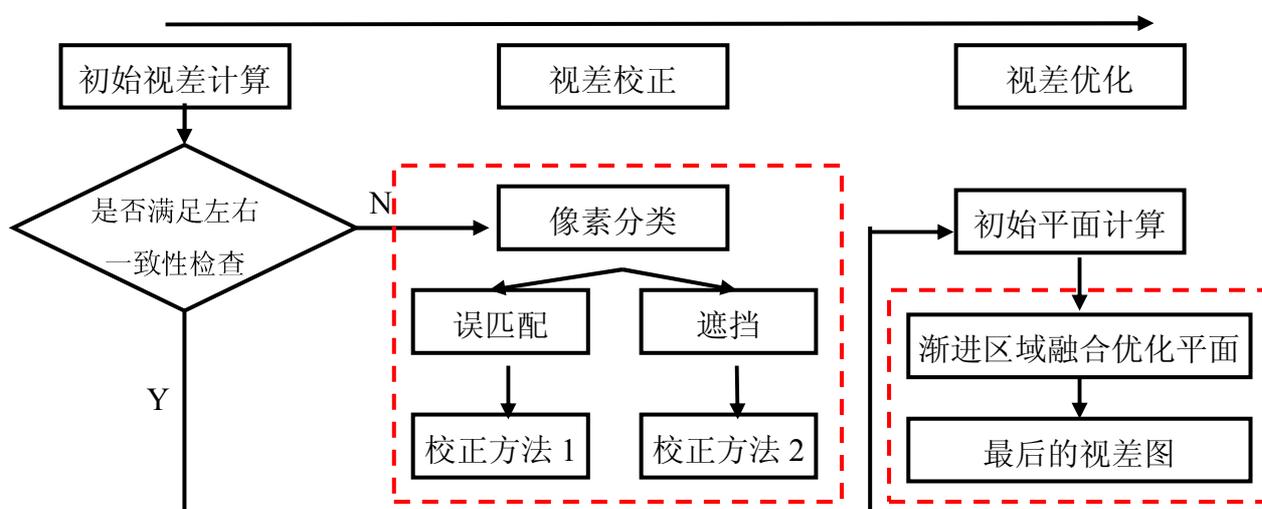


图 5-1 基于分割的全局立体匹配算法流程图

本章是这样安排的，5.1 节计算了初始视差图，5.2 节是视差校正部分，5.3 节提出了视差优化算法，算法的实验效果和比较在 5.4 节，最后是本章的小结部分。

### 5.1 初始视差计算

在这一节中，我们先利用置信传播算法对校正后的左右图像对计算初始视差图。计算过程中，仍然采用了 Birchfield 和 Tomasi 匹配代价下的数据约束和线性截断的光滑约束模型，实验过程中使用的参数设置和上一章是一致的。由于随着迭代次数的增加，BP 算法的计算时间线性增长，在本节中，我们设置 BP 算法的迭代次数是 8。图 5-2 给出初始视差图的实验结果。



(a) 原图

(b) 初始视差图

图 5-2 初始视差图

## 5.2 视差校正

在利用 BP 算法计算左右初始视差图之后, 本文提出了一个视差校正方法来校正初始视差图中的错误像素, 这些错误视差的像素可能是由低纹理、重复纹理、遮挡或者噪声等原因引起的。在整个算法中, 本文先利用左右一致性检验方法来检测初始视差图中的错误视差的像素, 然后提出了一种新的像素的分类方法来判断这些错误是由误匹配引起的还是由遮挡引起的, 最后本文对这两类错误视差分别校正。

### 5.2.1 左右一致性检查

设  $D_l$  和  $D_r$  分别是由置信传播算法得到的左、右初始视差图, 左右一致性检查方法要求左右视差图的视差值是一致的, 也就是说, 对左图  $I_l$  的任意一个像素  $p=(x_l, y)$ , 下面

的等式是成立的：

$$D_L(x_L, y) = D_R(x_L - D_L(x_L, y), y) \quad (5.1)$$

对于左图的每一个像素  $p = (x_L, y)$ ，若等式 (5.1) 成立，则说明该像素满足左右一致性检查，否则，我们设该点的视差值为 0。

$$D(x_L, y) = \begin{cases} D_L(x_L, y) & \text{if } D_L(x_L, y) = D_R(x_L - D_L(x_L, y), y) \\ 0 & \text{otherwise} \end{cases} \quad (5.2)$$

图 5-3 第二列给出了经过左右一致性检查后的视差图，我们称其为初始可靠的视差图，从图中可以看出，绝大多数的像素还是满足左右一致性检查的，且满足左右一致性检查的绝大部分像素的视差是可靠的。

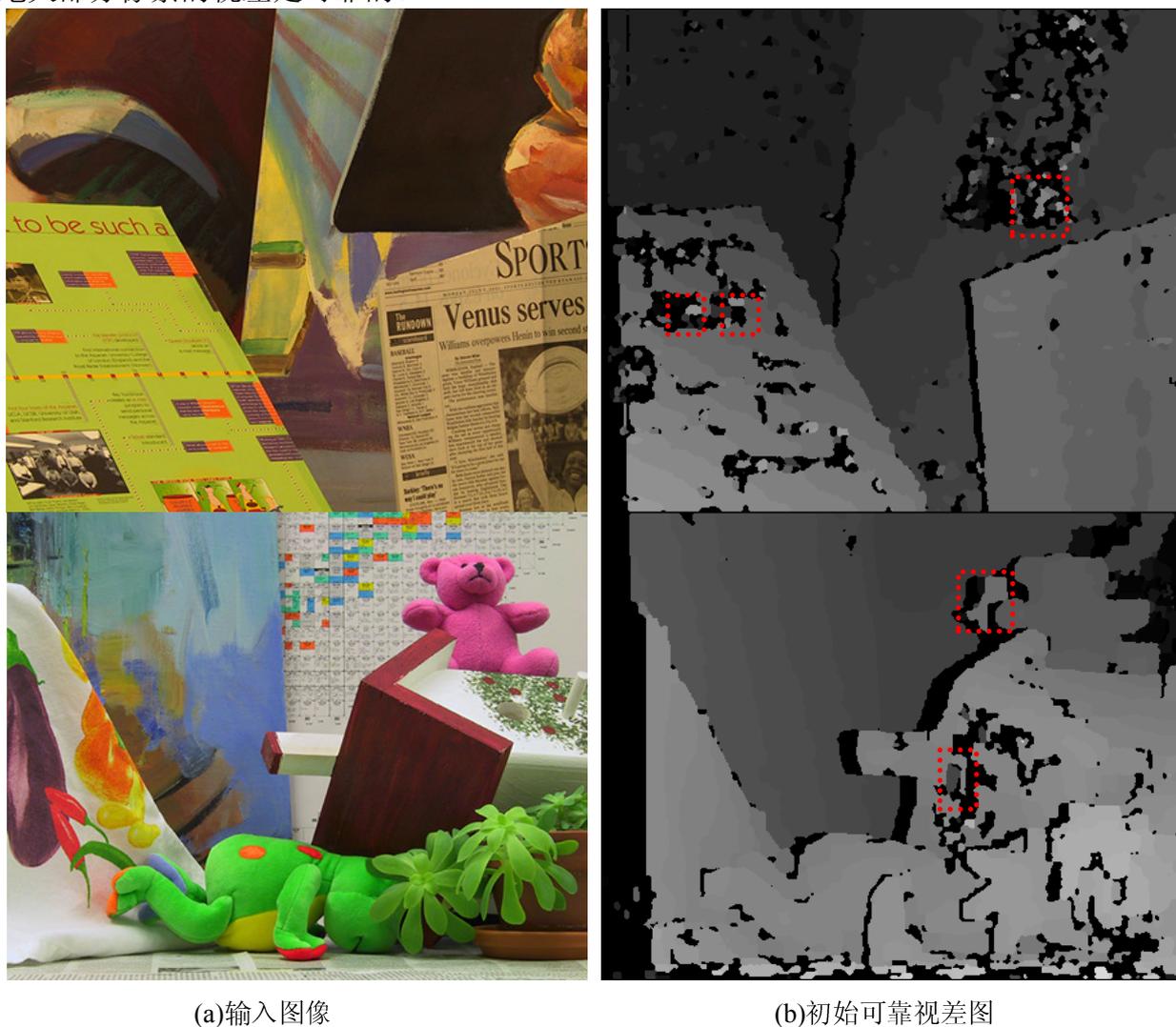


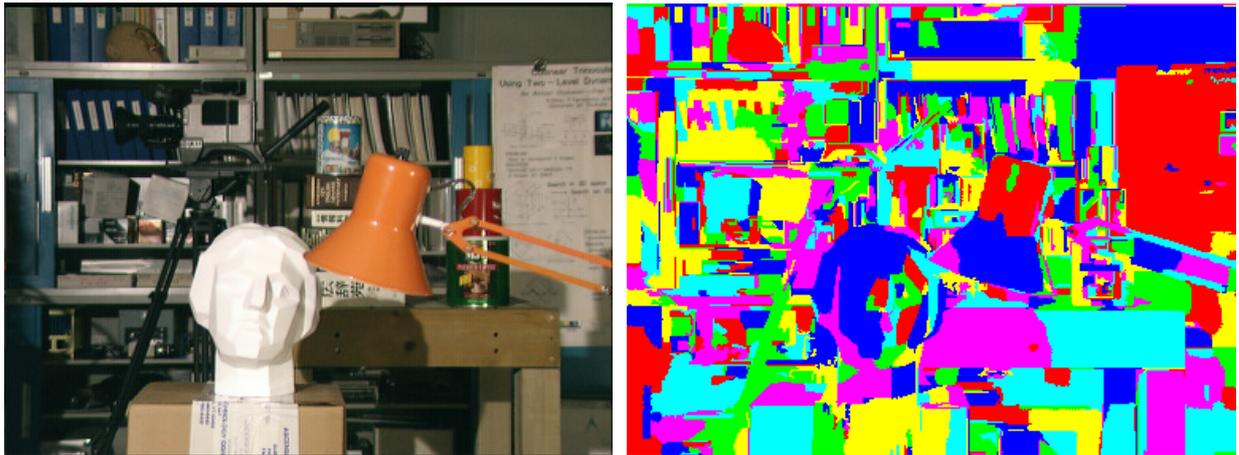
图 5-3 初始可靠像素的视差图

然而，还有一些像素，它们满足左右一致性检查，但是它们对应的初始视差值是错误

的，这一点可以从图 5-3 中可以看出，比如，上图中红色线框中的像素。幸运的是，这些对应着错误视差值的像素大都位于图像的低纹理区域，为了得到更加精确的视差图，我们先检测这些错误视差值对应的像素。

### 5.2.2 图像分割

由于这些错误视差值对应的像素位于图像的低纹理区域，于是本文先用均值漂移对图像进行分割，如图 5-4。



(a) Tsukuba

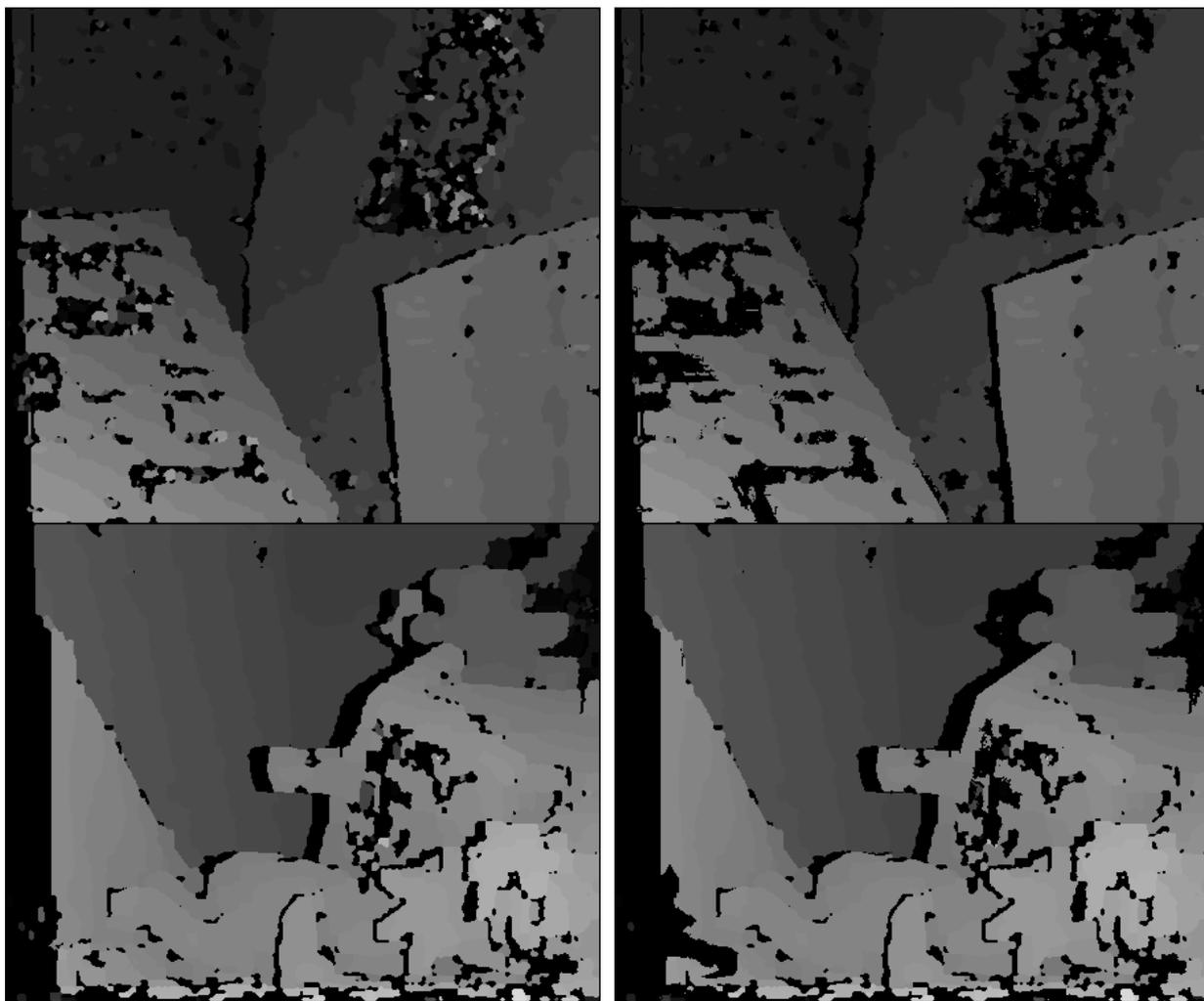
(b) Tsukuba 分割结果

图 5-4 图像分割

### 5.2.3 低纹理区域检测和误匹配像素检测

在本节中，我们还是用上一章中的低纹理检测方法来检测图像中的低纹理区域在得到低纹理区域后，本文将每一个低纹理区域当成一个前置的平面或者说是窗口，利用“WTA”策略来计算这些窗口的平均视差  $d'$ ，为了避免图像噪声的影响，我们在计算匹配代价的时候利用了高斯滤波器，其中带宽为 5，方差为 1。于是，对每一个像素  $(x, y) \in R_s$ ，如果  $|d(x, y) - d'| < th$ ，那么这个像素被认为是可靠像素，保持其原有的视差值不变，否则，该像素就是由低纹理区域引起的误匹配像素，并设该像素的视差值为 0。在本实验中， $th = 3$ 。

图 5-5 的第一列给出了经左右一致性检测后的初始可靠视差图，第二列是在初始可靠视差图的基础上经误匹配像素检测后的实验结果。对比实验结果可以看出，经误匹配像素检测算法后，初始可靠视差图中的剩余像素都是可靠像素，我们把此时得到的视差图记为可靠视差图。



(a) 初始可靠视差图

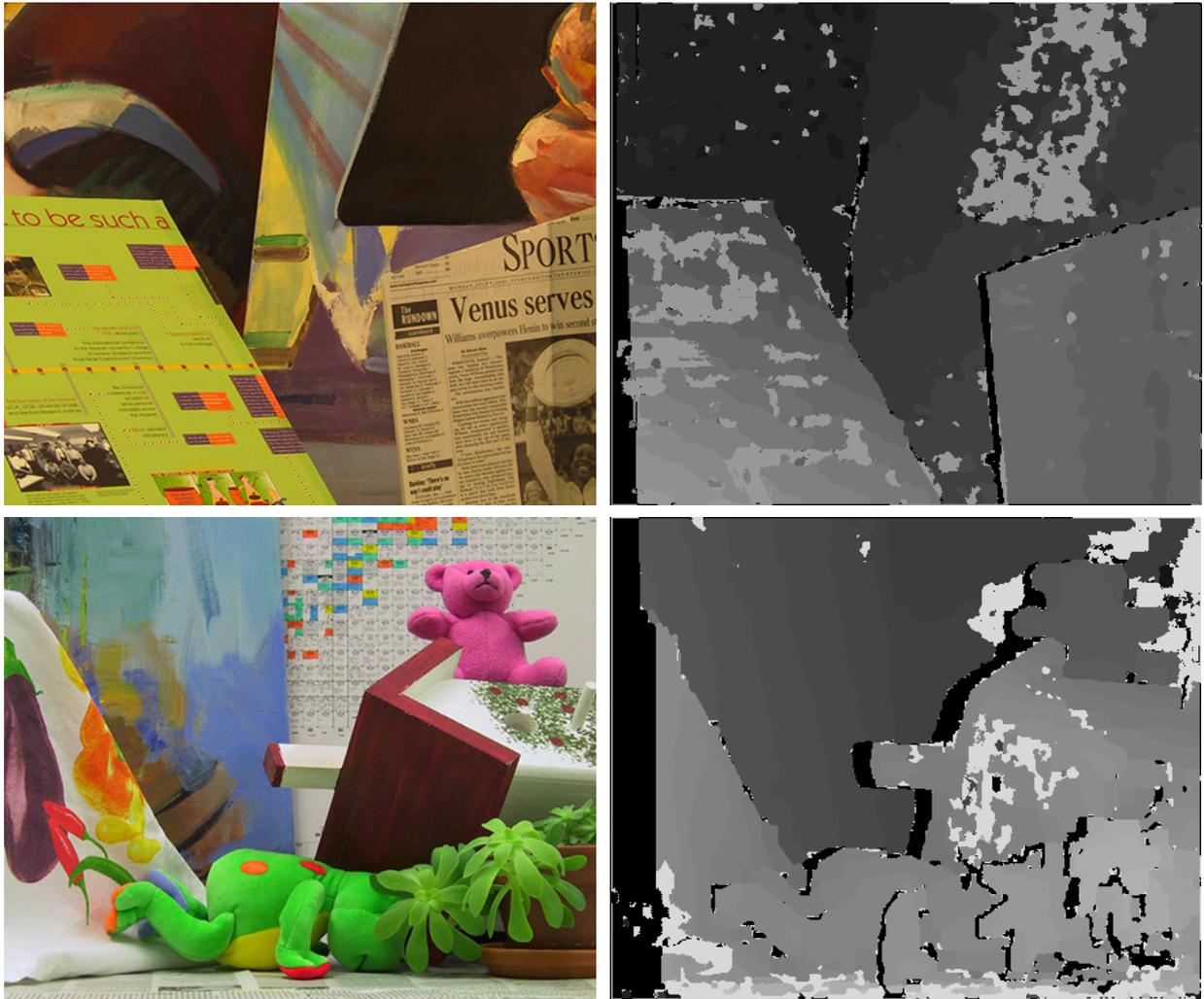
(b) 可靠视差图

图 5-5 视差图

#### 5.2.4 像素分类

对于立体图像对来说，遮挡像素一般在其中一幅图像中是可见的，在另一幅图像中是不可见的，而对于误匹配像素来说，它经常发生在图像的低纹理区域，于是为了区别遮挡像素和误匹配像素，我们采用了最简单的分类方法：对于图 5-5 中的每个错误视差的像素，也就是图像中视差值等于 0 对应的像素，若其位于图像中的低纹理区域，那么该像素被认为是误匹配的，否则，该像素属于遮挡像素。

图 5-6 的第二列给出了像素分类的结果，其中的白色像素表示误匹配，黑色像素表示遮挡。



(a)输入图像

(b)像素分类图

图 5-6 像素分类图

### 5.2.5 视差校正

将遮挡像素和误匹配像素进行分类后，本文利用 Hirschmüller 的方法对这两类像素进行校正，校正方法如下：

$$d'_p = \begin{cases} \text{second low } v_p & \text{若 } p \text{ 是遮挡像素} \\ \text{median } v_p & \text{若 } p \text{ 是误匹配像素} \\ d_p & \text{其他} \end{cases} \quad (5.3)$$

对每一个像素  $p$ ， $v_p$  是一个视差的集合，它包含像素  $p$  周围的 8 个方向对应的视差值。在等式 5.3 中，如果像素  $p$  的错误视差是由遮挡引起的，那么我们就用集合  $v_p$  中第二个最小的视差值来代替；如果像素  $p$  的错误视差是由低纹理引起的，那么我们就用集合  $v_p$  的均值来代替像素  $p$  的视差。利用这个校正方法，我们得到了校正后的视差图，见图 5-7 中的

第二列。



(a)输入图像

(b)校正后的视差图

图 5-7 校正后的视差图

### 5.3 视差优化

在这一节中，我们对校正后的视差图进行优化。优化的过程主要分为两个阶段：（1）利用 Ransac 方法计算每个分割区域的初始视差平面；（2）利用一种提出的渐进的区域融合算法优化初始视差平面。

#### 5.3.1 计算初始视差图

在得到校正后的初始视差图之后，我们将每一个分割区域  $R_s$  模拟成一个 3D 平面，表示为：

$$d = a_i x + b_i y + c_i \quad (5.4)$$

其中  $d$  表示像素  $(x, y) \in R_s$  的视差值。对每个分割区域，我们用 Ransac 算法求出其初始的视差平面参数  $[a_i, b_i, c_i]$ ，并比较拟合前后每个像素视差值的变化  $D_\Delta(x, y)$ ，除了小的分割区域（这里是指分割区域中像素的个数少于 100 个）：

$$D_\Delta(x, y) = |D_{before}(x, y) - D_{after}(x, y)| \quad (5.5)$$

如果  $D_\Delta(x, y) < 0.5$ ，那么该像素  $(x, y)$  是一个可靠像素。然后我们利用每个区域中的这些可靠像素去优化视差平面。

### 5.3.2 优化视差平面

在平面拟合的过程中，参与拟合的可靠像素越多，拟合后得到的平面参数将会越准确。由于我们在图像分割过程中采用了过分割技术，也就是说很多的大区域被分割成小区域，因此，如果这些小区域来自相同的分割区域，我们必须将这些小的区域合并成一个大的区域，那么用更多的可靠像素去拟合平面是合理的。

假设  $R_s$  和  $R_t$  是任意两个分割区域， $P_s$  和  $P_t$  是其相对应的平面参数，那么平面方程为：

$$\begin{aligned} P_s: \quad d_1(x, y) &= a_1x + b_1y + c_1 \\ P_t: \quad d_2(x, y) &= a_2x + b_2y + c_2 \end{aligned} \quad (5.6)$$

为了判断两个区域  $R_s$  和  $R_t$  是否来自相同的分割区域，本文通过判断与其对应的平面  $P_s$  和  $P_t$  是否相同来确定。主要有以下两个方面：

#### 1) 两个平面的夹角 $\alpha$

用两个平面  $P_s$  和  $P_t$  对应的法向量之间的夹角来计算  $\alpha$  的值。公式如下：

$$\cos \alpha = \frac{P_s \cdot P_t}{|P_s| \cdot |P_t|} = \frac{a_1a_2 + b_1b_2 + c_1c_2}{\sqrt{a_1^2 + b_1^2 + c_1^2} \cdot \sqrt{a_2^2 + b_2^2 + c_2^2}} \quad (5.7)$$

若夹角  $\alpha$  很小，那么我们认为这两个平面是近似平行的。

#### 2) 两个平面的空间距离

如果两个平面  $P_s$  和  $P_t$  是近似平行的，那么我们认为  $|c_1 - c_2|$  是这两个平面对应的空间距离。若  $|c_1 - c_2| < 1$ ，则这两个平面是近似相同的。也就是说，区域  $R_s$  和  $R_t$  是否来自相同的分割区域。

此外，在每个分割区域中，可靠像素的比例也是一个很重要的因素。定义  $m$  为分割区

域的置信系数:

$$m = N_r / N \quad (5.8)$$

其中  $N_r$  是区域中可靠像素的个数,  $N$  是区域中所有像素的个数。显然, 可靠像素越多, 置信系数会越大。

在给出了上述定义后, 本文提出了渐进的区域融合算法, 在该算法中, 我们首先选择置信度大、区域像素点多的区域作为种子区域, 利用这些区域一步一步去融合与其相邻的小区域, 并将该区域与其相邻的区域进行合并。下面给出了渐进的区域融合算法的具体步骤:

---

#### 算法 5.1: 渐进的区域融合算法

---

- 1、根据分割区域的大小降序排列平面参数  $P$ ;
- 2、顺序选择平面  $P_s$ , 若  $P_s = \phi$ , **break**; 否则, 计算平面  $P_s$  对应区域  $R_s$  的置信系数  $m$ , 若  $m$  大于一个给定的阈值  $T_1$  ( $T_1=0.6$ ), **continue**; 否则, 返回第二步;
- 3、对已经选定的种子区域平面  $P_s$

选择其相邻的区域  $R_t$ , 其平面记为  $P_t$

- (a) 比较平面  $P_s$  和  $P_t$  是否相同

若相同, 将  $R_t$  合并到  $R_s$  中并计算新区域的平面参数, 返回到第一步;

否则, **continue**;

- (b) 检查是否所有相邻的区域都已经访问过

若是, 将该平面移出并标记;

否则, **continue**;

- 4、利用式 (5.5) 计算最后的视差图。
- 

在渐进的区域融合算法中, 我们充分考虑了相邻分割区域之间的关系, 并将大区域、高置信系数的区域作为种子区域, 这样的话, 合并后的平面区域更加准确、鲁棒。这个算法有两个主要的贡献: 一是我们选择大的、置信系数高的区域作为种子区域, 可以大大降低平面拟合中的错误率; 二是我们利用了角度和距离来判断两个区域对应的平面是否相同, 与其他的计算匹配代价相比, 降低了算法的计算复杂度。

## 5.4 实验结果与比较

在本节中，我们对本章提出的算法在 Middlebury 的测试集‘Tsukuba’，‘Venus’，‘Teddy’和‘Cone’进行了详细评测，并将实验结果与其他算法的实验结果进行了比较。我们首先给出由本文算法计算得到的遮挡图的比较结果，然后给出本文算法和其他算法得到的视差图之间的定量比较。

### 5.4.1 遮挡图的定量比较

为了与真实遮挡图进行定量的比较，我们从两个方面：（1）精确度；（2）误检率对遮挡图进行了比较，其中精度和误检率的定义如下：

$$\text{Precision Rate} = \frac{\text{Number of detected occluded pixels}}{\text{Number of true occluded pixels}}$$

$$\text{False Accept Rate} = \frac{\text{Number of detected non-occluded pixels}}{\text{Number of all detected pixels}}$$

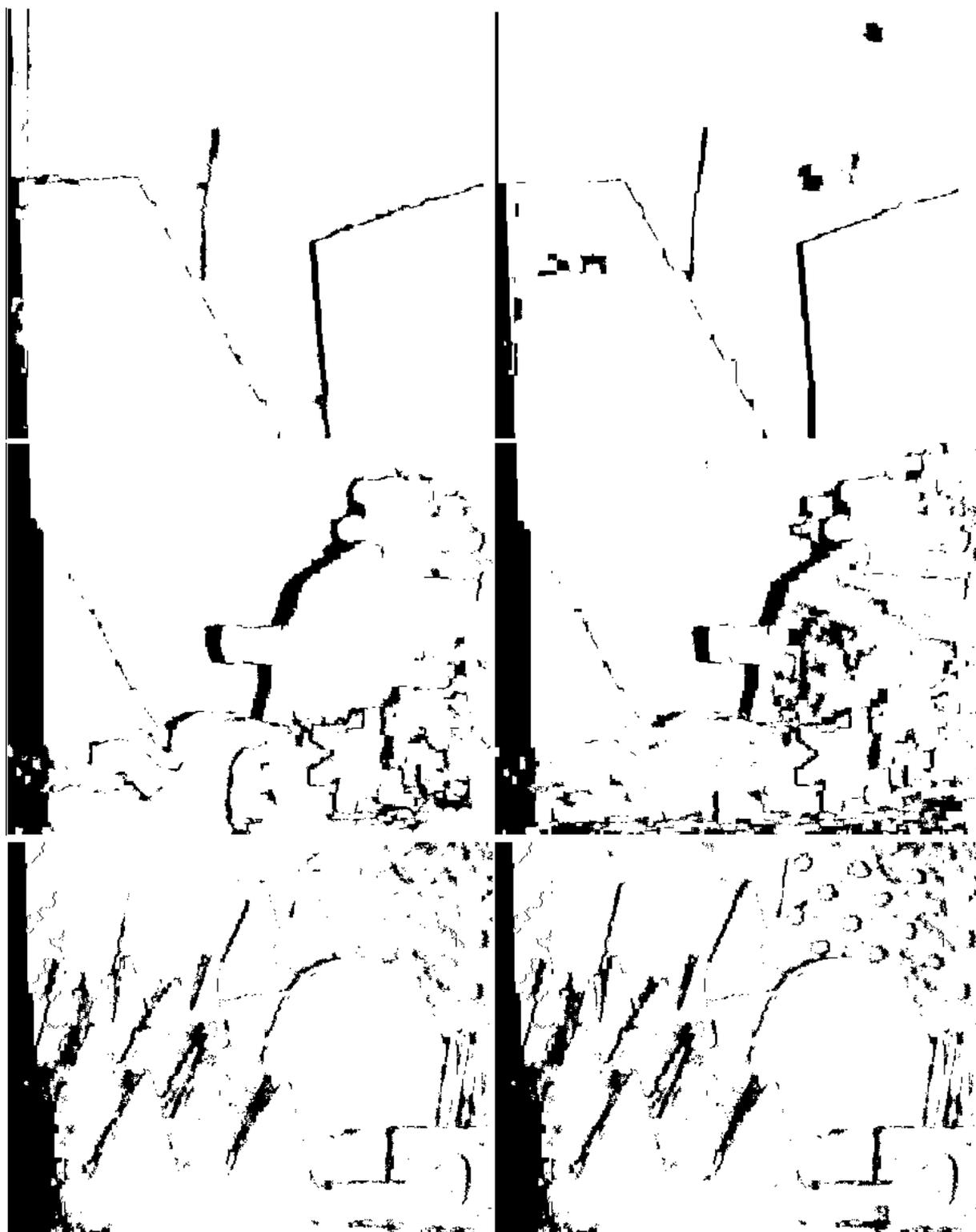
表 5-1 给出了本文提出的算法得到的遮挡图与真实遮挡图之间的比较结果。从表中我们可以看出，大约 60%的遮挡像素被检测到了，同时遮挡图中有大约 30%的像素不属于遮挡像素。

表 5-1 遮挡图的精度比较

Data set \ Occlusion map	Tsukuba	Venus	Teddy	Cones
Precision Rate	0.6343	0.6607	0.7463	0.6934
False Accept Rate	0.3785	0.2974	0.2566	0.2784

图 5-8 给出本文算法得到的遮挡图和 BP+ Ref. [36]方法得到的遮挡图的比较结果，如下：





(a) 本文算法的结果

(b) BP+ Ref. [36]方法的结果

图 5-8 遮挡图结果比较

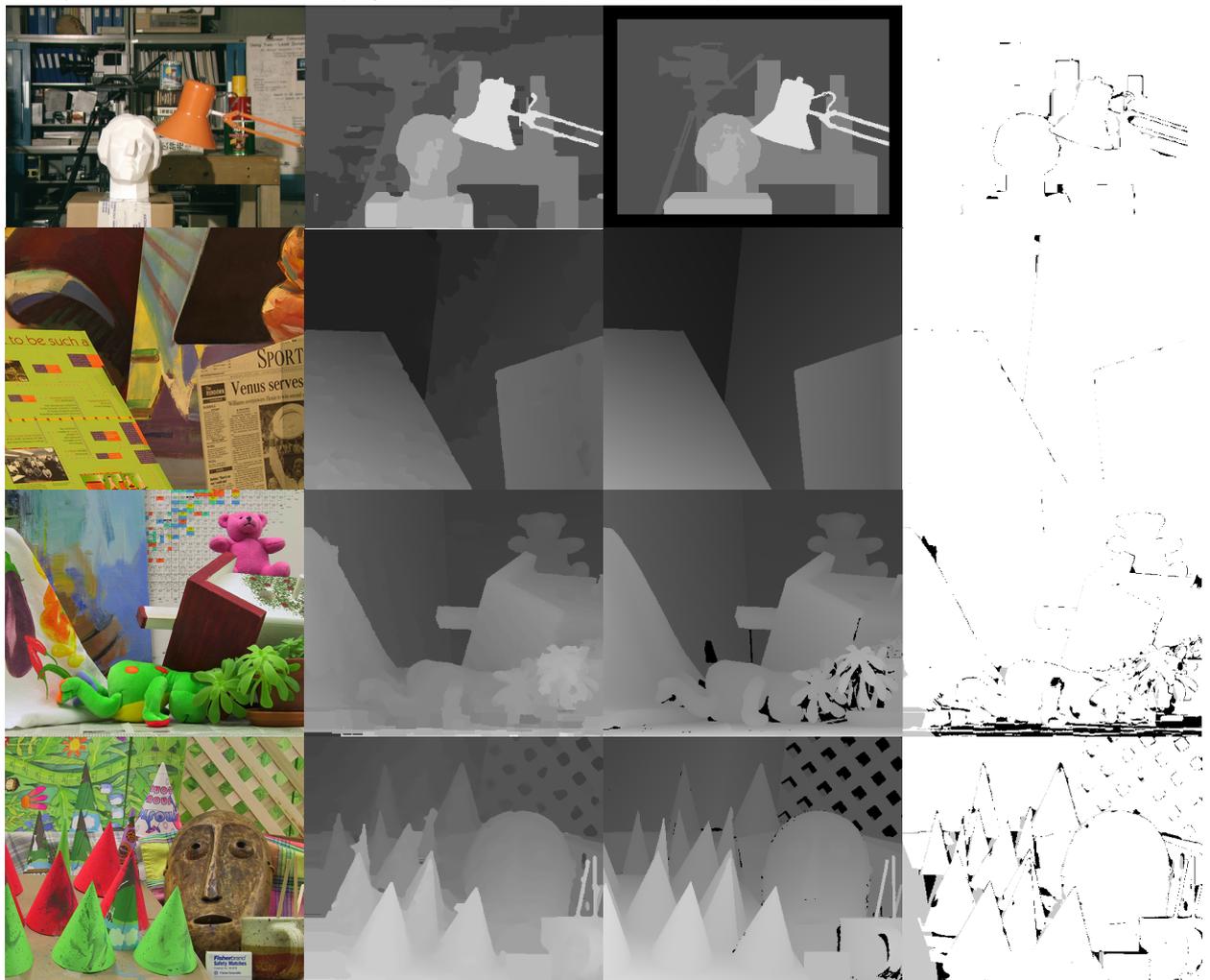
同样，我们也给出了这两个算法之间的定量比较结果，见下表 5-2。从表中可以看出本文方法得到的遮挡图精确度比 Ref. [36]得到的遮挡图的精度更高，误检率更低。

表 5-2 遮挡图的定量结果比较

Data set	Tsukuba		Venus		Teddy		Cones	
	P	F	P	F	P	F	P	F
Our method	0.6343	0.3785	0.6607	0.2975	0.7463	0.2566	0.6934	0.2784
Method of Ref.[36]	0.5935	0.7350	0.6372	0.3854	0.7128	0.4240	0.6366	0.2591

### 5.4.2 实验结果

在这一节中，我们将本文提出的算法的实验结果与当前在 Middlebury 测试集 ‘Tsukuba’, ‘Venus’, ‘Teddy’ 和 ‘Cone’ 上的实验结果进行了比较。主要从以下三个方面比较 (1) 非遮挡 (非遮挡区域中的像素点); (2) 全部 (非遮挡区域和半遮挡区域中的像素点); (3) 不连续 (接近遮挡区域的可视像素点)，通过比较每幅图像计算得到的视差图中错误匹配像素的百分比 (这里的错误像素是指其对应的视差值与真实值相差大于 1 个像素) 对视差图进行衡量。图 5-9 给出了本文算法的实验结果，其中第一列是输入图像，第二列是本文算法结果，第三列是真实视差图，第四列是错误像素视差图。



(a) 参考图像

(b) 本文算法结果

(c) 真实视差图

(d) 错误视差像素图

图 5-9 本文算法在 Middlebury 数据集上的结果

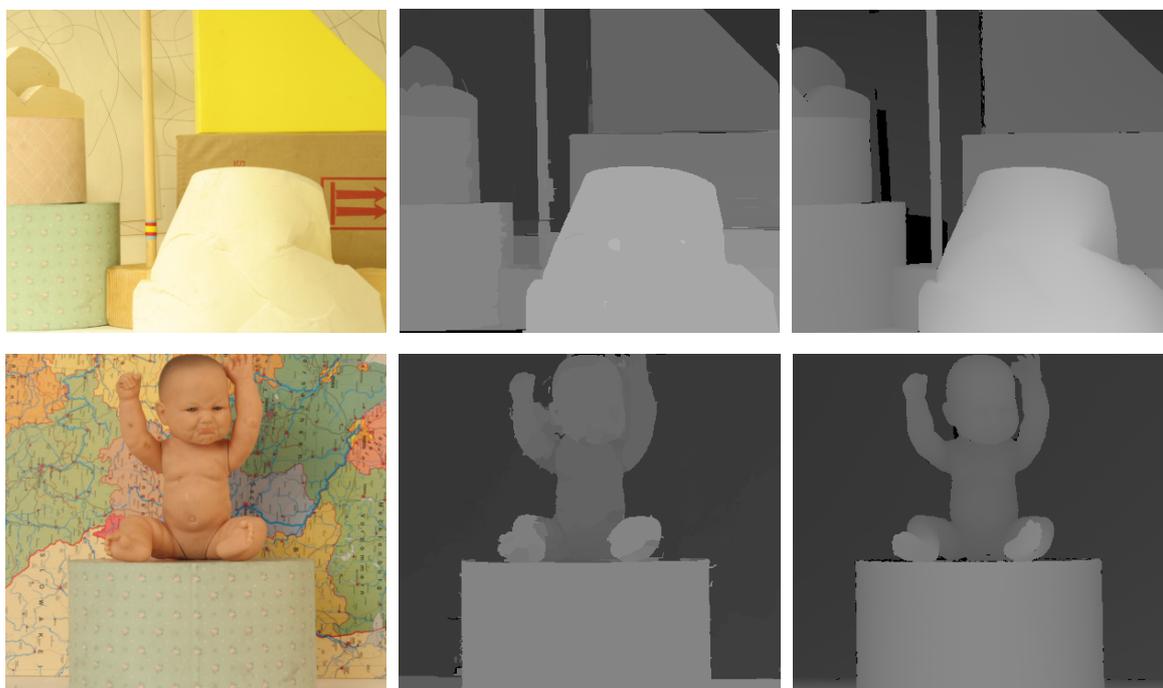
## 5.4.3 实验定量比较

为了定量的衡量本算法的实验结果，我们将本文提出的算法与当前最好的算法进行了定量的比较，表 5-3 给出了比较后的结果。

表 5-3 本文算法与其他算法定量比较结果

算法	Tsukuba			Venus			Teddy			Cones		
	Noncc	All	Disc	Noncc	All	Disc	Noncc	All	Disc	Noncc	All	disc
AdaptingBP <sup>[33]</sup>	1.11	1.27	5.79	0.1	0.21	1.44	4.22	7.06	11.8	2.48	7.92	7.32
CoopRegion <sup>[42]</sup>	0.87	1.16	4.61	0.11	0.21	1.54	5.16	8.31	13	2.97	7.18	8.01
DoubleBP <sup>[101]</sup>	0.88	1.29	4.76	0.14	0.6	2	3.55	8.71	9.7	2.9	9.24	7.8
<b>our approach</b>	<b>1.29</b>	<b>1.82</b>	<b>6.4</b>	<b>0.15</b>	<b>0.28</b>	<b>1.95</b>	<b>6.84</b>	<b>7.89</b>	<b>17.3</b>	<b>4.03</b>	<b>10.1</b>	<b>10.3</b>
Segm+visib <sup>[52]</sup>	1.3	1.57	6.92	0.79	1.06	6.76	5	6.54	12.3	3.72	8.62	10.2
C_SemiGlob <sup>[36]</sup>	2.61	3.29	9.89	0.25	0.57	3.24	5.14	11.8	13	2.77	8.35	8.2
GC+occ <sup>[27]</sup>	1.19	2.01	6.24	1.64	2.19	6.75	11.2	17.4	19.8	7.07	12.9	16.3
AdaptWeight <sup>[14]</sup>	1.38	1.85	6.9	0.71	1.19	6.13	7.88	13.3	18.6	3.97	9.79	8.26

此外，我们将本章提出的算法应用在 Middlebury 测试集的其他图像对上，如图 5-10，图中含有较多的低纹理区域和重复纹理区域，但是在本章算法的应用下，也取得了很好的实验结果。



(a) 参考图像

(b) 本文算法结果

(c) 真实视差图

图 5-10 本文算法在其他数据集上的结果

## 5.5 本章小结

本章重点研究了基于分割的全局立体匹配算法，由于遮挡多发生在区域的边界处且区域内的视差变化比较规整，同时分割后的区域数目远远小于图像的像素数，所以此类方法效率较高且取得了不错的效果，然而，该类算法一般是建立在区域内视差变化平缓和区域内不会出现视差跳跃两个假设的基础上，如果初始颜色分割发生错误，即视差在区域内部产生跳跃，则不可避免的会影响后期匹配结果，因此获得正确的初期匹配基元是该类方法需要解决的问题之一。

针对基于分割的立体匹配方法中存在的缺点，本文提出了基于视差校正和视差优化的立体匹配算法，该算法对立体视觉问题有两个主要的贡献：一是可以比较准确的计算出遮挡图的位置，这样我们可以更方便的处理图像中的遮挡像素。例如，我们可以在用全局最优化方法求解视差图的时候有针对性的加上遮挡约束；二是提出了一个渐进的区域融合算法，该算法可以很好的去优化视差平面。这个算法以大区域和置信系数高的区域为种子区域，然后逐步与其相邻的小区域进行比较与合并，形成一个大区域。在进行比较的过程中，本算法没有采用比较匹配代价的方法，而是用平面之间夹角和距离来衡量的，这样可以大大降低算法的计算复杂度。

效率较低始终是制约全局立体匹配算法的“瓶颈”问题，在以后的研究中，我们拟将采用多种手段来提高基于分割的全局立体匹配方法的效率。

## 总结与展望

计算机视觉是一门发展非常迅速的学科，现已形成了从图像获取到最终景物可视表面重构的较为完整的算法体系。经过 20 多年的研究，计算机视觉已在机器人定位与导航、精密工业测量、目标识别、虚拟现实以及军事领域中获得广泛应用。立体视觉的相关理论也正处于不断发展和完善之中。

本文主要针对双目立体视觉中的匹配问题进行了研究，主要研究成果和创新之处有以下几点：

1) 在双目立体视觉的校正系统中，提出了一种由粗到精的特征点匹配算法。该算法首先提取两幅图像的 SIFT 特征点，然后根据特征点之间的相似性约束以及同一幅图像中特征点对之间边的兼容性（称之为空间约束）建立相应的数学模型。该模型分两步求解。首先利用互相似度策略求相似性约束的局部最优解，然后利用改进遗传算法求空间约束的最优解。在应用遗传算法求空间约束的最优解的过程中，将相似性约束的结果作为遗传算法的一个初始解，同时对遗传算法做了一些改进。在三个图片测试集上进行的定量比较表明，与 Lowe's 方法以及 KD-Tree + RANSAC 算法相比，我们所提出的算法对于不同视角、不同光照、不同尺度下的宽基线图像的匹配精度更高。在得到更多、更准确的特征点对之后，利用这些特征点对计算了两个摄像机之间的相对位置关系：旋转矩阵和平移向量，最后给出了校正后的实验结果

2) 在双目立体视觉的匹配系统中，由于置信传播算法该算法对于含圈的图不能收敛到问题的全局最优解，且该算法存在计算冗余以及计算复杂度高等问题。因此我们提出了一种自适应置信传播算法。该算法也是建立在全局能量最小化框架下求最优解。在这个算法中，我们通过定义有限消息序列收敛与低纹理区域检测的方法，于是图像中的每个点都能近似自适应的收敛到最优解。与置信传播算法相比，自适应置信传播算法对立体视觉问题有两个比较大的贡献，其一是该算法的计算时间比置信传播算法的计算时间减少了大约 50%；其二是该算法的计算精度高于传统的置信传播算法的计算精度。

3) 针对基于分割的立体匹配方法中存在的缺点，我们提出了基于视差校正和视差优化的立体匹配算法。该算法对立体视觉问题有两个主要的贡献：一是可以比较准确的计算出遮挡图的位置，这样我们可以更方便地处理图像中的遮挡像素，比如我们可以在用全局最优化方法求解视差图的时候有针对性的加上遮挡约束；二是提出了一个可以优化视差平面的渐进的区域融合算法。该算法以大区域和置信系数高的区域为种子区域，然后逐步与其相邻的小区域进行比较与合并，最后形成一个大区域。在进行比较的过程中，本算法没有采用比较匹配代价的方法，而是用平面之间夹角和距离来衡量的，这样可降低该算法计算的复杂度。

图像匹配是应用非常广泛的技术，只要涉及到两幅图像、两个模型之间的比较，都会

用到图像匹配技术，本研究虽然两幅图像的匹配问题取得了一些阶段成果，但是仍有许多方面需要做进一步的研究：

1) 对宽基线下的图像匹配问题，进一步强调特征点对之间的空间约束。针对不同的应用目的，找到合适的理论模型以及方法对其进行求解。

2) 通过增加信息量来降低立体匹配的难度，包括如何充分利用多视图之间的相互关系来更好地预测两幅视图之间的视差，利用立体和运动之间的关系来得到更为准确的物体空间位置和运动方向等问题，进而提高计算效率。

3) 算法计算复杂度是全局算法中迫切需要解决的问题。全局算法必须有高效率支持才可能真正应用到实践当中。因此，需要考虑将算法向并行化发展，减少运算量，增强算法的实用性。

## 参考文献

- [1] David A.Forsyth, Jean Ponce. Computer Vision: A Model Approach. Prentice Hall, 2003.
- [2] Stefano Mattoccia. Stereo Vision: Algorithms and Applications. [http://www.vision.deis.unibo.it/smatt/fast\\_bilateral\\_stereo.htm](http://www.vision.deis.unibo.it/smatt/fast_bilateral_stereo.htm).
- [3] Marr,D. and T.Poggio. Cooperative Computation of Stereo Disparity. Science, 1976, 194(4262), 283-287.
- [4] Stephen T. Barnard and Martin A.Fischler. Computational Stereo. ACM Computing Surveys, 1982, 14(4), 553-572.
- [5] Dhond, U.R. and Aggarwal, J.K. Structure from Stereo-A Review. IEEE Transactions on Systems, Man and Cybernetics Society, 1989, 19(6), 1489-1510.
- [6] Andreas Koschan. What is New in Computational Stereo Since 1989: A Survey of Current Stereo Papers. Technical Report, University of Berlin, 1993.
- [7] Myron Z.Brown, Darius Burschka and G.D.Hager. Advances in Computational Stereo. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(8), 993-1008.
- [8] Scharstein,D. and R.Szeliski. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. International Journal of Computer Vision, 2002, 47, 7-42.
- [9] Inc,Point.Gray.Research.<http://www.ptgrey.com>.
- [10]Mulligan,J., V.Isler and K.Daniilidis. Trinocular Stereo: A New Algorithm and its Evaluation. International Journal of Computer Vision: Special Issue on Stereo and Multi-baseline Vision, 2002, 47, 51-61.
- [11]Hirschmuller, H.,P.Innocent and J.Garibaldi. Real-Time Correlation-Based Stereo Vision with Reduced Border Errors. International Journal of Computer Vision, 2002, 47, 550-580
- [12]T. Kanade and M.Okutomi. A Stereo Matching Algorithm with An adaptive Window: Theory and Experiment. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1994, 16(9), 920-932.
- [13]Veksler,O. Fast Variable Window for Stereo Correspondence Using Integral Images. IEEE Conference on Computer Vision and Pattern Recognition, 2003, 1, 556-561.
- [14]Yoon,K.-J. and I.-S.Kweon. Adaptive Support-Weight Approach for Correspondence Search. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(4), 650-656.
- [15]Veksler,O. Stereo Matching by Compact Windows Via Minimum Ratio Cycle. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(12), 1654-1660.
- [16]周秀芝, 文贡坚, 王润生.自适应窗口快速立体匹配. 计算机学报, 2006, 29(3):473-479.

- [17]刘正东. 计算机视觉中立体匹配技术的研究. 南京理工大学博士学位论文, 2005.
- [18]尹传历. 基于双目立体视觉的三维测量技术研究. 中国科学院博士学位论文, 2009.
- [19]周秀芝. 面向深度提取和形状识别的图像匹配. 国防科学技术大学研究生院博士学位论文, 2006.
- [20]H. Tao and H. Sawhney. Global Matching Criterion and Color Segmentation Based Stereo. IEEE Workshop on Applications of Computer Vision, 2000, 246-253.
- [21]H. Tao, H. Sawhney, and R. Kumar. A Global Matching Framework for Stereo Computation. IEEE International Conference on Computer Vision, 2001, 1, 532-539
- [22]S. A. Lloyd, E. R. Haddow and J. F. Boyce, A Parallel Binocular Stereo Algorithm Utilizing Dynamic Programming and Relaxation Labeling, Computer Vision, Graphics and Image Processing, 1987, 39, 202-225.
- [23]Lei,C.,J.Selzer, and Y.Yang. Region-tree based stereo using dynamic programming optimization. IEEE Conference on Computer Vision and Pattern Recognition, 2006, 2378-2385.
- [24]Olga Veksler. Stereo Correspondence by Dynamic Programming on a Tree. IEEE Conference on Computer Vision and Pattern Recognition, 2005, 384-390.
- [25]Hong L. and G.Chen. Segment-based Stereo Matching Using Graph Cuts. IEEE Conference on Computer Vision and Pattern Recognition.2004, 1, 74-81.
- [26]V. Kolmogorov and R. Zabih. Multi-camera Scene Reconstruction Via Graph Cuts. European Conference on Computer Vision, 2002, 82-96
- [27]V. Kolmogorov and R. Zabih. Computing Visual Correspondence With Occlusions Using Graph Cuts. International Conference on Computer Vision, 2001, 2, 508-515.
- [28]V. Kolmogorov and R. Zabih.. What Energy Functions Can be Minimized Via Graph Cut. European Conference on Computer Vision, 2002, 3, 65-81.
- [29]Boykov,Y., O.Veksler and R.Zabih. Fast Approximate Energy Minimization Via Graph Cuts. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23(11), 1222-1239.
- [30]P. F. Felzenszwalb and D. P. Huttenlocher. Efficient Belief Propagation for Early Vision, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004, I, 261-268.
- [31]J. Sun, N. N. Zheng, H. Y. Shum. Stereo Matching Using Belief Propagation, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(7), 787-800.
- [32]Pedro F. Felzenszwalb, Daniel P. Huttenlocher. Efficient Belief Propagation for Early Vision, International Journal of Computer Vision, 2006, 70(1), 41-54.
- [33]A. Klaus, M. Sormann, K. Karner. Segment-Based Stereo Matching Using Belief Propagation and a Self-Adapting Dissimilarity Measure, International Conference on Pattern

- Recognition, 2006, 3, 15-18.
- [34] T. Yu, R.-S. Lin, B. Super, B. Tang. Efficient Message Representations for Belief Propagation. *International Journal of Computer Vision*, 2006, 70(1), 41-54.
- [35] H. Hirschmüller. Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005, II, 807-814.
- [36] H. Hirschmüller. Stereo Vision in Structured Environments by Consistent Semi-Global Matching. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2006, II, 2386-2393.
- [37] S. Kirkpatrick, C.D. Gelatt, M.P. Vecchi. Optimization by Simulated Annealing. *Science*, 1983, 220(4598), 671-680.
- [38] S. T. Barnard. Stochastic Stereo Matching Over Scale. *International Journal of Computer Vision*, 1989, 3(1), 17-32.
- [39] J. Marroquin, S. Mitter, and T. Poggio. Probabilistic Solution of Ill-posed Problems in Computational Vision. *Journal of the American Statistical Association*, 1987, 82(397), 76-89.
- [40] Zitnick, C.L. and T. Kanade. A Cooperative Algorithm for Stereo Matching and Occlusion Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, 22(7), 675-684.
- [41] Luo, A. and H. Burkhardt. An Intensity-Based Cooperative Bidirectional Stereo Matching with Simultaneous Detection of Discontinuities and Occlusion. *International Journal of Computer Vision*, 1995, 15, 171-188.
- [42] Z. Wang and Z. Zheng. A Region Based Stereo Matching Algorithm Using Cooperative Optimization. *IEEE Conference on Computer Vision and Pattern Recognition*, 2008, 1-8.
- [43] Pearl, Judea. Reverend Bayes on Inference Engines: A Distributed Hierarchical Approach. *Proceedings of the Second National Conference on Artificial Intelligence*. 1982, 133-136.
- [44] <http://www.ptgrey.com/products/stereo.asp>
- [45] <http://www.nvela.com/products.html>
- [46] <http://www.minoru3dwebcam.com/>
- [47] Hartley, R.I. In Defence of the 8-point Algorithm, *IEEE international Conference on Computer Vision*, 1995, 1064-1070.
- [48] R. Hartley, A. Zisserman. *Multiple View Geometry in Computer Vision*, second edition, Cambridge University Press, Cambridge, 2003.
- [49] <http://vision.middlebury.edu/stereo/data/>
- [50] S. Birchfield and C. Tomasi. A Pixel Dissimilarity Measure That Is Insensitive to Image

- Sampling. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(4), 401-406.
- [51] M. Bleyer and M. Gelautz. Graph-based Surface Reconstruction from Stereo Pairs Using Image Segmentation. Proceeding of SPIE, 2005, 5656, 288-299.
- [52] Bleyer, M. and M. Gelautz. A Layered Stereo Algorithm Using Image Segmentation and Global Visibility Constraints. IEEE International Conference on Image Processing, 2004. 2997-3000.
- [53] Wei, Y. and L. Quan. Region-Based Progressive Stereo Matching. IEEE Conference on Computer Vision and Pattern Recognition, 2004, 1, 106-113.
- [54] Y. Zhang and C. Kambhamettu. Stereo Matching with Segmentation-based Cooperation. European Conference on Computer Vision, 2002, 2, 556-571.
- [55] M. Gerrits and P. Bekaert. Local Stereo Matching with Segmentation-based Outlier Rejection. Conference on Computer and Robot Vision, 2006.
- [56] F. Tombari, S. Mattoccia, and L. Di Stefano. Segmentation-based Adaptive Support for Accurate Stereo Correspondence. Pacific-Rim Symposium on Image and Video Technology, 2007.
- [57] H. Moravec. Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover. Technical Report, Carnegie-Mellon University, Robotics Institute. 1980
- [58] C. Harris and M. Stephens. A Combined Corner and Edge Detector. Proceedings of the 4th Alvey Vision Conference, 1988, 147-151.
- [59] J. Shi and C. Tomasi. Good Features to Track. IEEE Conference on Computer Vision and Pattern Recognition, 1994, 593-600.
- [60] C. Tomasi and T. Kanade. Detection and Tracking of Point Features. International Journal of Computer Vision, 1991.
- [61] David G. Lowe, Object Recognition from Local Scale-invariant Features. International Conference on Computer Vision, 1999, 1150-1157.
- [62] Matthew Brown and David G. Lowe. Invariant Features From Interest Point Groups, British Machine Vision Conference, 2002, 656-665.
- [63] David G. Lowe, Distinctive Image Features from Scale-invariant Keypoints. International Journal of Computer Vision, 2004, 60, 91-110.
- [64] Lisa Gottesfeld Brown. A Survey of Image Registration Techniques. ACM Computing Surveys, 1992, 24(4), 325-376.
- [65] Lee, D. T.; Wong, C. K. Worst-case Analysis for Region and Partial Region Searches in Multidimensional Binary Search Trees and Balanced Quad Trees. ACTA Informatics, 1977, 9(1), 23-29.

- 
- [66] Jacob E. Goodman, Joseph O'Rourke and Piotr Indyk (Ed.). Nearest Neighbors in High-dimensional Spaces. Handbook of Discrete and Computational Geometry, CRC Press, 2004
- [67] Cormen, Thomas H.; Leiserson, Charles E., Rivest, Ronald L. Introduction to Algorithms. MIT Press and McGraw-Hill, 2001.
- [68] Beis, J. and Lowe, D.G. Shape Indexing Using Approximate Nearest-neighbor Search in High Dimensional Spaces. Conference on Computer Vision and Pattern Recognition, 1997, 1000-1006.
- [69] Chanop Silpa-Anan and Richard Hartley. Optimised KD-trees for Fast Image Descriptors Matching. Conference on Computer Vision and Pattern Recognition, 2008, 1-8.
- [70] J.Xie and P.Heng. Shape Modeling Using Automatic Land Marking. International Conference on Medical Image Computing and Computer Assisted Intervention, 2005, 709-716.
- [71] A.Myronenko, X.Song, and M.Carreira-Perpinan. Non-rigid Point Set Registration: Coherent Point Drift. Advances in Neural Information Processing Systems, 2007, 1009-1016.
- [72] Alexander Toshev, Jianbo Shi, and Kostas Daniilidis. Image Matching Via Saliency Region Correspondences. Conference on Computer Vision and Pattern Recognition, 2007, 1-8.
- [73] Thomas Schoenemann and Daniel Cremers. Matching Non-rigidly Deformable Shapes Across Images: A Globally Optimal Solution. IEEE Conference on Computer Vision and Pattern Recognition, 2008, 1-6.
- [74] B.Zitova and J.Flusser. Image Registration Methods: A Survey. Image and Vision Computing, 2003, 21(1), 977-1000.
- [75] A.W.Fitzgibbon. Robust Registration of 2d and 3d Point Sets. Image Vision and Computing, 2003, 21(13), 1145-1153.
- [76] S.Gold and A.Rangarajan. A Graduated Assignment Algorithm for Graph Matching. IEEE Transaction and machine intelligence, 1996.18(4), 377-388
- [77] P.N. Suganthan. Structural Pattern Recognition Using Genetic Algorithms. Pattern Recognition, 2002, 1883-1893.
- [78] Ravindra K. A Greedy Genetic Algorithm for the Quadratic Assignment Problem. Computers and Operations Research, 2000, 27(10), 917-934.
- [79] Linzhong Liu. The Maximum Fuzzy Weighted Matching Models and Hybrid Genetic Algorithm. Applied Mathematics and Computation, 2006, 662-674.
- [80] Xuejun Tan, Bir Bhanu. Fingerprint Matching by Genetic Algorithms. International Journal on Pattern Recognition, 2006, 39(3), 465-477.
- [81] Surapong Auwatanamongkol. Inexact Graph Matching Using a Genetic Algorithm for Image

- Recognition. *Pattern Recognition Letters*, 2007, 28, 1428-1437.
- [82] M. Singh, A. Chatterjee, S. Chaudhury. Matching Structural Shape Descriptions Using Genetic Algorithms. *International Journal on Pattern Recognition*, 1997, 30 (9), 1451-1462.
- [83] A.D.J. Cross, R.C. Wilson, E.R. Hancock, Inexact Graph Matching Using Genetic Search. *International Journal on Pattern Recognition*, 1997, 30 (6), 953-970.
- [84] Yongzhong Wu, and Ping Ji. Solving the Quadratic Assignment Problems by a Genetic Algorithm with a New Replacement Strategy. *Proceedings of world academy of science, engineering and technology*. 2007, 24, 1307-1310.
- [85] R. Szeliski. *Iccv2005 computer vision contest*.  
<http://research.microsoft.com/iccv2005/Contest/>, 2005.
- [86] <http://coe.gucas.ac.cn/SDL-Homepage/resource.html>
- [87] P.H.S. Torr and A. Zisserman. A New Robust Estimator with Application to Estimating Image Geometry. *Computer Vision and Image Understanding*, 2000, 78, 138-156.
- [88] Chojnacki, W., Brooks, M.J., van den Hengel, A., Gawley, D. A New Approach to Constrained Parameter Estimation Applicable to Some Computer Vision Problems. *Statistical Methods in Video Processing Workshop*. 2002, 1-2.
- [89] Armangue, X., Salvi, J. Overall View Regarding Fundamental Matrix Estimation. *Image and Vision Computing*, 2003, 21, 205-220.
- [90] Z. Zhang, R. Deriche, O. Faugeras. A Robust Technique for Matching Two Uncalibrated Images Through the Recovery of the Unknown Epipolar Geometry. *Artificial Intelligence*, 1995, 78, 87-119.
- [91] Quan-Tuan Luong and Olivier D. Faugeras. The Fundamental Matrix: Theory, Algorithms, and Stability Analysis. *International Journal of Computer Vision*, 1996, 17(1), 43-75.
- [92] P. Torr, A. Zisserman and S. Maybank. Robust Detection of Degenerate Configurations for the Fundamental Matrix. *International Conference on Computer Vision*, 1995, 1037-1042.
- [93] G. Csurka, C. Zeller, Z. Zhang, and O. Faugeras. Characterizing the Uncertainty of the Fundamental Matrix. *Computer Vision and Image Understanding*, 1997, 68, 18-36.
- [94] Pearl, Judea. Reverend Bayes on Inference Engines: A Distributed Hierarchical Approach. *Proceedings of the Second National Conference on Artificial Intelligence*, 1982,
- [95] Kim, Jin H., Pearl, Judea. A Computational Model for Combined Causal and Diagnostic Reasoning in Inference Systems. *Proceedings of the Eighth International Joint Conference on Artificial Intelligence*, 1983, 133-136.
- [96] Pearl, Judea. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco, CA: Morgan Kaufmann, 1988.
- [97] Murphy, K., Y. Weiss, and M. Jordan. Loopy Belief Propagation for Approximate Inference:

- an Empirical Study. *Uncertainty in Artificial Intelligence*, 1999, 467-475.
- [98] Wainwright, M.J., T.S. Jaakkola and A.S. Willsky. Tree-reweighted Belief Propagation and Approximate Estimation by Pseudo-moment Matching. *Workshop on Artificial Intelligence and Statistics*, 2003.
- [99] Meltzer, T., C. Yanover, and Y. Weiss. Globally Optimal Solutions for Energy Minimization in Stereo Vision Using Reweighted Belief Propagation. *International Conference on Computer Vision*, 2005, 428-435.
- [100] D. Comaniciu, P. Meer. Mean shift: A Robust Approach toward Feature Space Analysis. *IEEE Transactions on Pattern and Machine Intelligence*, 2002, 24(5), 603-619.
- [101] Q. Yang, L. Wang, R. Yang, H. Stewénius and D. Nistér, Stereo Matching with Color Weighted Correlation, Hierarchical Belief Propagation and Occlusion Handling. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2006, II. 2347-2354.



## 发表文章目录

- [1] Zhihua Liu, Zhenjun Han, Qixiang Ye, Jianbin Jiao. A New Segment-Based Algorithm for Stereo Matching. IEEE International Conference on Mecharonics and Automation, 2009, 999-1003.
- [2] Zhihua Liu, Qixiang Ye, Lu Ke, Jianbin Jiao. A Progressive Region-Merging Algorithm for Stereo Matching. IEEE Youth Conference on Information, Computing and Telecommunications, 2009, 142-145.
- [3] Zhenkai Zhang, Zhihua Liu, Jianbin Jiao. DSP Implementation of a Multi-Channel Video Display System with Image Stitching. IEEE Youth Conference on Information, Computing and Telecommunications, 2009, 204-207.
- [4] Lin Zhang, Zhihua Liu, Jianbin Jiao. An Improved RANSAC Algorithm for Image Stitching. SPIE, 2011, 7080.
- [5] Zhihua Liu, Qixiang Ye, Lu Ke, Jianbin Jiao. Robust Feature Point Matching Using a Coarse-to-Fine Strategy. Chinese Journal of Electronics. (under review)
- [6] Zhihua Liu, Qixiang Ye, Jianbin Jiao. Segment-based Stereo Matching Algorithm with Disparity Correction and Optimization. Computer Vision and Image Understanding. (under review)
- [7] Zhihua Liu, Qixiang Ye, Jianbin Jiao. An Adaptive Belief Propagation for Stereo Matching Problem. Image and Vision Computing (under review)



## 致谢

衷心感谢导师焦建彬教授对我的精心指导，焦老师宽厚的学术功底、严谨的治学态度、强烈的责任心、平易近人的工作作风和不辞辛苦的敬业精神使我终身受益，是我前进道路上的楷模。

衷心感谢叶齐祥副教授对我的悉心指导，叶老师给我提供了很多的建议和观点，与他进行深入而细致的讨论使我开拓了视野，受益匪浅。

衷心感谢中科院数学与系统科学所的刘振宏研究员和韩继业研究员对我的指导，他们渊博的数学知识和谦逊的教学态度是我今后努力的方向。

衷心感谢实验室的师弟师妹在科研上给我的帮助，与他们的交流使我收获很多。

最后，我要感谢我的家人，是他们默默地在精神和物质上给予我无私的关心和帮助，使我能顺利地完成学业。

刘志花

2011年4月