

摘要

立体匹配是计算机视觉的一个重要分支，作为立体视觉中研究最活跃的主题之一，在三维场景重建、移动机器、对象识别、智能控制、三维测量等领域有广泛的应用场景。

立体匹配的实质是以同一场景从不同角度拍摄的两幅图像作为输入，提取图像对应的视差图。通过视差与深度的关系，进一步获取场景中物体的深度信息，从而应用于各个领域。立体视觉技术的实现一般包括：图像获取，图像校准，立体匹配以及三维重建四个步骤。立体匹配是最重要也是最困难的一步。

本文在经典局部匹配算法 SAD 的基础上，分析算法的不足，进行算法优化，提出了基于遮挡检测的SAD算法，并为检测出来的遮挡区域的像素应用本文提出的遮挡区域填充算法，另外通过添加 Census 灰度变换和基于几何距离权重的算法，提高立体匹配的匹配准确率。用标准测试图像作为对比，将实验结果通过评估函数进行测评。测评结果证明，优化算法提高了图像遮挡区域的匹配精度，同时整体提高了视差图的匹配质量。

关键词：立体视觉，局部匹配方法，算法优化

ABSTRACT

ABSTRACT

As one of the most active stereo vision research theme, stereo matching is an important branch of computer vision and has been widely used in the three-dimensional scene reconstruction, the field of mobile robotics, object recognition, intelligent control, three-dimensional measurement.

Based on the substance of the two images of the same scene from different angles, stereo matching is to extract the corresponding disparity map of the input image. With the disparity map, we can parallax the depth map to obtain further information on the depth of objects in the scene, which has been applied to various fields. Stereo scopic vision technology generally includes four steps: the image acquisition, the image calibration, the stereo matching and the three-dimensional reconstruction. Among them, the stereo matching is the most critical and most difficult step.

Based on the classic SAD local algorithm, by analyzing the lack of SAD algorithms we proposed the improved SAD local algorithm which is based on the occlusion detection. By using the occlusion detection algorithm to deal with the two disparity maps, we use the filling algorithm to fill in the occlusion region. Finally, we used Census conversion and changed SAD local algorithm to base on the distance of block pixels in order to improve the result of matching. Using standard test images as comparsion, experimental results will be input to the function for evaluation. Experimental results show that the implementation of algorithm improves the accuracy of image matching block area, while improving the overall matching quality disparity map.

Keywords: Stereo vision, Local matching algorithm, Algorithm improving

目 录

第1章 绪论	1
1.1 计算机视觉技术概述	1
1.1.1 计算机视觉简介	1
1.1.2 计算机视觉的发展现状	2
1.1.3 计算机视觉的关键问题	3
1.2 双目视觉系统简介	3
1.3 立体匹配技术简介	4
1.4 本课题主要研究内容	4
1.5 本文结构安排	5
第2章 双目视觉系统	6
2.1 双目视觉系统概述	6
2.2 双目视觉系统的基本原理	7
2.3 双目视觉的数学模型	8
2.4 摄像机的标定	10
第3章 立体匹配方法及优化	12
3.1 立体匹配方法简介	12
3.2 立体匹配方法的分类和基本过程	13
3.2.1 基于特征的立体匹配方法	14
3.2.2 全局立体匹配方法	15
3.2.3 局部立体匹配方法	16
3.3 立体匹配的难点和算法优化方向	18
第4章 双目视觉系统的构建及实验总结	20
4.1 摄像机离线标定	20
4.2 实验过程匹配效果的评估	23
4.3 局部立体匹配SAD算法实现	25
4.4 匹配窗口的选择	28
4.5 优化局部立体匹配算法	29

目录

4.5.1 增加Census变换	30
4.5.2 双向匹配遮挡检测	32
4.5.3 基于几何距离的自适应权重	33
4.6 实验结论及全文最后总结	34
参考文献	37
致 谢	39
外文资料原文	40
外文资料译文	42

第1章 绪论

1.1 计算机视觉技术概述

1.1.1 计算机视觉简介

众所周知，生物可以通过自身的视觉系统，获取外界景象的形状大小、明暗亮度、行为变化、位置颜色等各种信息，这一功能对生物的自身发展，心理行为甚至生存具有重要意义。据统计人类获取的信息至少有超过五分之四的外界信息是经由视觉系统获得，这也就是为什么说视觉系统对于信息采集至关重要。随着科学技术的不断发展，为了让机器能够获得更加丰富的外界信息，从而为人类提供更加可靠高效的服务，人们提出了视觉理论的框架，并不断地完善计算机视觉的相关知识架构，加上计算机硬件水平和摄像机采集能力的提高，人们逐渐开始通过计算机及相关设备模仿人类的视觉系统。实现具有图像信息获取能力的系统，利用它对采集的图片或视频进行处理，经过一系列的算法运算，实现机器对现实中各种应用场景的分析判断，提取有用信息并根据这些信息进行对应的机器决策，这就是计算机视觉。

人类的视觉系统不外乎两个主要的功能，第一是适应外界环境，第二是根据信息决定自身行为。相应的，要实现一个计算机视觉系统，该系统必须能够准确识别图像信息，确定物体的位置和形状，以及能根据一段时间内物体的位置判断物体的运动状态。因此，目标的识别定位、三维形状建模以及运动状态分析就组成了计算机视觉的主要研究内容。计算机视觉是科学领域中的一个重要研究领域，它所具备的科研价值与其巨大的应用潜力是相匹配的。

由于是对人类视觉的模仿，那么既然要实现计算机视觉，就需要分析提炼出人类视觉系统的识别过程，每当我们用眼睛接收到现实世界中物体的信息时，人类的大脑需要对这个物体存在某种表达。目前，计算机视觉领域中主要存在两种主流的表达理论。一种是由视觉理论研究的创始人马尔（Marr）提出的三维表达理论（三维重建理论）^[1]，马尔从信息处理的角度出发，综合了心里物理学、神经生理学和图像处理学的研究成果，首次提出了著名的 Marr 视觉理论。它也是最早的计算机视觉理论框架。Marr 的视觉理论其实就是把计算机视觉的过程看作

一个信息处理过程，并将这个过程分为三个阶段层次：计算理论层次、算法与数据结构层次和硬件实现层次。另一种是本世纪初发展出的基于图像的表达理论。三维表达理论的基本思想是：物体是以通过几何形状的方式在大脑中被表达的。由于改变物体的观察视角是不会改变物体的三维形态的，所以三维表达是一种与视角无关的表达，对此，Marr在其同时期出版的书中作了详细介绍，可以说正是Marr的三维物体表达模型打开了计算机视觉研究的大门。但随着研究的深入，人们发现Marr的三维重建理论与很多生理和心理实验有出入，针对这些问题，一种基于图像表达的全新理论被提了出来。2004年Poggio等人对基于图像表达的理论做出了比较全面的论述。时间再往后，随着神经科学的研究发展，基于图像表示方式的弊端也渐渐凸显，一些学者开始研究将二维和三维混合起来建模。综上，物体表达是计算机视觉中避免不了但至今仍没有得到很好解决的一个任务。一般来说，计算机视觉的发展有马尔视觉计算理论、目的视觉理论、分层重建理论和基于学习的视觉理论这4个历程。而无论在哪个阶段，对现实场景的精确记录都是计算机视觉研究的基础。

1.1.2 计算机视觉的发展现状

计算机视觉是近代的一门新兴学科，伴随计算机技术的快速发展，在上世纪中后期后开始迅速发展，逐渐成为近代计算机科学最为热门的研究主题之一。计算机视觉指的是利用计算机和相关摄像设备来仿真和模拟生物的视觉，通过对采集的图像或视频信息进行处理，计算出相应场景的三维信息，这是计算机视觉的核心任务。如同生物视觉系统的工作机制，让机器也拥有视觉信息获取的能力，通过这些信息帮助机器进行更加准确的判断，让机器代替人类从事一些基础的服务工作。计算机视觉在许多领域有着广泛的应用场景，例如在地质勘测、交通监测、犯罪监控、增强现实、虚拟现实、人工智能等领域都是项目的核心技术，而且占据研究力度的比例越来越高，因此各领域的研究人员对计算机视觉技术的重视程度也是日渐提高。

科学技术不断发展，到了二十一世纪，在各领域学者的共同努力之下，计算机视觉逐渐形成并完善了一套系统的理论框架，如何快速实现一个具备视觉识别功能的系统已经逐渐体系化。计算机视觉技术已经可以通过计算机和相关的摄像机设备来对各种各样的视频图像信息进行识别，提取，匹配，处理。围绕着计算机视觉系统实现的四个模块：图像提取，摄像标定，立体匹配，信息重建，已经有了较好的实现方案，其中立体匹配作为双目视觉中的核心问题，是目前业界共

同努力的方向，主要围绕如何提高立体匹配方法的匹配精确度和算法的运行速度这两个方面。

1.1.3 计算机视觉的关键问题

在客观现实空间中，物体深度信息的获取是计算机视觉技术研究领域中的核心问题。目前随着计算机视觉技术和立体匹配技术的迅速发展，为了能够更加准确，快速地获取真实空间中景物的深度信息，已经研究出很多方法和理论，并逐渐成为了立体视觉领域中的一个热门研究主题。双目视觉通过立体匹配方法对图像信息进行一些列处理，还原出显示世界坐标系中的深度坐标，是计算机视觉的基础，如果要获得深度信息，进行一系列后续信息处理和操作判断，就必须做好双目视觉中最核心的一个功能，立体匹配。

双目视觉是计算机视觉的一个基础而重要的话题，而在双目视觉系统的实现模块中，立体匹配方法决定着计算机匹配精准度和匹配速度，是直接影响到双目视觉系统的匹配效果最重要的因素。可以说立体匹配方法研究是双目视觉领域中的最核心问题。立体匹配方法根据分类基准的不同有多种方式，主要有注重匹配精确度而牺牲算法运算速率的全局匹配方法，反之也有注重匹配速度，对匹配精确度要求不高的局部匹配方法，它们有各自不同的应用场景。无论是哪一种匹配算法，都要解决图像匹配过程中的遮挡区域，低纹理区域，深度不连续区域的误匹配问题。

1.2 双目视觉系统简介

双目立体视觉(Binocular Stereo Vision)，也简单称作双目视觉，是计算机视觉的基础研究分支。由于要获取图像的三维信息，必须至少用有两幅图像，单目系统在稳定性上也远远不如双目系统，所以相对于单目系统，双目系统更具有研究意义。类似人类拥有一双眼睛，双目系统对于同一物体可以获取不同角度的两幅图像，提取其像素信息，利用摄像机内部参数，根据匹配算法计算两幅图像之间像素点的对应关系，得出图像视差图，再根据三角测量原理，利用视差图信息计算出图像各个位置的三维深度坐标，这就是双目视觉系统的实现过程。

一个完整的双目立体视觉系统通常由相机安装、摄像头标定、立体校正、图像预处理、立体匹配和深度信息计算这几个部分组成。其中立体匹配是双目视觉系统实现的关键部分。

1.3 立体匹配技术简介

立体匹配方法是本论文主要研究的内容。立体匹配是通过获取同一物体的两幅或者多幅图像的信息，计算物体三维几何信息的一种技术。立体匹配的基本过程是：首先获取现实场景中同一物体同一时刻不同角度的两幅图像，经过图像校准后，建立图像对同一世界坐标像素点之间的映射关系，最终得到匹配视差图。立体匹配主要包括四个步骤，匹配代价计算、匹配代价聚集、视差最优化计算和视差细化。

如果根据双目系统提取计算出的视差图稀疏程度分类，立体匹配方法主要有基于灰度信息和基于特征信息这两种匹配方法。还可以根据匹配基元的不同，将立体匹配分为基于特征，基于相位，基于区域匹配等多种各具特色的方法。基于特征的立体匹配方法特点是受图像特征影响大，并且该方法只能获得稀疏的视差图，匹配的准确度相对而言较差；基于相位的匹配方法算法复杂度则比较大，匹配时间往往过长，这就决定了它不适用于需要快速匹配的场景；另外根据优化规则的区别，还可分为全局匹配和局部匹配方法，局部匹配算法和基于区域匹配算法其实是相同的原理，只是分类的出发角度不同而已。

立体匹配算法实现的指导思想是实现以下三个目标：输出图像对视差图，以此为基础如何提高匹配精确度和算法速度。立体匹配在立体视觉中占有重要的作用，立体视觉的应用推动匹配算法不断的发展，尤其是在速度和精度两个方面，这也是立体匹配过程中要解决的一个时间和精度的矛盾。

1.4 本课题主要研究内容

本课题将研究计算机在视觉领域中的关键问题：双目视觉系统立体匹配方法。基于双目视觉系统的成像原理和立体匹配过程的学习，研究匹配算法，实现双目视觉系统的基本模块，包括摄像机的离线标定，图像获取，立体匹配，内容重点是研究立体匹配方法的实现。

上文已经提到，立体匹配方法是以两幅图像对应像素点的邻域作为参考单元，通过某一相似性函数进行相似度匹配的方法。在各种不同的立体匹配方法中，局部匹配算法既能得到稠密的视差图，又能极大程度地缩小搜索范围还有运算时间。考虑到实验条件和算法在现实场景中的实用性，本课题研究的立体匹配算法主要是基于区域的立体匹配方法，即局部匹配算法。

本课题将在实现经典的局部固定窗口算法基础上，从匹配窗口大小，遮挡图像检测，自适应权值等多个方向优化算法以提高立体匹配的效率和精准度，利用匹配评估函数对比算法的匹配结果，得到匹配效果良好的立体匹配方法，对比不同实例图像的匹配效果，得到稳定的匹配优化算法。

1.5 本文结构安排

论文结构将由整体到局部，依次对双目视觉系统的框架原理，系统各个模块的实现方法，立体匹配技术实现，不同场景和图像情况下的算法优化方法，匹配评估函数，实际场景匹配效果试验进行分析和展现。具体章节结构安排如下：

第1章：绪论，这一章主要是简单介绍计算机视觉的发展现状，以及其核心问题双目视觉立体匹配技术的一些概念。

第2章：介绍双目视觉系统的 basic 框架和实现方法，研究双目视觉系统的数学模型，以及双目视觉系统中摄像机标定的过程及意义

第3章：介绍多种立体匹配方法，分析对比算法各自的实现难点，对不同图片区域的匹配精确度，以及实际应用的可行性。在介绍完立体匹配基础算法后，针对不同的图像情况，研究立体匹配过程中精确度和运算时间要达到一定的标准，所要解决的一些难点，为算法优化提供可行方向。

第4章：构建双目视觉系统，离线标定，通过编程语言完成立体匹配算法，并以此为基础进行算法优化，通过实验对匹配结果进行评估，对比，综合优化过程得到一种匹配效果较好，误匹配率低的立体匹配方法。最后，总结实验结果和本文。

第2章 双目视觉系统

2.1 双目视觉系统概述

双目视觉系统如同人的双眼一样，我们观察外部世界，感知到外界三维信息，对物体远近的位置信息进行判断，在大脑中加工这些信息以使我们掌握真实的客观世界形态。双目视觉系统对人类双眼的模拟就是依据于此，为了使计算机能够获得客观世界中的物体三维信息，就需要利用双目视觉系统，将图像的二维信息进行一系列处理，还原世界坐标下该物体的三维信息。

如果只有一幅图像，那么摄像机无法获取到图像中物体的深度信息，因为单一图像只能保存平面在某一时刻的信息，摄像机无法识别某一像素点在与摄像机平面的垂线方面上进行移动的过程，因此无法得到完整的物体信息，但是如果用两台摄像机，就可以捕捉到这些信息，双目视觉系统由此而来。

双目立体视觉获取深度信息前需要先得到图像对的视差图，视差图就是两幅输入图像之间像素点的偏移量信息图，根据这些偏移量信息和摄像机标定得到的内部参数就能求出深度信息。视差图的获取的基本过程一共包括三个步骤：摄像机标定，获取成像模型所需内外部参数；图像获取，用标定好的摄像机系统拍摄目标场景图像，采集二维信息；立体匹配方法，寻找输入图像对之间的对应像素点，输出视差图；最后就是根据视差图来算深度，获取物体深度信息。综上可以按照功能可以将双目视觉系统分为图像获取，摄像机标定，立体匹配，深度计算四个主要的模块^[2]。如图2-1所示。



图 2-1 双目视觉系统的主要模块

2.2 双目视觉系统的基本原理

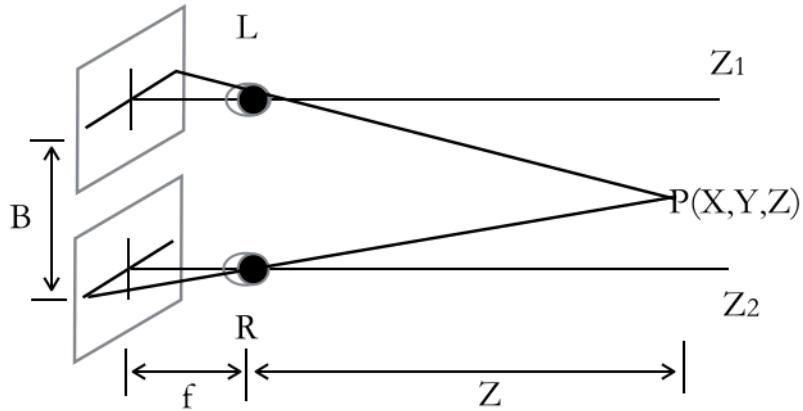


图 2-2 人类视觉感知深度的原理示意图

如图2-2所示为人通过双眼获取物体深度信息的原理图。图中，L表示人的左眼，R表示人的右眼，投影平面可以看做人的视网膜成像平面，B为两眼之间的瞳距，通常人双眼之间的距离大概是6厘米，P为待观察目标。两眼在同一时刻观察同一目标时，由于两只眼睛的观察角度不同，所以左右眼看到的图像不是完全相同的，两眼成像像素点之间的偏移量就是视差值。将观察到的图像所有像素点一一映射对应，可以将所有图像的视差值记录成像，这个就是视差图。大脑通过对两眼观察到的图像信息进行处理，就可以对现实场景进行立体感知。双目视觉系统就是模拟人眼视觉立体感知过程的，视差原理是双目视觉系统的一个重要原理。根据人类视觉的特性，研究其成像原理，是双目视觉领域的重要研究方向，可以说双目视觉技术的实现目标就是实现接近于人的双目视觉系统和人的三维感知能力。

简单来说，双目视觉系统模拟人眼识别，首先获取图像信息，随后通过立体匹配方法输出图像对的视差图，利用视差信息和摄像机标定得到的内外部参数，经过一系列计算得出三维信息，双目视觉系统的数学模型将在下一小节展开介绍。

双目视觉系统模拟人的视觉，对物体按照一定的规格采集两幅图像，以此分析计算出视差图。如图2-3所示为一个双目摄像机获取像素信息的简单示意图，PQ是一个像素点的移动轨迹，对物体的某一点的深度感知无法在只有单目摄像机时感知，当物体从P点移动到Q时，对于做摄像机而言物体没有发生变化，而右摄

像机则记录下了深度的信息，根据几何成像的原理即可得出物体深度变化信息，如果只有单目摄像机就无法复原丢失的三维信息。左右摄像机匹配对应像素点的过程就是立体匹配方法，立体匹配方法作为本论文的重点研究内容将在下一章详细介绍。

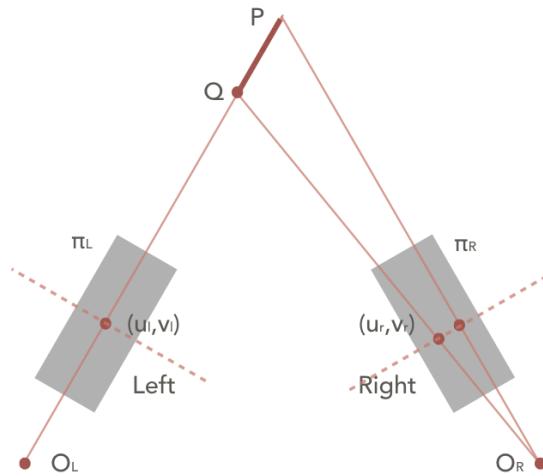


图 2-3 双目摄像机获取像素点信息的示意图

2.3 双目视觉的数学模型

双目视觉系统是利用视差图和摄像机内外部参数求解图像深度信息的过程，本节将简单介绍双目视觉系统在提取所需信息，求解现实场景三维信息的过程。

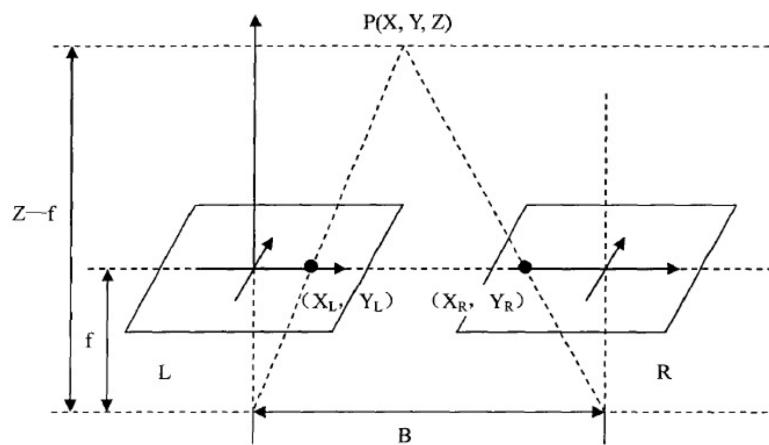


图 2-4 双目测距原理图

为了方便理解和公式求导，将成像面翻转到目标点与视点之间，目标点与视点之间的连线在左右两个成像平面的交点就是该目标点在摄像机坐标系中所成的像，双目测距原理示意图如图2-4所示。

在双目摄像机安装，调节完成之后，通过摄像机标定获取相关参数，然后通过双目摄像机同时拍摄在焦距范围内的待拍目标P，获取包含目标P的左右两幅图像，分别用L和R表示，由于双目摄像机必须确保前后位置相同，所以这两幅图像处于同一垂直空间平面内，所以 $Z_L = Z_R$ ，P在左右图像上的投影坐标分别表示为 (X_L, Y_L) 和 (X_R, Y_R) ，又因为摄像机水平平整，所以有 $Y_L = Y_R$ 。假设现实坐标系与摄像机左坐标系重合，那么P点与左摄像机中心光轴的距离为X，B表示两摄像机光轴中心之间的距离，所以P点与右摄像机中心光轴的距离为 $(B - X)$

根据相似三角形的计算原理，可得公式2-1：

$$\frac{X_L}{X} = \frac{f}{Z-f} = \frac{X_R}{B-X} \quad (2-1)$$

其中，目标距离Z和目标横坐标X都是未知数，变换公式可得公式2-2：

$$BX_L = (X_L + X_R)X \quad (2-2)$$

由公式2-2可解得观测目标横坐标X：

$$X = \frac{BX_L}{X_L + X_R} \quad (2-3)$$

再将观测目标横坐标X值，即2-3代入公式2-1得：

$$\frac{X_L + X_R}{B} = \frac{f}{Z-f} \quad (2-4)$$

令 $X_L + X_R = D$ ，D即为视差，经过公式变化可得：

$$Z = \frac{Bf}{D} + f = f\left(1 + \frac{B}{D}\right) \quad (2-5)$$

在公式2-5中，可以通过测量，标定获得所需要的系统参数，如像距f、轴间距B，视差 $D = X_L + X_R$ 通过立体匹配方法在两图像中找取对应点坐标计算现实场景中同一像素点的偏移量得到。完成标定和立体匹配后，f、B、D已知，就可以计算得目标点P与成像平面之间的距离Z，即深度信息，将距离Z代入公式2-1，可得目标横坐标X：

$$X = \frac{B}{D}X_L \quad (2-6)$$

同理可得，目标纵坐标Y为：

$$Y = \frac{B}{D}Y_L \quad (2-7)$$

公式(2-5) (2-6) (2-7)分别求出了空间目标像素点在世界坐标下的坐标位置，实现双目系统的功能，这也就是双目立体视觉测距技术的基本原理。如何准确的获取内部参数和视差图是重点，这也是确保最终测距精确度的难点，必须有足够准确的视差图和摄像机参数。下一节将介绍摄像机标定的过程，而视差图的获取将在下一章展开。

2.4 摄像机的标定

双目视觉系统在对获取图像信息进行加工得出深度信息的过程中，必须得到摄像机的一些内部固定参数，另外由于存在畸变，输入图像在进行立体匹配之前必须经过摄像机参数的校正，摄像机内外部参数获取的过程就是摄像机标定，摄像机标定结果直接影响立体匹配效果以及最终的信息还原情况。

摄像机标定主要包括摄像机成像系统内外的几何及光学参数的标定，双目摄像机之间相对位置关系的标定，其主要目的是确定目标物体的世界坐标与图像坐标的映射关系。摄像机标定的精度直接影响到双目视觉系统的精确度，是决定立体匹配效果的关键因素，具有重要的研究意义。

获取图像的每个像素点坐标与空间中物体表面对应点的实际位置具有相关性，这种位置上的映射关系由摄像机的成像几何模型唯一确定。通常把这个成像的几何模型参数称为摄像机参数，计算这些参数的过程也就是摄像机定标的过程。由上一节推导可知，这是双目视觉系统所依赖的公式计算的基础。

摄像机标定最初是在摄影测量学中逐渐研究发展起来的，摄影测量学提出了四个标定问题^[3]：

(1)绝对方位：确定两个坐标系统之间的位置和方向等一系列传感器在绝对坐标系的坐标点的校正。

(2)相对方位：确定相对方向和位置在两个摄像头从预测的校准点在现场。

(3)外部方位：确定方向和位置，相机在绝对坐标系统中的投影校准点。

(4)内部方位：确定内部的几何参数，其中主要包括摄像机常数，位置的主要观察点，纠正透镜畸变。

这些标定问题最先起源于通过航拍提取信息，创建相关地形地图的技术，是一个经典的摄影测量学难题。值得一提的是摄影测量学也包括双目视觉系统立体匹配和深度恢复的问题。

摄像机的标定存在许多方法，传统摄像机标定方法精度比较高，但需要足够的标定模板，过程复杂繁琐；相比之下，基于主动视觉的摄像机标定方法不仅不需要标定模板，而且算法简单且稳定性较高，不过设备比较昂贵，并且不能够用于摄像机运动状态不确定的情形；自标定法是一种利用环境的刚体性，依靠图像间的对应关系进行标定的方法，这种算法灵活性强，但是稳定性不高。对于大多数双目视觉系统而言，通常只需要在实验或实际应用之前将摄像机固定好，进行一次标定即可，计算出相关参数，实现离线标定。虽然只进行一次标定即可，但是由于标定结果对图像纠正和立体匹配效果有直接影响，所以标定过程要尽可能精确。

在一次离线标定的前提下，主动视觉标定法和自标定法就不如传统标定算法有优势了，传统的标定算法，如张正友标定法仅需要一个平面模板，对摄像机和模板的位置没有限制。张正友标定法通过摄像机拍摄同一个平面靶标，多次改变靶标的位置和角度，提取相关信息，因为在改变靶标的过程中假定摄像机的内参数不变，因此可以根据多个平面数据计算出摄像机的内部参数。通常张正友标定法使用的是黑白棋盘格作为靶标，其作为靶标的优势是，棋盘格的特征，即黑白格形成的角点非常容易提取，而且特征点之间的位置关系是确定的。

第3章 立体匹配方法及优化

立体匹配是双目视觉系统的核心问题，通过立体匹配寻找两幅图像中对应的像素点的位置，得出匹配结果，通常用视差图表示，从而进一步求出深度信息。本章将着重介绍立体匹配方法，基于多种分类方式研究各种立体匹配方法的实现优缺点，以及如何提高立体匹配的精确度，在立体匹配过程中存在的各种各样的难题。

3.1 立体匹配方法简介

从根本上讲，立体匹配方法其实就是通过构建一个能量代价函数^[4]，来最小化估计像素点视差值的一个过程。立体匹配方法实质是一个最优化求解问题，算法通过建立一个合理的能量函数，根据匹配约束条件增加相应的约束公式，结合最优化理论方法进行方程求解，这种求解方法也是典型的病态问题求解。在实际的双目视觉系统中，对采样图像进行拍摄获取时往往受到许多不确定的环境因素干扰，曝光不一致，抖动等问题都提高了立体匹配的难度，如何减少不利因素带来的影响，获得具有高准确匹配率的图像匹配结果是立体匹配技术的难点也是要攻克的核心问题。

随着各方面条件成熟，立体匹配方法迅速发展，现在已经出现了各种各样的匹配算法，它们各具优势各具特色，不过这些算法都离不开几个重要的组成部分，只是在各个部分存在实现的差异。立体匹配方法的这几个重要部分为匹配代价计算，匹配代价聚集，视差最优化计算，视差细化四个部分^[5]。



图 3-1 立体匹配常见算法分类

匹配代价计算：基于灰度差值的绝对值匹配代价计算方法是这个步骤常用的一种代价方法，此外还有基于灰度值差值平方的绝对值来计算匹配代价，归一化

互相关等，不过这往往只是算法基础，通常还要在对匹配代价函数添加一些参数因子或者进行一些变换以提高匹配准确度。

匹配代价聚集：输入单个参考像素进行匹配代价计算后，由于存在较大的误差，所以必须按照一定的区域计算匹配像素点周围若干像素的匹配代价结果，再对统计结果进行比较判断，找出对应像素点，提高匹配准确度，聚集像素匹配代价结果的通常为有固定视差的二维区域，窗口有简单的固定窗口，能够适应不同图像区域的可变窗口和自适应窗口等等。

视差最优化计算：通过匹配代价聚集之后得到一个最优的匹配像素点，得到视差值，这是局部匹配方法的实现过程，即简单地通过一个像素找到参考图像最优匹配的像素求出视差，即所谓的赢家通吃法(Winner Take All)。这个方法没有考虑到其他像素点的匹配情况，所以匹配正确率较差，通过视差优化可以平衡匹配效果，达到整体最优。

视差细化：通过视差细化可以进一步提高立体匹配效果，输出更加标准的视差图，通过一些常用方法对，如中值滤波法，消除图像中的噪点；用中间像素视差精练法，用曲线拟合视差变化情况，去除可能出现的毛刺，在图像中的像素灰度变化平滑区域有很好的效果。这一步骤通常是全局匹配方法的最后一步。

3.2 立体匹配方法的分类和基本过程

立体匹配方法作为双目视觉系统中的核心问题，经过多年的研究发展，涌现了各种各样具有不同优势的算法。

若根据匹配基元的不同，主要分为基于特征的立体匹配方法，基于相位的立体匹配方法，还有基于区域的立体匹配方法。基于特征的立体匹配方法中，匹配基元是基于全局优化的图像特征集，如点，边缘，区域特征，从图像特征集中寻找对应关系实现匹配；基于区域的立体匹配方法中，匹配基元是具有一定尺寸的模板窗口，采用相关函数测量视差范围内待测窗口相似性实现匹配；基于相位的立体匹配方法，其匹配基元是图像相位，利用多尺度的空间频率分析方法，提取图像不同频段的信息进行匹配。匹配原理基本相同，都是基于基元的相似性测量，但适用方法的搜索策略各不相同。

若根据匹配结果的稠密程度，则可分为稠密匹配和稀疏匹配方法。稠密匹配以图像的灰度、邻域相关程度等为判断依据进行匹配，为绝大多数像素找到匹配

结果。稀疏匹配是以边缘、轮廓、线段等图像特征为匹配内容进行匹配，也就是基于特征的立体匹配，仅得到特征点处匹配结果。

若根据匹配的优化规则，可以分为全局立体匹配方法和局部立体匹配方法。局部匹配方法在经典的窗口匹配基础上，还有自适应窗体，自适应权值，多窗口匹配等优化方向，局部匹配方法其实也是一种基于区域的立体匹配方法，它们只是分类角度不同而已。全局立体匹配的主要过程是在局部匹配的基础上进行匹配结果优化，主要的全局优化方法有图割法，置信传播，动态规划等。

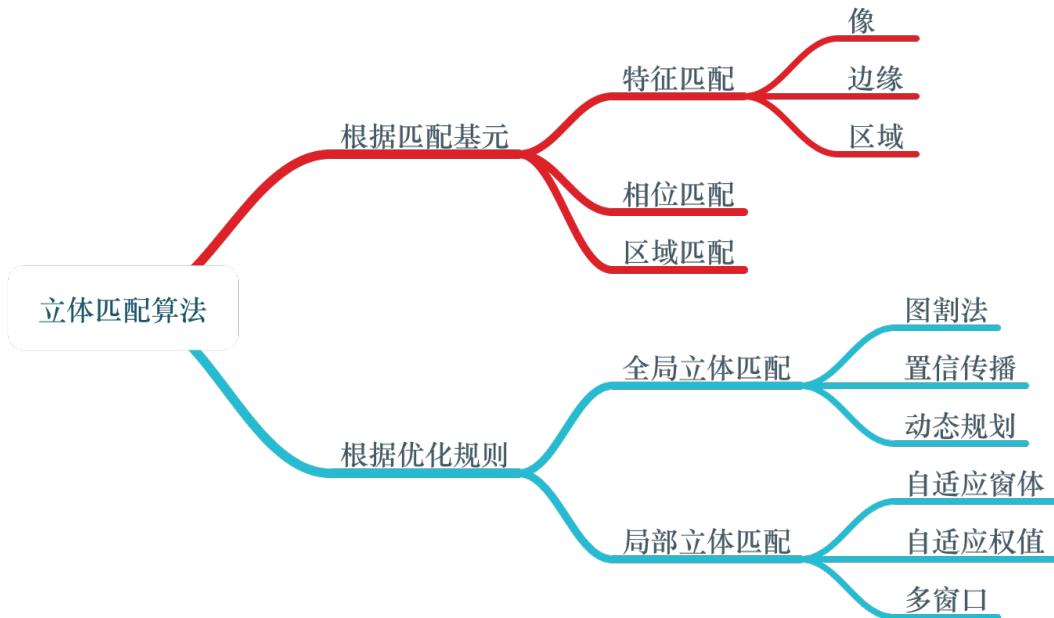


图 3-2 立体匹配常见算法分类

本文将介绍三种比较常见的立体匹配方法：基于特征的立体匹配方法，全局立体匹配方法，局部立体匹配方法（基于区域的立体匹配方法）。

3.2.1 基于特征的立体匹配方法

像素灰度值是图像的基本特征，不过单一像素的灰度值不能提供足够的信息，无法实现立体匹配，而图像特征可以提供一个更高层次的图像描述。图像特征是一幅图像中局部的、有意义的、可检测的部分。基于特征的匹配算法，主要是基于图像的几何特征信息，如边缘、线、角点、兴趣点和几何基元等，相应不同的图像信息和图像特征选取合适的特征信息，针对这些几何特征进行视差估计，首先必须提取图像的特征点，特征越明显越好，然后获取图像特征点区域的对应的

视差图，由此可以看出基于特征的立体匹配方法是一种稀疏的立体匹配方法，它只能对拥有明显特征点的特征区域进行立体匹配。

基于特征的立体匹配方法所需要的主要步骤：图像预处理、提取特征、特征点的匹配，最后获取稀疏视差图，可以通过采用差值估计，插值的方法来获取稠密的视差图。提取特征时可提取图像的点、线、面这些布局特征的分布情况，也可提取图像结构或者多边形等全局特征。基于特征的立体匹配方法运行速度快，缺点是依赖于图像的特征表现，对特征不明显的图像信息难以匹配，此外特征提取过程易受到拍摄环境的影响，遮挡、光线条件、重复纹理等影响较大；差值估计计算量大，对于特征不明显以及要求得到稳定输出视差的双目系统，基于特征的立体匹配方法适合用于一些简单且特征明显的应用场景。

3.2.2 全局立体匹配方法

基于局部区域的立体匹配算法主要通过优化局部窗口内累积代价的计算方式来估计视差值，这种方法既能在一定程度上保证匹配速度，又能照顾到匹配精度。这些特质使得局部匹配算法在继特征立体匹配算法之后成为立体匹配领域的研究热点。但近些年得益于计算机计算能力的大幅提升和计算成本的下降，使得研究者能够越来越放手的去追求匹配精度。

全局立体匹配算法主要是在局部立体匹配的基础上，根据约束条件，采用一些全局的优化理论方法来估计视差，细化误差，建立全局能量函数，通过最小化全局能量函数得到最优的视差值。可以说全局立体匹配将独立的各个像素点的匹配结果统一起来进行分析，从而得到更加精确的匹配结果，要实现这个过程，十分消耗运算时间。近些年来随着计算机计算能力的大幅提升，计算成本也逐渐降低，全局匹配方法也得到一定程度的发展，往往运用在对匹配结果精确度要求高的场合。

利用匹配约束方法可以解决全局匹配过程中匹配点的选择问题，在中心像素点寻找到多个符合的匹配参考点时，结合约束方法，往往能够得到准确的结果，常用的约束条件包括：极线约束，唯一性约束，顺序一致性约束，视差连续性约束，相似性约束等

全局立体匹配方法常用的全局优化方法有动态规划等，图割法（graphcuts）、信念传播（belief propagation）等等^[6]。全局匹配算法特点明确，通过牺牲算法运算时间，得到匹配正确率更高的视差图，因此不适合实时运行，实时匹配的应用场景。

3.2.3 局部立体匹配方法

局部立体匹配算法是立体匹配算法中发展迅速并且应用领域非常广泛的一类算法，局部立体配算法的实质是：通过直接对输入的图像对之间的彩色信息或灰度信息进行匹配，从而得到其中一幅图像对应的视差图。不同于基于特征的匹配算法得到的视差图为稀疏的视差图，大多数局部匹配算法都能得到图像对应的稠密视差图，保留了图像中丰富的视差信息和深度信息。基于以上原因，本课题的研究方向定位局部立体匹配算法的研究和改进，而局部立体匹配算法中，匹配窗口尺寸和形状的选择是一个很重要的因素，一个大小形状合适的窗口对算法的结果影响很大。窗口不能太小，否则，窗口中包含的可用信息太少，导致无匹配的可能性增大。窗口也不能太大，否则，一方面严重影响到匹配效率，一方面可能受到投影畸变的影响。在固定窗口的局部匹配算法中通常使用以中心像素为中心点的正方形窗口。

由于窗口的选择对局部立体匹配算法的效果影响很大，由此形成了以SAD算法为代表的基于窗口的局部立体匹配算法。除了基于窗口的局部匹配算法以外，以此为基础，近年来也出现了一些优化算法，如基于Census变换的局部算法为代表的基于图像变换的局部匹配算法，还有基于傅立叶变换的局部算法、基于Rank变换的局部算法等。此外还有基于自适应权值的局部立体匹配算法，这一类算法的表面采用固定的匹配窗口进行匹配，实际上则是通过自适应的权值为每一个参考像素选择最佳形状和尺寸的匹配窗口，以ASW算法为代表，正是基于自适应权值的局部立体匹配算法将局部立体匹配算法推向了顶峰。

总体而言，局部立体匹配算法可大致分为三类：基于窗口的局部立体匹配算法、基于变换的局部立体匹配算法和基于自适应权值的局部立体匹配算法。

基于窗口的局部立体匹配算法：顾名思义，即匹配窗口的大小是固定不变的。是局部算法中最为经典也是最为简单的一类。算法简单易于实现，而且匹配效率很高，缺点是算法提取的视差图受窗口大小影响很大，而且视差图中物体的轮廓不清晰，对图像中深度不连续区域的匹配效果较差。

基于变换的局部立体匹配算法：在进行匹配之前，先对匹配窗口进行变换，对经过变换的窗口进行立体匹配，代表算法有基于Census变换的局部立体匹配算法。其他的算法还包括：基于傅立叶变换的局部算法、基于Rank变换的局部算法等。由于该类算法在基于窗口的局部算法的基础上加了图像变换，因此算法复杂

度和基于窗口的局部算法略高。由于以上原因，这一类算法效率较高，但提取的视差图中噪点仍然较多，物体边缘轮廓不够清晰。

基于自适应权值的算法为每一个像素寻找匹配像素点时，不是单独地进行对应窗口像素的灰度值的差值作累积，而是为窗口中的每一个像素对中心像素的贡献设置一个权值（常用的权值有基于颜色的权值和基于几何距离的权值），在窗口区域内与中心像素点关联越大的点对整个窗口匹配代价的影响就越大，通过这种自适应权值的方法可以大大提升图像的整体匹配效果，尤其是在图像的边缘轮廓区域，获取较为清晰的视差图。在匹配效率方面，由于采用固定大小的窗口，因此匹配效率也相对于可变窗口的匹配算法效率高，因此，在双目立体匹配领域被广泛的使用。

局部立体匹配算法使用简单的赢者全胜(Winner Take All)的视差选择策略，与全局立体匹配算法相比，它通过匹配窗口的全部像素的累积量来进行匹配。相比全局立体匹配算法，局部立体匹配算法有以下优点：

匹配稳定性好：区别于基于特征的立体匹配方法，局部立体匹配方法拥有较好的稳定性，不依赖于拍摄目标的特征，针对不同图像区域的局部匹配算法都有较好的匹配效果。

算法复杂度低，运行速率快：目前越来越多的应用环境对系统的响应时间要求很高，甚至很多应用场景要求算法的效率要接近实时运算，因此，虽然全局算法的匹配效果不错，但单从效率上讲，已经是毫无用武之地。鉴于这个原因，局部算法的研究和学习也越来越广泛和深入。

匹配准确率较好：早期，全局算法备受关注，原因是全局算法获取的视差图质量相比当时的局部算法获取的视差图表现出色很多，但是随着局部算法的进一步研究和改进，目前，许多局部算法提取的视差图的效果已经很好，甚至超出某些全局算法的匹配效果，随着局部算法匹配效率的提高，相应的匹配复杂度也不断提升，即局部算法效果的提升是以牺牲匹配效率为代价的。

尽管局部算法的复杂度不断提升，但是相比于全局算法，在速度方面的优势仍旧是非常显著的，加上优化过后的全局算法也有不俗的匹配输出效果，因此，在某些对实时性和匹配效果并求的应用场景下，局部算法无疑是最佳选择。

3.3 立体匹配的难点和算法优化方向

在双目系统的立体匹配过程中，通常对于左图像中的某一个像素点，通过匹配方法找到的位于右图的匹配候选点往往有多个，而实际上真正的匹配点最多只有一个，遮挡区域的像素则实际上是不能找到匹配点的。而产生多个匹配候选点则会提高误匹配的概率，导致匹配准确度的下降，在图像的遮挡区域，低纹理区域以及深度不连续区域是最容易产生误匹配的地方，这也是立体匹配需要克服的几个难点。也是本文算法优化所要解决的问题。

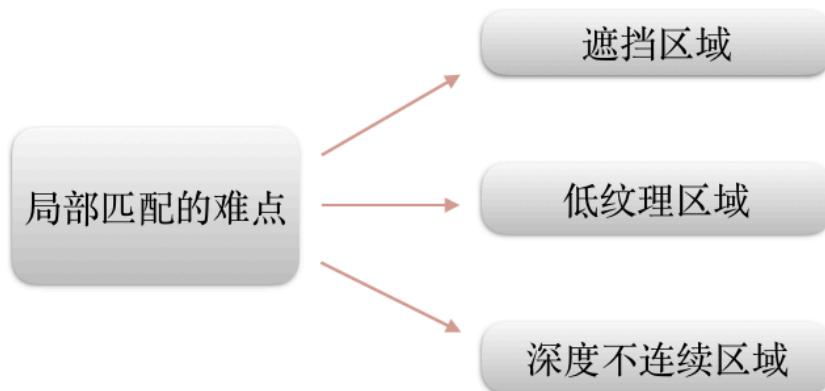


图 3-3 局部匹配的难点

遮挡区域：由于摄像机在同一水平面平移的原因且光圈大小一致，所以某些区域的像素点不可能同时出现在输入的两幅图像中，存在着左图像有而右图像被遮挡的情况，也就不可能找到对应的匹配点。这种现象就是遮挡，遮挡住的图像区域就是遮挡区域。遮挡是立体匹配过程中面临的主要困难之一，由于前景和背景在不同观察点观察的时候，位置偏移量会有所不同，部分的背景在参考图像上存在，而在目标图像中被遮挡了，使得参考图像中的像素点失去了对应的匹配点，从而造成误匹配。解决遮挡的方法是添加遮挡检测，可以通过多幅图像进行处理，也可以利用交叉检验的方法判断被遮挡像素点，主要是采用后者进行检测遮挡的像素点，再通过后景视差填充，输出遮挡区域的视差图像。

低纹理区域：该区域内的像素高度相似，比如一张纯色的木桌，通过一般的相似性测量来寻找匹配点，通常会找到多个匹配点与之对应，因为该区域相似性太高，导致难以得到正确匹配结果。弱纹理或重复纹理区域主要是指在某些区域中的像素点，其颜色特征比较相近，即纹理特征并不丰富。通常可以通过增加窗

口大小以获取更多中心图像周围的灰度信息来解决，不过在匹配的过程如果窗口的过大也会导致视差图噪点增加，降低其他区域的匹配准确度，所以对于匹配窗口大小的选择必须经过一定的实验验证。

深度不连续区域：这种区域普遍存在于场景中的物体的边缘处，由于物体之间的像素差异而存在的自然区分带，匹配输出视差图常常出现的边缘模糊情况一般就是由于深度不连续区域引起的。一般认为连续性约束只适用于物体表面，在物体边缘处是不满足的。针对这类区域中的像素匹配过程，要提高匹配精度，可以通过采用多个窗体，在每一个窗体中，像素不是总处于窗体的中间位置，而是在不同的位置。然后计算在每类窗体下，计算匹配代价，为每一个像素选择一个最优的窗口。

下一章将通过具体的实验环境，实现双目视觉系统的基本功能，包括标定求出内外参数，实现立体匹配功能，为了得到更好的立体匹配效果，得到精确的视差图，将在基础的局部立体匹配方法基础上增加一些优化模块，最后通过评估模块和标准图像验证试验结果。

第4章 双目视觉系统的构建及实验总结

经过前文对双目视觉系统的介绍，对双目视觉和立体匹配有了初步的了解，本章将在实验中实际搭建一个双目视觉系统，并实现一种基础的局部立体匹配方法，通过几种优化方法，提高立体匹配效果，降低匹配错误率。

在实验中，立体匹配方法的逻辑实现主要使用了框架OpenCV^[7]，OpenCV是一个跨平台计算机视觉开源库，用C++语言编写，其主要的接口编程语言也是C++，但仍保留了大量的C语言接口，也有许多Python, MATLAB, Java接口可以使用。利用这个框架实现的函数可以快速搭建视觉系统，实现立体匹配功能。

4.1 摄像机离线标定

首先对摄像机进行离线标定，以获取摄像机的内外部参数，摄像机标定结果将用来纠正输入的左右图像，使图像只在水平方向上存在偏移，简化计算过程，这对于后面立体匹配的匹配效果影响很大。

离线标定方法采用张正友标定法，“张正友标定”又称“张氏标定”，是指张正友教授于1998年提出的单平面棋盘格的摄像机标定方法^[8]。张氏标定法已经作为工具箱或封装好的函数被广泛应用，如OpenCV函数库中封装了张正友标定法的实现函数cvCalibrateCamera2。此文中所提到的方法，为相机标定提供了很大便利，并且具有很高的精度。用该方法标定可以不依赖于特殊的标定物，只需要一张打印出来的棋盘格。操作简单可行，本实验使用OpenCV实现张正友棋盘法来进行标定。

实验之前必须明确标定的三个问题：

相机标定的目的：获取摄像机的内部参数和外部参数矩阵，内参和外参系数可以对之后相机拍摄的图像就进行矫正，以得到畸变相对很小的图像。

相机标定的输入：标定图像上所有内角点的图像坐标，标定板图像上所有内角点的空间三维坐标（一般情况下假定图像位于Z=0平面上）。

相机标定的输出：摄像机固定不变的内部参数和外参系数。

这三个基础的问题就决定了使用Opencv实现张正友法标定相机的标定流程、标定结果评价以及使用标定结果矫正原始图像的完整流程：首先要准备标定图片，

棋盘法常用棋盘如图4-1所示。用摄像机拍摄多张拍摄图片，图片的位置和角度适当做一些改变，棋盘平面与摄像机坐标平面夹角不宜太大，获取多张样本可以减少实验误差，然后输入图像，提取每张照片的角点信息，进一步提取亚像素角点以提高标定精确度，在棋盘标定图上绘制内角点，对摄像机进行标定，查看标定结果并对其进行评估，获取内外参数矩阵。

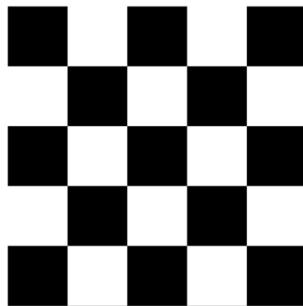


图 4-1 实验棋盘图

具体实验过程，将双目摄像机安装固定好后，确保摄像机安置平面水平平整，由此假定摄像机的内参（包括焦距长度，透镜畸变参数等）和外参（旋转矩阵和平移矩阵）均固定。利用张正友标定方法，通过采集多组不同角度、距离的棋盘格图片，即可计算出摄像机的内外参数。这个步骤可以通过Matlab的标定工具离线完成。采用Matlab平台对双目相机进行标定，标定方法采用张正友标定法，具体实现为OpenCV的cvCalibrateCamera2函数。

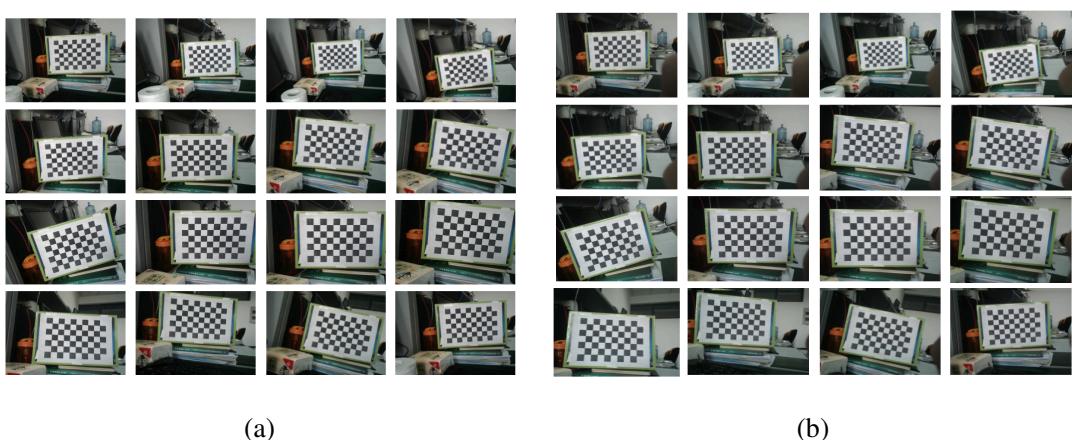


图 4-2 采集的图像(a)左摄像机;(b)右摄像机

打印一张9x7的棋盘格，每个方格的边长为25mm，铺平贴到硬纸板上以便拍

摄。尽量让棋盘格在左右图像中占30%以上的比例，以便在标定过程中更加准确检测到角点，减少标定误差。用双目摄像机从不同角度、距离拍摄，得到多组图片。采集图片如图4-2(a)和4-2(b)所示。

将图像提取焦点后标定，Matlab输出定结果如图4-3所示。

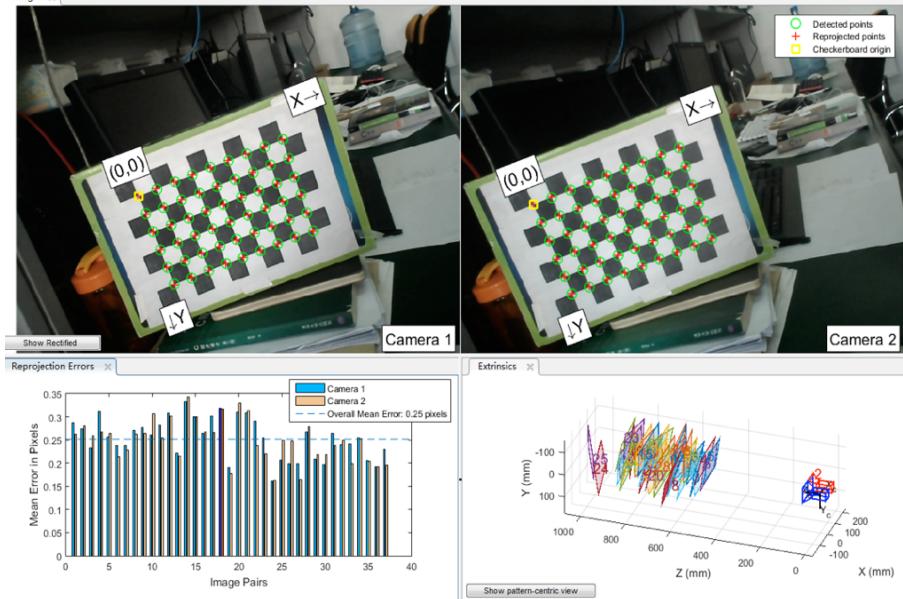


图 4-3 标定结果

整合所有角点信息，计算出摄像机的内外参数表4-4

参数记录		左摄像机	右摄像机
内参数	内参矩阵	$\begin{bmatrix} 719.8 & 0 & 0 \\ -0.871 & 724.5 & 0 \\ 0.216 & 257.9 & 1 \end{bmatrix}$	$\begin{bmatrix} 722.8 & 0 & 0 \\ -0.294 & 727.4 & 0 \\ 337.2 & 237.3 & 1 \end{bmatrix}$
	畸变向量	$[0.216 \ -0.714 \ 0.902]$	$[-0.256 \ -1.162 \ 2.24]$
外参数	旋转矩阵	$\begin{bmatrix} 1 & 0.00042 & -0.0036 \\ -0.00041 & 1 & 0.00502 \\ 0.00336 & -0.00502 & 1 \end{bmatrix}$	
	平移向量	$[-68.714 \ -0.7531 \ 0.2987]$	

图 4-4 内外参数记录表

通过标定参数对采集图像进行纠正，可以使图像在垂直方向的坐标相同，这

样视差就转变为计算左右视图水平坐标的距离，方便了视差的计算。

从图4-5(a)和图4-5(b)可以看到纠正图像后，图像的棋盘格子边缘与水平直线重合。

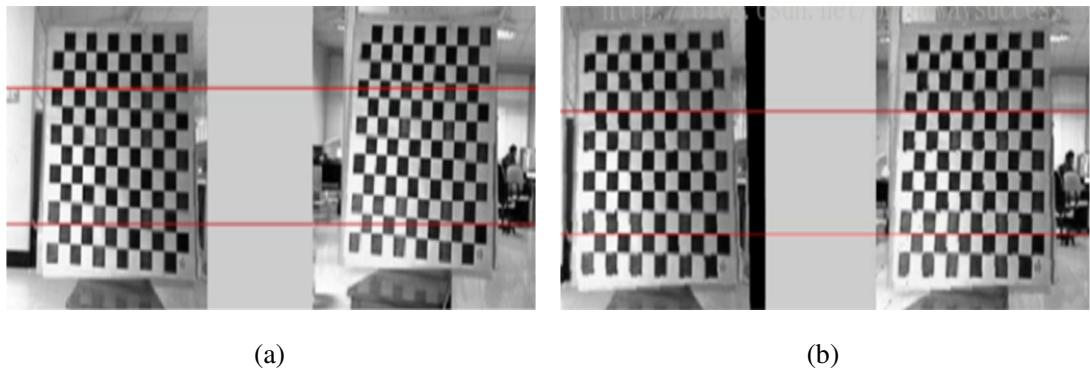


图 4-5 采集的图像(a)标定前左右图像位置;(b)标定后左右图像位置

4.2 实验过程匹配效果的评估

立体匹配的过程，除了主观观察图像立体匹配效果，大多时候匹配误差没有办法直接主观地进行判断，需要通过某种客观的评估方法对匹配结果进行评估，以分析算法匹配精确度。

双目视觉系统在过去的很长一段时间里，业内一直没有一个统一的度量立体匹配算法优劣的评判标准。直至 2002 年，Daniel 发表文章，首次提出了立体匹配算法的框架并建立了 Middlebury 网站的测评系统，并给出了一系列的标准测试图像。利用标准的测试图像，我们可以判断匹配效果与标准输出视差图的差距，从而得到算法的匹配精确度。该测评系统和测试图像自建立以来，一直被沿用至今，并且已经被广泛认可和采纳。

Daniel等人不仅提出了立体匹配算法的度量标准，还提供了四种用于测试算法的标准测试图像序列，代号分别为Tsukuba、Cones、Venus、Teddy，如图4-6。这四种标准图涵盖了许多拍摄场景的布局和光照情况，为了避免标定的内外参数误差对立体匹配效果的影响，已经方便客观地输出立体匹配精确度，实验过程中将先采用Middlebury网站提供的标准左右图，最后将匹配结果与网站提供的标准视差图对比评估。



图 4-6 四种标准图像

Middlebury 网站除了提供纠正过的标准左右图像，还有提供图像准确的遮挡区域图4-7所示，也提供了准确的视差图4-8，这些图像信息都是经过严格匹配的结果。有了这些标准图像，我们就可以利用评估函数对匹配方法的准确性进行测试。

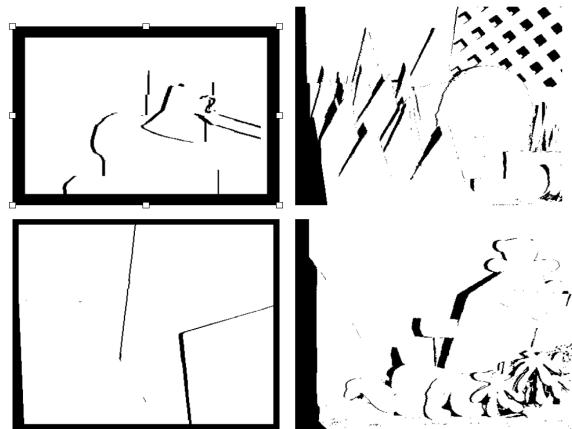


图 4-7 四种标准图像的遮挡区域图

有了标准左右图像和标准视差图，只要将匹配结果与标准视差按某种度量方法进行比对，就能得到一个客观的立体匹配精确度。常用的度量方法有RMS (root-mean-squared) 和 POBMP (Percentage of bad matching pixels)

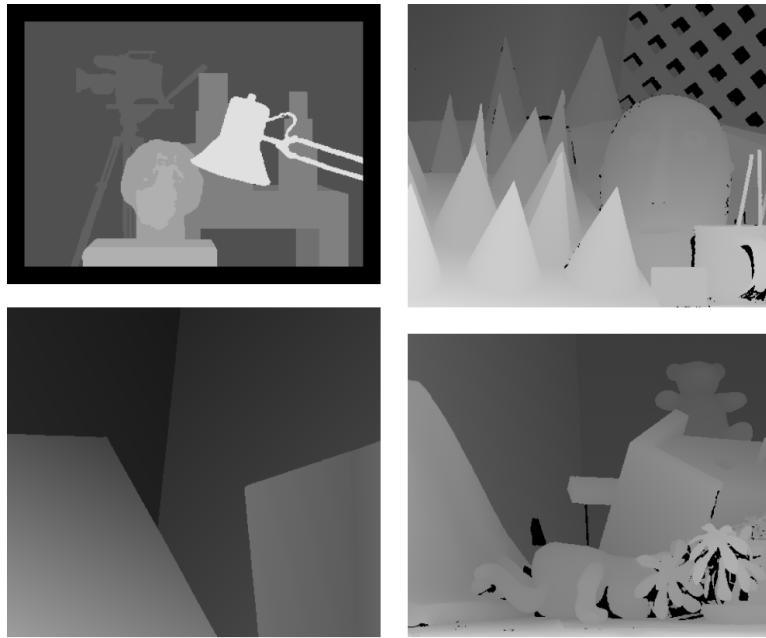


图 4-8 四种标准图像的准确视差图

本实验将通过编程实现一个基于POBMP度量标准的立体匹配算法评估函数，直接统计误匹配的像素占图像整体像素的百分比，即误匹配率，POBMP公式：

$$POBMP = \frac{1}{N} \sum_{(x,y)} (|d_c(x,y) - d_t(x,y)| > \delta_d) \quad (4-1)$$

其中， $d_c(x,y)$ 是计算出来的视差图中像素 (x,y) 的视差值， $d_t(x,y)$ 是真实视差图中像素 (x,y) 的视差值， δ_d 是视差值的误差阈值，通常取1， N 是图像中像素的总个数。通过输入middlebury的标准匹配视差图和实际匹配输出视差图进行对比，计算得出误匹配率。

算法执行速度的评估则直接根据C语言头函数time.h库自带的时间打点记录函数来计算算法的执行时间。

4.3 局部立体匹配SAD算法实现

双目视觉系统的关键模块就是立体匹配方法的实现，本节将以经典的局部立体匹配方法SAD为例，实现一次完整的局部立体匹配过程。

局部匹配算法在匹配代价聚合过程中有多种相似性测度函数，如绝对误差和(the Sum of Absolute Differences,简称SAD)，平方误差和(the Sum of Squared Differences,简称SSD)，归一化互相关(normalized cross correlation,简称NCC)等，SAD是

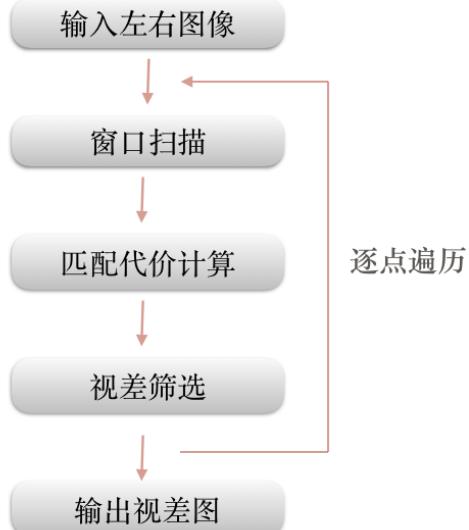


图 4-9 局部匹配算法流程图

最简单的一种，其的计算公式为：

$$SAD(x, y, d) = \sum_{(x,y) \in N_L} |I_1(x, y) - I_2(x, y + d)| \quad (4-2)$$

其中， N_L 是以参考点 (x, y) 为中心的参考窗口， R_N 是以目标点 $(x, y + d)$ 为中心的目标窗口， d 为视差， I_1 为参考图像， I_2 为目标图像。该算法分别以参考点 (x, y) 和目标点 $(x, y + d)$ 为中心，在参考窗口 L_N 和目标窗口 R_N 中统计窗口内对应位置像素的灰度值和的差，以此作为参考窗口 L_N 和目标窗口 R_N 的误差。在给定的视差取值范围内遍历所有的目标窗口，使误差SAD取值最小的目标窗口对应的视差值 d 即为参考点 (x, y) 的视差值。SSD算法只是更改相似性函数的计算方法，将绝对值误差替换为平方值误差，其他方法步骤相同。

SAD算法具体的执行步骤为：

- (1) 以左图像像素点为中心构造一个小窗口，类似于卷积和。
- (2) 用同等大小的窗口覆盖右图像同一垂直坐标的区域，水平移动遍历所有可能窗口。
- (3) 在每一个可能窗口内，遍历依次选取窗口内的像素点。
- (4) 左图像像素点坐标减去右图像像素点坐标，求出同一窗口内所有像素点差的绝对值之和。
- (5) 同理，依次计算所有参考窗口。

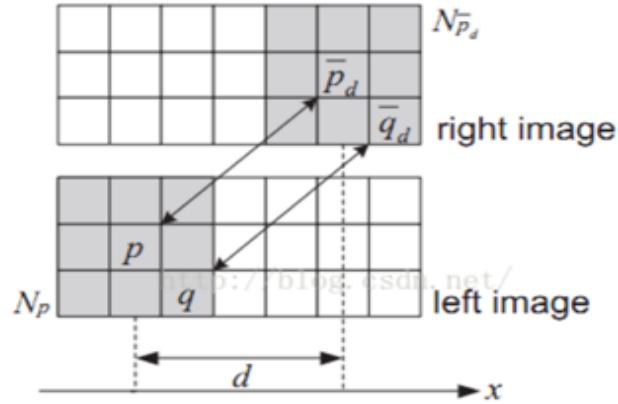


图 4-10 SAD算法过程示意图

(6) 找到同一垂直坐标范围内SAD值最小的窗口，该窗口的中心像素点就是左图像中心像素的最佳匹配，它们的差值就是视差。

(7) 同理执行所有左图像素点，得到所有像素的视差，输出视差图。

用OpenCV函数读入标准图像信息，利用C语言编写代码实现上述逻辑过程，以Tsukuba标准图，构造窗口大小取11为例，实现的SAD算法立体匹配视差图像如4-11所示。

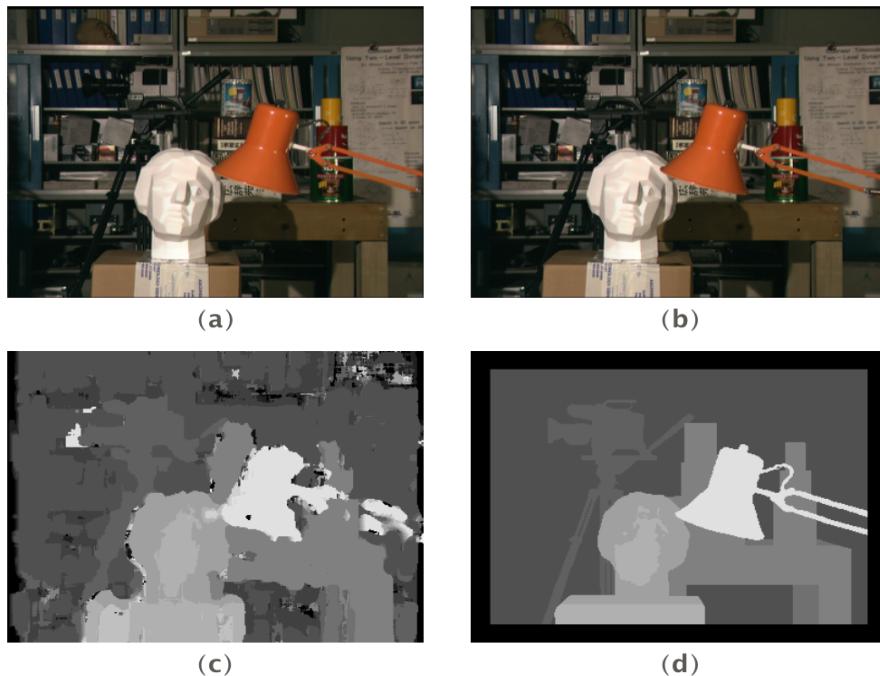


图 4-11 Tsukuba立体匹配效果：(a)左图像 (b)右图像 (c)SAD匹配视差图 (d)标准视差图

4.4 匹配窗口的选择

在研究与应用中发现，基于区域相似度的匹配算法的可靠性会受所选取支持窗口的大小影响：支持窗口越大，匹配信息越丰富，在低纹理区域或重复区域匹配效果越好，但是在细节复杂、视差不连续区域或者闭塞区域误匹配率越高；支持窗口越小，在视差不连续区域和闭塞区域的误匹配越少，但是在低纹理区域信息采集越不充足、误匹配率越高。出现这种矛盾的情况是因为，在匹配过程中需要收集尽可能多的正确有效的参考信息用于相似度测量，为了采集更多参考信息，支持窗口需要增大，但是随着支持窗口增大，支持窗口内的闭塞区域信息也会增多，闭塞区域的信息是无效信息甚至是干扰信息，参考信息可靠性降低将直接影响匹配准确率。

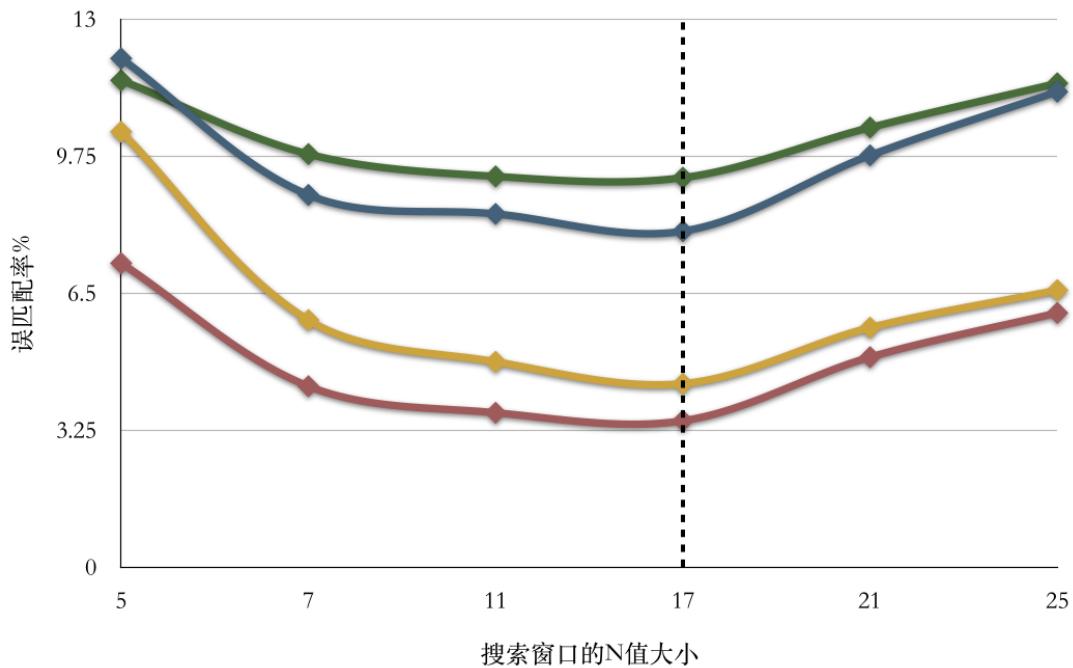


图 4-12 各标准图窗口大小与误匹配率的关系表

利用实现的评估函数算法和标准视差图，我们可以通过输入不同的窗口大小值，得出误匹配率，输入常用的窗口边长大小，进行多次试验，对比各个窗口SAD算法的误匹配率，可以得出其关系如表4-12所示，又该折线图可以判断标准图像的输出视差图在 $N=17$ 时误匹配率最低，当 N 大于8之后误匹配率上升，而且随着选择窗口的变大，算法每一轮循环需要遍历的像素点越多，算法的时间复杂度提高，导致运行时间开始增加，局部匹配的优点之一就是匹配速度较快，综上

可以得出标准图的匹配窗口大小 $N=17$ 时图像匹配效果最好。各标准图在不同窗口大小下的输出视差图效果如图4-13(a)-(d)所示。

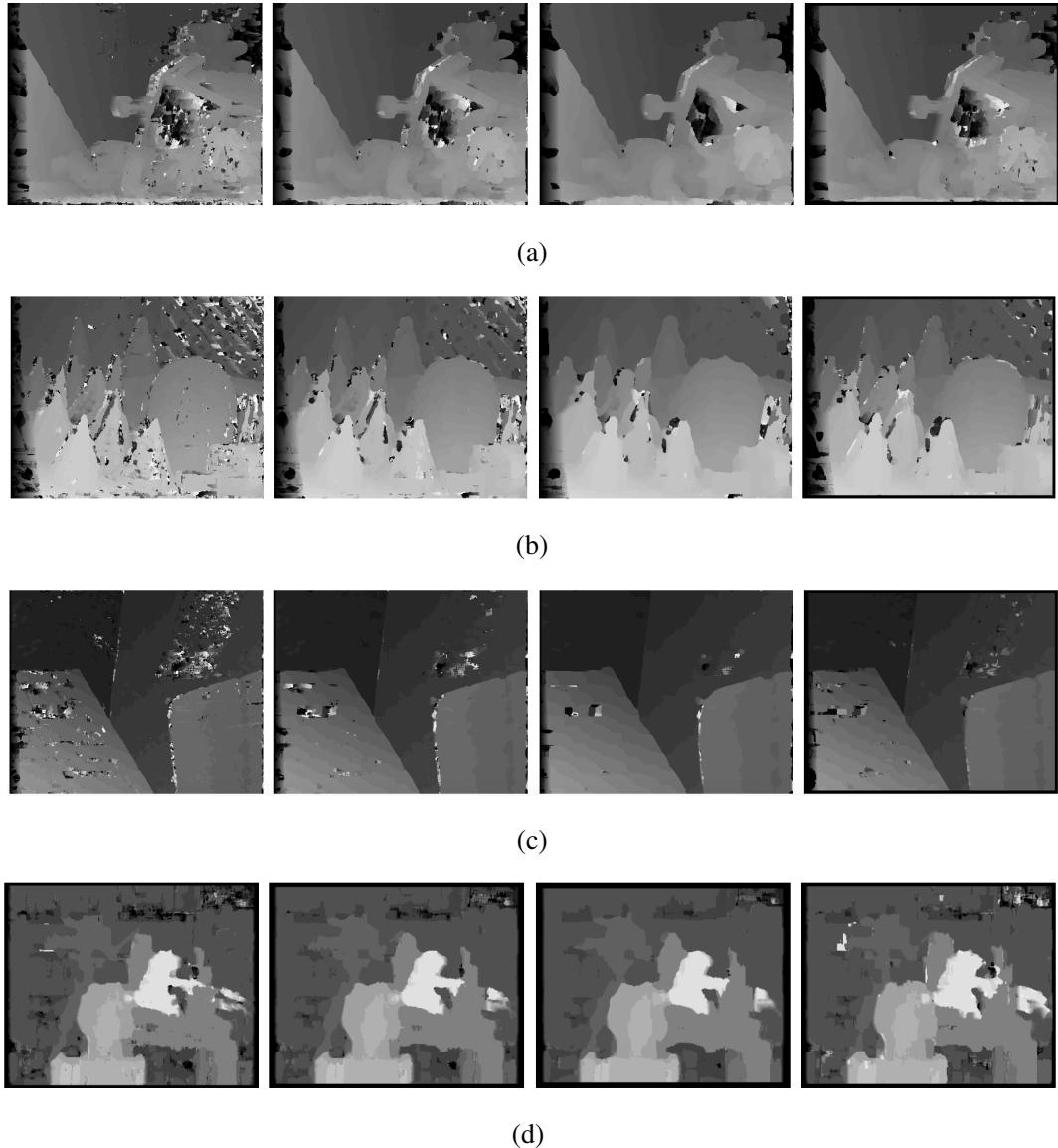


图 4-13 不同窗口大小视差图(a)Teddy;(b)Cones;(c)Venus;(d)Tsukuba

4.5 优化局部立体匹配算法

上文通过调用相关函数库，编写算法逻辑，已经实现了基础的局部立体匹配方法SAD，本节将研究立体匹配过程中的常见难题，如左右输入图像曝光不一致，

导致灰度信息缺乏可信度，遮挡现象影响匹配，中心像素点对匹配窗口内所有像素点依赖一致等等，对这些可能遇到的问题进行算法优化，提高算法的稳定性，降低误匹配率。

4.5.1 增加Census变换

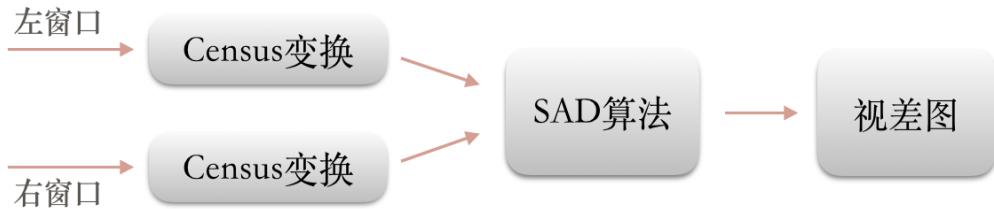


图 4-14 Census 变换示意图1

由于SAD算法的相似性测度函数采用窗口对应点的灰度值直接计算，没有考虑窗口各点与中心点的相关性，因此受窗口大小和环境影响较大。比如左右图像曝光时间或光照条件不同导致的左右图像对应区域灰度值差距大。可以采用Census变换方法，先将窗口内各点的灰度值转换为与窗口中心相关的数值，将绝对的灰度条件，转换为相对灰度条件，以此解决图像的灰度信息不同步的问题，提高精确度。

具体过程是在参考窗口（即左窗）和目标窗口（即右窗）中分别作如下计算：遍历窗口中的每一个像素，若该像素灰度值大于中心像素灰度值，则该像素置 1，否则置 0，如公式 4-3 所示

$$Census(q) = \begin{cases} 1 & I(q) > I(p) \\ 0 & otherwise \end{cases} \quad (4-3)$$

其中， p 为非中心像素， q 为非中心像素。 $I(q)$ 为非中心像素的灰度值， $I(p)$ 为非中心像素的灰度值。如图 4-15 所示， N_L 和 $CensusN_L$ 分别为变换前后的窗口。

在分别对参考窗口 N_L 和目标窗口 N_R 进行 Census 变换之后，就得到了两个仅由 0 和 1 组成的二进制串 N_{L01} 和 N_{R01} ，如图 3.6 所示。由于汉明距离表示两个码串之间对应位置不同码字的位的个数，这样，就可以通过比较 N_{L01} 和 N_{R01} 之间的汉明距离来度量参考窗口 N_L 和目标窗口 N_R 之间的相似度。易知， N_{L01} 和 N_{R01} 之间的汉明距离为 13。同理，在视差值的取值范围内移动目标窗口，分别计算对应的汉明距离，最后再用赢者通吃算法（WTA）筛选出误差值即可。

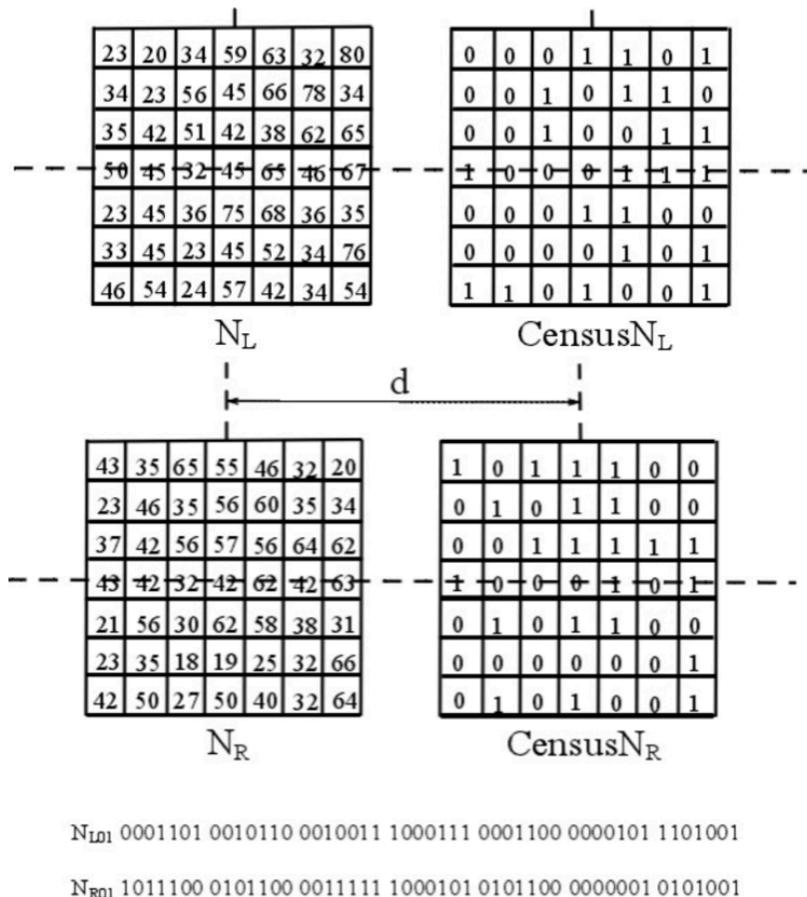


图 4-15 Census变换示意图2

在传统的基于 Census变换的局部立体匹配算法中，在对窗口中的像素灰度值进行 Census变换时参考的是中心像素的灰度值。由于图像中像素的灰度值受很多外界不确定因素的影响，因此，如果中心像素不够准确，则比较结果的误差将会增加。为了减小误差，在改进的基于Census 变换的局部立体匹配算法中，在对窗口中的像素进行 Census 变换时，不直接采用窗口中像素的灰度值直接计算，而是采用像素灰度值相对中心像素的大小来综合对比，这样就可以减少由于窗口曝光的不一致而造成的匹配误差。

以Teddy标准图为例，通过对图像输入灰度值的调整，模拟左右摄像图片曝光不一致的情形，通过对输入像素进行Census变换，转换为二进制串进行对比，经过多次试验，基础的SAD算法误匹配率在15.23%左右，经过Census变换后的误匹配率为12.89%左右，实际的匹配效果如图4-16所示，添加Census变换后立体匹配在面对曝光不一致的情况下噪点情况得到改善，降低了误匹配率。

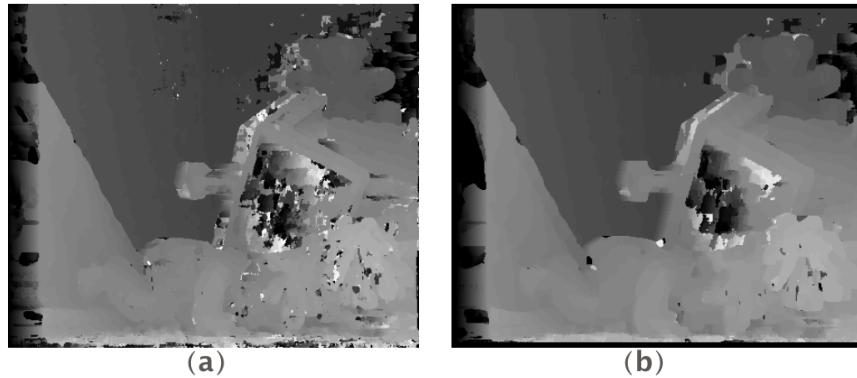


图 4-16 Teddy 匹配效果: (a)SAD 算法 (b)Census 变换优化

4.5.2 双向匹配遮挡检测

立体匹配过程中存在遮挡现象。由于摄像机平移的特点，在相机平移过程中必然会出现前景物体将后景物体边缘遮挡的情形，此时左图像的点在右图中无法找到准确的匹配点，这部分区域就是遮挡区域。

添加遮挡区域检测，以提高图片遮挡区域的匹配准确率。通过对换参考图像和目标图像，进行双向匹配，比较两次输入得到的视差图，视差值大于一定阈值的区域就是遮挡区域。

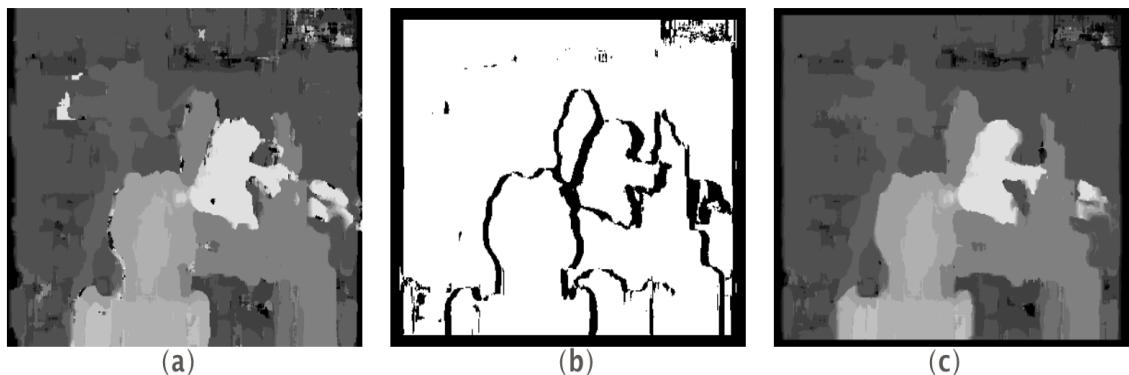


图 4-17 Tsukuba 匹配效果: (a)SAD 方法 (b)检测到的遮挡区域 (c)添加遮挡填充的 SAD

以左图为参考图像，右图为目標图像，即在右图中寻找左图中每一个像素对应的像素点，从而获取左图对应的视差图。但是，这样单向匹配的结果就是，在左图中，有些像素得到了正确的匹配，有些像素得到了错误的匹配。如果匹配的结果是正确无误的，那么不管以左图和右图中的哪一副为参考图像，另外一副为目标图像，那么匹配的结果都应该是一致的，而在图像的遮挡区域，肯定得不到

正确的匹配，因此两次匹配的结果应该是不一致的，这就是双向匹配算法的思想。以此改进算法，添加遮挡检测，并在遮挡区域选择视差较小的背景视差对视差图进行遮挡填充。以Tsukuba标准图为例，实验过程得到的SAD匹配效果，遮挡检测输出的遮挡区域，以及添加遮挡填充的SAD算法如图4-17所示。

分析图4-17中可以得知，在没有添加遮挡检测的SAD算法输出视差图中噪点较多，且大部分落在了遮挡区域内，通过检测遮挡区域，降低了在该区域内误匹配率，从而得到匹配效果更准确的视差图。通过评估函数计算，未优化的SAD算法误匹配率在15.22%左右，添加遮挡检测后降低到13.52%左右。

4.5.3 基于几何距离的自适应权重

在SAD算法中，窗口中每一个像素点计算出来的基元对于窗口中心点来说权重是一致的，这其实并不符合常理，距离窗口中心比较近的区域应该比比较远的地方的像素点信息更有参考意义，如果不通过一些方法去除这种一致性，那么算法的匹配准确度将会非常依赖于窗口大小的选择。

在双目立体视觉中，图像对中的遮挡等闭塞、视差不连续情况是普遍存在的，而且这种存在的可能性随支持窗口的增大而增大，邻域点离中心点越远匹配可靠性和参考价值越小，这种特性也是符合人眼视觉特性的。

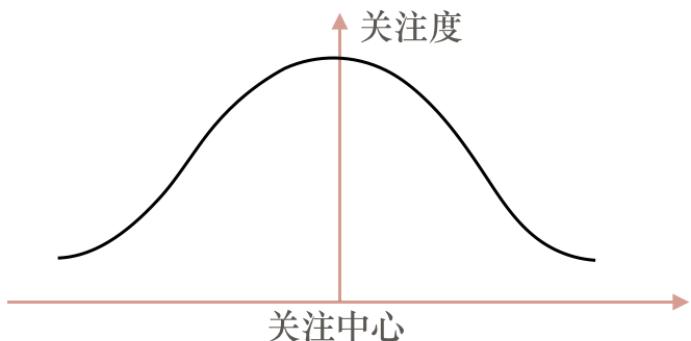


图 4-18 人类双眼关注度与关注中心的相关性示意图

根据邻域点至窗口中心的几何距离分配匹配权重的模型，几何距离权重 f_w 的表达式如公式：

$$f_w(i, j) = a\omega^{f^s} \quad (4-4)$$

式中， (i, j) 为邻域点在窗口中的坐标， $f = \sqrt{i^2 + j^2}$ 是邻域点至窗口中心的几何距离。 a 和 ω 为权重调节因子： a 为幅值调节因子，可避免计算结果下界溢出；

ω 为指数衰减速度调节因子， ω 越大特征曲线越平滑； s 为峰度调节因子， s 越大特征曲线越窄。 s 和 ω 共同作用控制核心区的范围和权重系数，本实验中这些常数因子取常用经验值进行计算。

由于距离窗口中心点距离越远其提供的信息参考价值越小，通过基于几何距离自适应权重，距离中心像素点越远，相似性函数计算得出的结果被权重系数削弱影响就越明显，以此解决窗口内像素点贡献无差异的问题，从而提高低纹理和深度非连续区域的匹配准确性。

在SAD算法公式的基础上，每一个窗口匹配像素点在计算灰度值差值后，乘以基于几何距离计算出来的权重值，然后再保存，最后将窗口所有像素点统计相加，得到最佳的匹配点，输出视差图。本实验以Venus标准图为例子，经过几何距离自适应权重之后的匹配效果如图4-19所示，观察分析可以看出图像噪点明显减少，误匹配的情况有所改善。将输出的视差图分别用评估函数进行计算，得到优化后的匹配算法误匹配率从原来的15.39%下降到了11.77%。

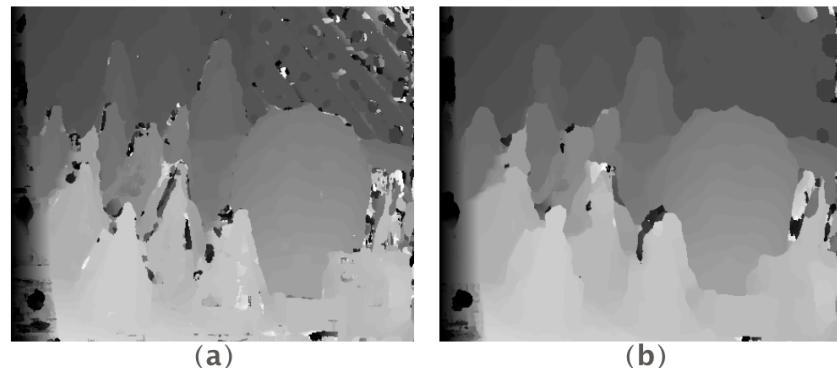


图 4-19 Tsukuba匹配效果：(a)SAD方法 (b)基于几何距离自适应权重

4.6 实验结论及全文最后总结

上一节通过多个角度，针对立体匹配过程中存在的问题，局部匹配方法SAD存在的缺陷进行优化，分析实际的匹配效果以及通过评估函数计算得到的误匹配率变化，验证了优化方法的有效性。综合所有优化过程步骤，即给SAD算法添加Census变换，遮挡检测以及依据几何距离的自适应权重，对输入图像进行立体匹配，得到如图4-20所示的匹配效果，可以看到输出视差图噪点明显减少，在深度不连续和低纹理区域的匹配能力都有所改善。

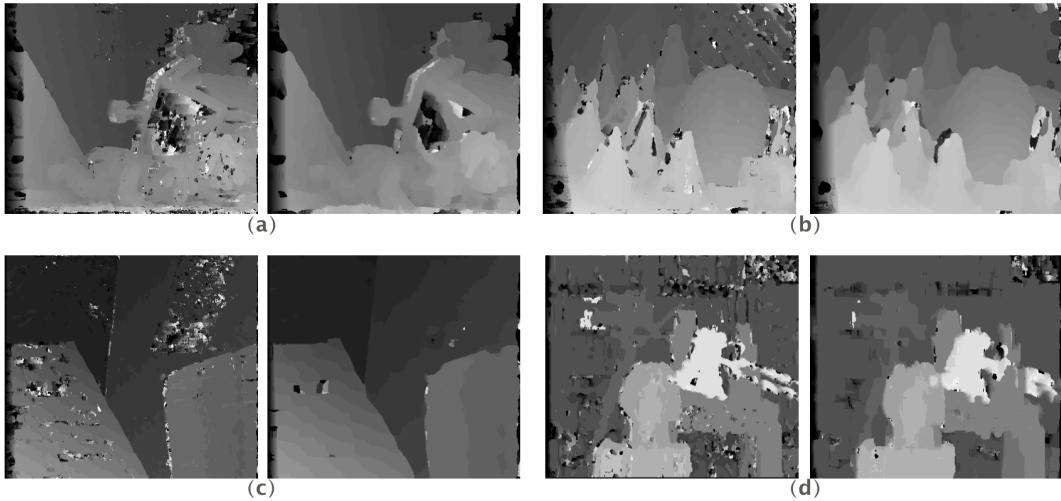


图 4-20 优化效果对比: (a)Teddy (b)Venus (c)Cones (d)Tsukuba

将优化前后的输出视差图，输入到评估函数与标准视差图进行比对，对前后匹配效果进行量化，得到的误匹配率变化如图4-21所示，相比基础的SAD算法，优化后的立体匹配方法在不同的标准图测试过程中都在一定程度上降低了误匹配率，从而验证了优化的真实有效。

标准图	Tsukuba	Venus	Cones	Teddy
SAD算法	8.39%	9.22%	10.34%	12.65%
优化的SAD算法	6.77%	7.03%	8.05%	10.56%
优化程度	1.62%	2.19%	2.29%	2.09%

图 4-21 算法优化前后的误匹配率

至此，实验最终实现了一个双目视觉系统的基本功能，通过标定纠正采集的图像，实现了基于局部立体匹配的图像匹配效果，通过多种方式优化局部匹配方法，利用评估函数对比算法匹配效果，最终得要一种匹配效率较高的立体匹配方法，达到实验目的。

经过本课题的学习研究，了解了一些计算机视觉的基础知识，掌握了一个双目视觉系统的基本原理和实现方法，通过对立体匹配的学习，对匹配难点的思考，针对匹配过程中的常见问题，本课题从多个角度来优化算法，并得到满意结果，提高了匹配准确率。从这个过程中锻炼了思考解决实际问题的能力，完成匹配要

求的同时也提高了编程能力。

计算机视觉领域现在正在快速发展，经过课题的研究学习，我对双目视觉有了粗浅的入门，掌握了基本的三维重建原理。我对计算机视觉的发展前景和应用场景普及十分看好，希望在各领域学者和企业的共同努力下，加速计算机视觉的发展，在多个领域为人类提供便捷的服务。

参考文献

- [1] Marr. Vision Computing Theory.[D]. Higher Education Press, 1988
- [2] 胡汉平. 双目立体测距关键技术研究[D]. 中国科学院, 2014
- [3] 荆鑫. 计算机视觉中双目立体匹配技术的研究[D]. 南京理工大学, 2013
- [4] 裴聪. 基于计算机视觉中双目立体匹配技术研究[D]. 江苏大学, 2010
- [5] 马源. 基于双目立体视觉的深度感知技术研究[D]. 北京理工大学, 2015
- [6] RedMud. 立体匹配十大概念综述-立体匹配算法介绍[EB/OL]. <https://zhuanlan.zhihu.com/p/20703577>, 2016
- [7] Bradski G, Kaehler A. 学习OpenCV[M]. 北京: 清华大学出版社, 2009
- [8] 张正友. A Flexible New Technique for Camera Calibration[D]. 浙江大学, 1998
- [9] 刘波译. 深入理解OpenCV——实用计算机视觉项目解析[M]. 北京: 机械工业出版社, 2014
- [10] Richard, Szeliski. 计算机视觉-算法与应用[M]. 北京: 清华大学出版社, 2012
- [11] 许雯. 立体视觉中局部立体匹配算法研究[D]. 西安电子科技大学, 2014
- [12] 查志强. OpenCV三种立体匹配求视差图算法总结[EB/OL]. <http://blog.csdn.net/zhangzhiqiang/article/details/50595850>, 2016
- [13] D. K. Hoa, L. Dung, N. T. Dzung. Efficient determination of disparity map from stereo images with modified Sum of Absolute Differences (SAD) algorithm[C]. Hanoi University of Science and Technology, Vietnam, 2014
- [14] 于舒春. 双目立体视觉和三维重建关键技术研究[D]. 哈尔滨工业大学, 2008
- [15] 孟晓桥, 胡占义. 摄像机自标定方法的研究与进展[D]. 自动化学报, 2003
- [16] Scharstein D, Szeliski R. A Taxonomy and Evaluation of Dense Two- Frame Stereo Correspondence Algorithms[D]. Springer, 2002
- [17] Humenberger M, Zinner C, Weber M, 等. A fast stereo matching algorithm suitable for embedded real-time systems[D]. CVIU, 2010
- [18] 刘小群. 基于立体视觉的立体匹配算法研究[D]. 中南大学, 2011
- [19] 王逸林. 基于区域匹配算法的双目立体匹配技术的研究[D]. 大连海事大学, 2012
- [20] Barzigar N, Roozgard A, Cheng S, 等. Dense image registration using sparse coding and belief propagation[D]. JVIS, 2012

- [21] Yoon K J, Kweon I S. Locally Adaptive Support-Weight Approach for Visual Correspondence Search[D]. IEEE, 2005
- [22] Z.Lee, J.Juang, T.Nguyen. Local disparity estimation with three-moded cross census and advanced support weight[D]. IEEE, 2013

致 谢

时光飞逝，岁月如梭，转眼间，四年的本科大学生活即将结束，本次论文作为我在大学期间做的最后一次学习研究任务，经历多个月时间的查找资料，学习基础知识，组织实验，撰写论文，最终顺利完成。

值此我首先要感谢我的导师李宏亮教授和研究生孙文龙学长。李宏亮导师专业知识渊博，治学态度严谨，工作一丝不苟、精益求精，具有诲人不倦的高尚师德，孙文龙学长具有和蔼可亲的人格魅力，为我耐心讲解各个知识点，给我解决问题提供了许多思路。

除此之外，能够顺利完成本次论文，也离不开大学四年母校和多位老师对我的指导，培养了我独立思考和自主学习的能力，在此一并感谢。感谢535教研室各位学长的关心与帮助，使我能够顺利完成论文的实验部分，感谢248寝室同学在我论文书写期间对我的支持和帮助。

通过本次论文的学习，除了对本课题有了深刻的理解，对科研流程的初步熟悉之外，查阅大量的文献资料，提取相关内容，锻炼了我自学的能力，组织论文结构，进行课题探讨，设计实验提高了我的思维能力，总体而言，这次论文的完成过程使我各方面得到很大的提升，再次感谢所有给予我帮助的人！

Efficient determination of disparity map from stereo images with modified Sum of Absolute Differences (SAD) algorithm

Dang Khanh Hoa, Le Dung, Nguyen Tien Dzung

This paper proposes modification of the conventional Sum of Absolute Differences (SAD) for performance improvement in depth-map estimation from stereo images captured by a camera in a stereo system. The conventional SAD is commonly search in whole stereo images to find out the difference in pixels between the left and right captured images, and then obtains the corresponding disparity map and this may lead to high elapsing time. In order to reduce the number of searching pixels, the proposed modified SAD tries to estimate the difference only from edge pixels which are referred as pixels-of-interest and bring significant information about depth map. The number of pixels being searched is reduced to about 17% on the total pixels, hence the total elapsing time is saved up to around 89% compared to that of the conventional SAD. This results is promising for implementation of a real-time vision system.

Stereo matching is a problem to find correspondences between two input images,. It is one of fundamental computer vision problems with a wide range of applications, and hence it has been extensively studied in the computer vision field for many recent years . Stereo matching consists to find for each point in the left image, its corresponding in the right one. The difference between horizontal distances of these points is the disparity. A disparity map consists of all the possible disparity values in an image. Such a map is basically a representation of the depth of the perceived scene. Therefore, the disparity maps have been used to address efficiently problems such as 3D reconstruction, positioning, mobile robot navigation, obstacle avoidance and many other domains.

There are three broad classes of techniques, which have been used for stereo matching: area-based, featurebased, and phase-based. SAD-based implementations are the

most favorable area-based techniques in real-time stereo vision, since they can be straightforwardly implemented in hardware. The calculations required in terms of design units are simple, since only summations and absolute values are performed. Parallel design units can be utilized in order to handle various disparity ranges, in order to reduce the computational time required. SAD correlation algorithm can be applied to solve the problem automatically detect mold applications in robot control and mapping. The research results show that the SAD correlation algorithm can be a potential replacement for SIFT method proposed by Lowe in the landmark selection problem. The purpose of the SIFT method is quoted as a full-featured constant rate, rotate, move but criticized the selection of stable characteristics proved more effective when reproducing 3D or service of navigation.

Typically, the disparity is computed as a shift to the left of an image feature when it is viewed in the right image, it is calculated by determining a measure of the SAD, that is used to calculate disparities at each pixel in the right image. After this SAD “match strength” has been computed for all valid disparities, the disparity that yields the lowest SAD is determined to be the disparity at that location in the right image. Hence, the computation of a disparity map is performed on all pixels of stereo images so this approach may be influenced by an object’s position and large elapsing time to determine the corresponding points between the two images, especially for huge size of captured images or frames. In this paper, the proposed SAD algorithm tries to modify the way of disparity determination by utilization of edge information extracted from the captured stereo images. Since edges operators used in the proposed method require fewer computational load, the disparity map can be represented with low elapsing time while maintaining reasonable quality of the map reconfiguration.

The rest of this paper is organized as follows: Section 2 describes modifications in the proposed SAD method to determine a disparity map. In Section 3, experimental results obtained on real images are presented and discussed. Finally, Section 4 concludes the paper with some remarks and future works.

基于优化的匹配算法高效获取视差图

本文提出改进的常规误差绝对和（SAD）算法，从而改进立体相机感应物体深度距离的性能。基础的SAD算法通常在整个立体图像中搜索信息，以找出左右拍摄图像之间的像素视差，然后获得相应的视差图，这可能导致大量的运算时间损耗。为了减少数量搜索像素，建议改进的SAD算法尝试仅仅计算被称为兴趣像素的边缘像素的视差，并提供有关深度图的重要信息。该方法能够使正在搜索的像素数减少占总像素大约17%的像素数，因此与传统SAD相比总时间节省了89%左右。这个结果有望实现实时的视觉系统。

立体匹配是一个关于如何找到两个输入图像之间的对应关系的问题。它是广泛应用的基础计算机视觉问题之一，因此近年来在计算机视觉领域得到了人们广泛的关注和研究。立体匹配目标是为左边图像中的每个像素点，找到它在右图像的对应像素点。这些点的水平距离之间的差值就是视差值，视差图由图像中的所有像素点的视差值组成。实际应用中的地图基本上是感知场景的深度距离，视差图可以用来解决诸如3D重建，定位，移动机器人导航，障碍物回避等许多领域的问题。

如今有三种广泛的技术被用于立体匹配：基于区域的，基于特征和基于相位的立体匹配方法。基于SAD的立体匹配实现是实时立体视觉中最有利的一种基于区域匹配的技术，因为它可以在硬件中直接实现。设计算法很简单，仅执行求和和绝对值。可以使用并行设计单元来达到误差范围，以减少所需的运算时间。可以应用SAD相关算法来解决机器人控制和映射中自动检测模具应用的问题。研究结果表明，SAD算法可以成为Lowe在地标选择问题中提出的SIFT方法的潜在替代方法。SIFT方法的目的是为全功能恒定速率，旋转，移动物体提供匹配方法，但SAD方法被证明在再现3D或导航服务时更有效。

通常将视差定义为以右图像坐标为基准观察到的左图像对应像素点的偏移量，通过确定右图像中每个像素处对应的SAD值来计算视差。在对所有的有效差值进行SAD匹配相似度计算之后，产生最低SAD的像素点将被视为图像中该位置像素点的匹配值。对立体图像的所有像素点执行视差计算，由于该方法可能受到

目标对象的位置和大小的影响，必须先确定两个图像之间的对应点，特别是对于捕获的动态图像或帧的巨大尺寸。

本文所提出的SAD算法尝试通过利用从捕获的立体图像提取的边缘信息，通过这些信息来修正视差值的输出结果。由于在所提出的方法中使用的边缘算子可以减少运算时长，所以可以以较短的时间来得到视差图，同时保持视差图匹配效果。