

摘要

立体匹配是立体视觉中研究最活跃的主题之一，作为计算机视觉的一个重要分支，在三维场景重建、移动机器人、对象识别、智能控制、三维测量等领域得到了广泛应用。

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。在以上四个步骤中，立体匹配是最重要也是最困难的一步。一般地，立体匹配算法分为两大类：局部立体匹配算法和全局立体匹配算法。通过对这两种算法的优缺点，本文将重点研究和讨论局部立体匹配算法。

本文在经典局部匹配算法 SAD 的基础上，通过分析算法的不足，进行了算法优化，提出了基于遮挡检测的改进的SAD算法。通过对输入图像进行自左向右和自右向左的双向匹配得到的左图和右图对应的两幅视差图进行遮挡检测，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效果。

为了验证算法的有效性，以 Middlebury 网站提供的标准测试图像作为输入，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 an important branch of computer vision, stereo matching is one of the most active stereo vision research theme 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. In general, the stereo matching algorithm is divided into two categories: local stereo matching algorithm and global stereo matching algorithm. Through comparative analysis of the advantages and disadvantages of these two algorithms, this paper will focus on and discuss local stereo matching algorithm.

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 matching the input image from left to right and from right to left, we can get the disparity map of the left image and the disparity map of the right image. 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.

In order to verify the effectiveness of the algorithm, with the standard test images provided by Middlebury website as input, 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

ABSTRACT

quality disparity map.

Keywords: Stereo vision, Local Matching Algorithm, improve Algorithm

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	2
1.2 双目视觉系统简介	3
1.3 立体匹配技术简介	3
1.4 本课题主要研究内容	4
1.5 本文结构安排	4
第2章 双目视觉系统	5
2.1 双目视觉系统概述	5
2.2 双目视觉系统的基本原理	5
2.3 双目视觉的数学模型	7
2.4 摄像机的标定	9
第3章 立体匹配方法及优化	10
3.1 立体匹配方法简介	10
3.2 立体匹配方法的分类和基本过程	11
3.2.1 基于特征的立体匹配方法	12
3.2.2 全局立体匹配方法	13
3.2.3 局部立体匹配方法	13
3.3 立体匹配的难点和算法优化方向	15
第4章 双目视觉系统的构建及实验总结	17
4.1 摄像机离线标定	17
4.2 实验过程匹配效果的评估	19
4.3 局部立体匹配SAD算法实现	21
4.4 匹配窗口的选择	22
4.5 优化实验算法	23

目录

4.5.1 增加Census变换	23
4.5.2 双向匹配遮挡检测	24
4.5.3 基于几何距离的自适应权重	25
4.6 实验结论及全文最后总结	26
参考文献	28
致 谢	29
外文资料原文	30
外文资料译文	32

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

人类和动物通过自身的视觉系统，可以获得外界物体的形状大小、明暗亮度、行为变化、位置、颜色等各种对交际发展，机体生存具有重要意义的信息。据统计至少有80%以上的外界信息经由视觉系统获得，由此可见视觉是人和动物最重要的感觉。随着科学技术的不断发展，视觉理论框架的提出和完善，计算机硬件和摄像机采集能力的提高，人们逐渐开始通过利用计算机及相关的其他设备模仿人类的视觉系统，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，实现机器对客观世界中的三维场景进行感知、识别和理解，这就是计算机视觉。

人类视觉的主要功能可概括为：适应外界环境，控制自身的运动。相应的，计算机视觉系统需要：能够准确识别物体，确定物体的位置和形状以及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极其富有挑战性的重要研究课题和研究领域。

当我们对现实中的某个物体进行肉眼的识别时，人类的大脑需要对这个物体存在某种表达。目前，计算机视觉界主要存在两种主流的表达理论。一种是上世纪八十年代初马尔（Marr）提出的三维表达理论（三维重建理论），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：物体是以其几何形状的方式在大脑中被表达的。由于物体的三维形状和观察角度无关，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，随着神经科学研究的发展，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，对三维世界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

计算机视觉学作为一门新兴学科，自二十世纪中期开始迅猛发展，成为当时最为热门的研究主题之一。计算机视觉使用计算机及相关的其他设备对生物视觉进行仿真和模拟，它的重要功能和主要任务是通过对采集的图片或视频进行处理以获得相应场景的三维信息，就像人类和许多其他类生物每天所做的那样。由于计算机视觉在工农业生产、地质学、虚拟现实、航空航天、天文学、气象学、医学及军事并学等领域都发挥了很大的作用以及扮演着非常重要的角色，因此越来越受到学者们的重视。

如今，经过众多学者的不懈努力，已经形成了一整套系统的理论与算法，并且得到了迅速的发展和广泛的研究。计算机视觉技术已经具备了通过计算机来对数字图像信息进行各种各样地处理和加工的功能。集数字图像处理、光学、几何学、物理学、模式识别及数字信号处理等知识于一体，计算机视觉技术作为一门迅速发展起来的科学领域，计算机视觉已经涉及到图像处理、计算机图形学、计算几何等诸多应用领域。

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解决部分。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。在立体视觉技术中，双目立体视觉技术就是主要针对以上问题的一个专门领域，通过对真实场景不同角度的图像的分析和匹配，从而获取真实场景中物体的视差信息，最后通过物体的视差信息和深度信息的关系，计算出物体的深度信息，为进一步的三维重建、机器人导航等热门应用领域的应用做好准备。因此，无论是从视觉生理的角度还是在工程应用中，双目立体视觉的研究都具有十分重要的意义。

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。按照提取的视差图的稀疏程度，立体匹配方法主要有基于灰度信息的匹配算法和基于特征信息的匹配算法。

1.2 双目视觉系统简介

双目立体视觉(Binocular Stereo Vision)，简称双目视觉，是计算机视觉的一项重要研究内容。双目立体视觉技术模拟人眼视觉特性，从两个不同的角度观察同一个场景的同一目标，分别获得其投影图像，匹配对应点并获取偏差信息，根据三角测量原理，利用该偏差信息计算对应点的距离信息，从而获取景物的三维信息。

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。

1.3 立体匹配技术简介

立体匹配是由场景中同一物体的两幅或多幅图像获取物体三维几何信息的一种技术。它是这样的一个过程：以三维场景中同一物体不同角度拍摄得到的两幅图像作为输入，通过图像配准，寻找真实三维场景中某一物体在两幅图像中对应的像素点的过程。立体匹配通常包括匹配代价计算、匹配代价聚集、视差最优化计算和视差细化四个步骤。

按照提取的视差图的稀疏程度，立体匹配方法主要有基于灰度信息的匹配算法和基于特征信息的匹配算法。根据匹配基元的不同，主要分为基于特征，基于相位，基于区域匹配多个不同方向。基于特征的立体匹配方法受图像特征影响大，并且只能获得稀疏的视差图，匹配的准确度较差；基于相位的匹配方法算法复杂度较大，匹配时间过长，不适用于需要快速匹配的场景；根据优化规则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。

立体匹配算法必须解决以下三个问题：选择正确的匹配特征，寻找特征间本质属性及建立能正确匹配所选特征的稳定算法。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，文章简要分析了匹配算法的分类，较详细地综述了匹配算法在实时性和准确性两个方面的进展。

1.4 本课题主要研究内容

本课题研究计算机视觉领域中的关键问题：双目视觉立体匹配方法。基于双目视觉系统成像原理，要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，核心是研究立体匹配方法的实现。

局部匹配算法是以两幅图像对应像素点的邻域为参考单元进行相似度匹配的方法，局部匹配算法既能得到稠密的视差图，又能极大地缩小搜索范围和运算时间。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，最后通过摄像机采样实际图片作为输入，检验匹配效果。

1.5 本文结构安排

文章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，评估匹配函数，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，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：双目视觉系统的基本框架和实现方法，立体视觉标定，讨论离线的摄像机标定技术，以及畸变模型及参数的计算，设计实验验证相应算法的可行性。

第3章：介绍多种立体匹配方法，分析对比其各自的实现难点，着重解决的图像匹配问题，匹配效果和实际应用的可行性。实现立体匹配基础算法，针对不同的图像情况，讨论多种可执行的优化方案，评估对比不同方法的匹配效果，提高立体匹配算法的匹配精确度和执行速度。

第4章：构建双目视觉系统，采用实际场景的图像对前面研究的算法匹配效果进行验证。最后，总结和展望，对全文的工作进行总结，提出未来进一步改进和研究的方向。

第2章 双目视觉系统

2.1 双目视觉系统概述

人类观察外部世界，能够感知物体的立体性，分辨所观察的场景的深度以及景物的远近，人眼获取的场景信息是立体信息。双目立体视觉系统是立体视觉最常见的系统，是一种模拟人类双眼视觉观察特性的计算机视觉系统。利用双目立体视觉系统，能够实现对待测场景目标的距离感知与被动测距，以及实现目标场景的三维信息恢复与重构。双目立体视觉系统是，通过借助摄像机获取的图像信息，计算世界坐标系下物体的位置和形状等信息。

单幅图像只能够提供图像中每单个点的视线，即垂直平面的世界坐标，不能得到在该视线上世界坐标系中相应的三维空间点与探测垂直平面的距离，因此单幅图像所获取的信息只能提供世界坐标系中某平面上二维测量，无法重构场景的三维几何信息，这是将3D的空间转换成2D的图像发生的信息损失。双目立体视觉技术是如何来恢复这些三维信息，从而重构三维场景呢？通过采集同一时刻，同一位置两幅二维图像的图像信息，利用立体匹配技术，结合视差测距，摄像机标定等先验知识，获取物体的三维表面形状信息的技术，其工作原理主要是基于人类双眼视觉中视差测距的原理。

双目立体视觉实现三维重建的主要包括如下步骤：第一步，摄像机标定，获取成像模型所需内外部参数；第二步，用标定好的摄像机系统拍摄目标场景图像，采集二维信息；第三步，双目立体匹配，寻找两幅二维图像中待测点的点对点对应关系，若为稠密匹配则进入第四步，若为稀疏匹配则进入第五步实现稀疏点测距；第四步，生成视差图；第五步，根据视差计算深度，实现场景的三维重建。按照功能可以将双目视觉系统分为四个主要的模块，如图2-1所示。

2.2 双目视觉系统的基本原理

人类获取目标深度信息的原理如图2-2所示。图中，L和R分别代表人的左眼和右眼以及两眼的视网膜图像，B为两眼间的瞳距(一般人眼的瞳距约为6厘米)，P为观察的目标。由于瞳距的关系，两只眼睛同时观察同一目标，由于两只眼睛的



图 2-1 双目视觉系统的主要模块

观察角度不同，所以会有左眼看到左边多一些、右眼看到右边多一些的情况。若两眼同时观察目标点P，则点P分别投影成像于两眼，这两个投影是有位置差异的，这个差异就是视差，场景中所有点的视差组成一幅图像，称为视差图。大脑对视差图进行处理，可以产生对场景的立体感知。人眼视觉的这个立体感知的过程就是基于视差的原理，视差原理也是计算机视觉的原理基础。

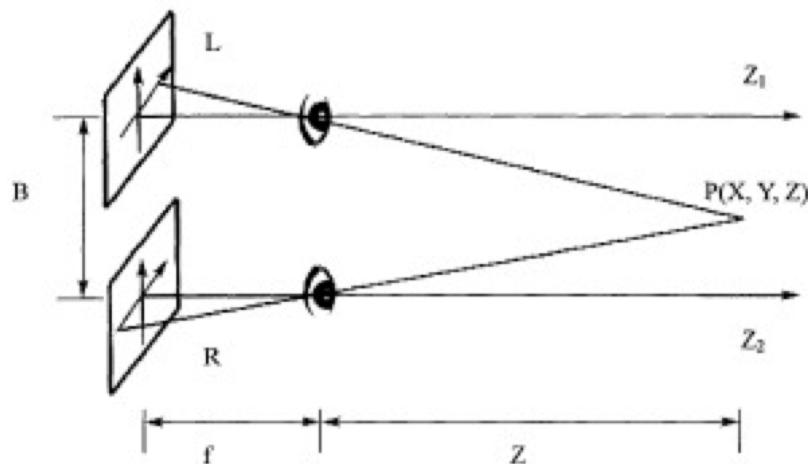


图 2-2 人类视觉感知深度的原理示意图

基于人类视觉特性的研究是立体视觉技术的重要研究方向，趋近于人类视觉能力是立体视觉技术的研究目标。目前，在计算机立体视觉技术的研究中，基于单目多焦距图像序列的立体视觉技术、基于人类感知特性的立体视觉技术以及基于双目成像的立体视觉技术都有研究人员从事相应工作。根据环境条件需求和重构效果需求的不同，选择的立体视觉技术也不同。其中，双目立体视觉技术是最接

近人类视觉系统、适应于机器实现且重构效果较好的立体视觉技术。

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，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，对于左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刻得出物体深度变化信息。而左右摄像机匹配对应像素点的过程就是立体匹配方法，将在下一章详细介绍。

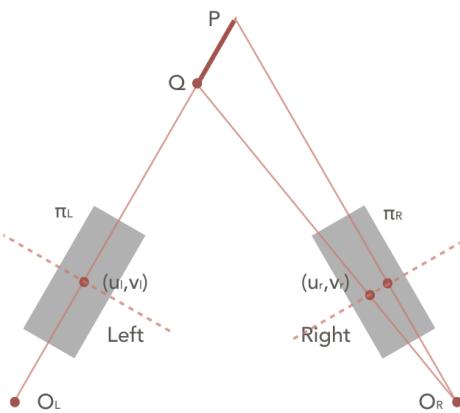


图 2-3 双目摄像机获取像素点信息的示意图

2.3 双目视觉的数学模型

为了便于理解和计算方便，通常成像面翻转至视点与目标点之间，这样视点和目标

点连线与两成像面的交点即为目标点在成像面的像，如图2-4所示为经翻转成像面的双目测距原理示意图。调节与标定完成后，两台摄像机同时对焦距范围内待测场景目标P拍摄，分别获得 L 和 R 两幅图像，两幅图像在同一个与Z轴垂直的空间平面，即 $Z_L = Z_K$ ，两幅图像中P的对应点(即P在两成像平面的投影)坐标分别为 (X_L, Y_L) 和 (X_R, Y_R) ，首先以X轴方向为例推导目标距离，令 $Y_L = Y_R$ 。假定世界坐标系与左成像图坐标系重合，则P点与左摄像机光轴距离为X，与摄像机光轴距离为 $(B - X)$ 。

根据相似三角形计算原理，可得公式2.1

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

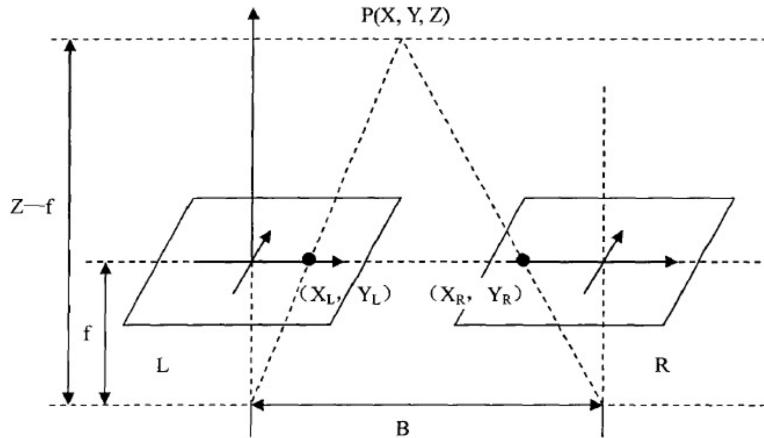


图 2-4 双目测距原理图

公式中，目标距离Z和目标横坐标X是未知数，公式变换可得公式2.2

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

由此可解得目标横坐标X，如公式2.3所示

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

再将目标横坐标X(即公式2.3)代入公式2.1得公式2.4

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

令 $X_L + X_R = D$ ，经公式变化可得公式2.5

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

在公式2.5中，像距f、轴间距B是系统参数，可以通过测量获得，视差 $D = X_L + X_R$ 可以在两图像中找取对应点坐标计算得到。f、B、D已知，即可计算得目标点P与成像平面之间的距离Z，将距离Z(即公式2.5)代入公式2.1，可得目标横坐标X如公式2.6所示

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

同理，目标纵坐标Y如公式2.7所示

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

公式(2.5) (2.6) (2.7)组成空间点P的坐标。这就是双目立体视觉测距的基本原理。该方法实现点测距的前提是已准确完成对应点匹配，通过两对应点的坐标实现点

测距。当待测场景的大多数点完成测距，以其三维坐标构建空间坐标系，即可实现三维重建

2.4 摄像机的标定

双目视觉系统根据摄像机获取的图像信息计算出待测点的深度信息，重建目标场景的表面三维形状与空间位置。三维空间点与其在两二维成像平面中的投影点是对应的，该对应关系取决于摄像机参数，这些参数需要通过实验和计算来确定，该过程就称为摄像机标定。实现三维重建首先需要采集准确的目标场景图像，摄像机标定正是获取准确图像的必要过程。

摄像机的标定最初在摄影测量学中研究和发展，摄影测量学提出了四个标定问题：

- (1) 绝对方位：确定两个坐标系统之间的转换或位置和方向的一系列传感器在绝对坐标系的坐标点的校正。
- (2) 相对方位：确定了相对位置和方向在两个摄像头从预测的校准点在现场。
- (3) 外部方位：确定位置和方向，相机在绝对坐标系统的投影在现场校准点。
- (4) 内部方位：确定内部几何相机，其中包括摄像机常数，位置的主要观点，并更正为透镜畸变。

这些标定问题源于由航拍照片创建地形地图的技术，是经典的摄影测量学问题。除了这四个基本标定问题，摄影测量学也涉及双目立体场景差异深度确定和立体影像重建。

摄像机标定包括摄像机成像系统内外几何及光学参数的标定和多个摄像机之间相对位置关系的标定，其目的是确定目标物体世界坐标与图像坐标之间的映射关系。摄像机标定的精度直接关系到计算机视觉系统的系统误差，是决定三维重建精度的关键因素，非常具有研究和改进意义。

图像上每点的坐标与空间中物体表面对应点的实际位置相关，这种位置上的映射关系是由摄像机的成像几何模型所唯一确定的。一般把这个几何模型的参数称作摄像机参数，计算这些参数的过程就是摄像机定标的过程。作为双目立体视觉技术赖以计算的基础，摄像机标定的精度直接影响整个系统的测量效果。

第3章 立体匹配方法及优化

3.1 立体匹配方法简介

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。

从根本上讲，立体匹配算法主要是通过建立一个能量代价函数，通过此能量代价函数最小化来估计像素点视差值。立体匹配算法的实质就是一个最优化求解问题，通过建立合理的能量函数，增加一些约束，采用最优化理论的方法进行方程求解，这也是所有的病态问题求解方法。现实生活中，对物体进行拍摄时难免受到某些不确定环境因素干扰的影响，在不同角度下拍摄的同一场景的图片会有所不同。进行立体匹配时难以获得具有很高准确度的结果。为了最大限度地减小这些不利因素的影响，选择恰当的匹配算法是立体匹配技术的重点。

经过多年的研究和发展，已经出现许多各具优势的匹配算法，但是这些算法都不外乎又几个常见的部分构成。立体匹配方法通常可以分为匹配代价计算，匹配代价聚集，视差最优化计算，视差细化四个部分。

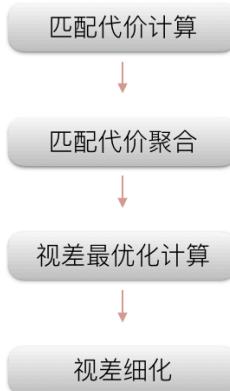


图 3-1 立体匹配常见算法分类

匹配代价计算：这一步最常用的方法有基于灰度值差值的绝对值的匹配代价计算以及基于灰度值差值平方的绝对值的匹配代价计算，还有基于可变权值的灰度值差值的绝对值的匹配代价计算。

匹配代价聚集：通过累加邻域内的匹配代价或取其平均值来代替单个参考像素的匹配，就是匹配代价聚集，通常用这种方式来增加匹配结果信噪比。局部匹配方法主要通过代价聚集步骤来提高最终的匹配质量。近来，越来越多的全局匹配方法也利用代价聚集步骤来提高匹配质量，并且已经达到了不错的效果。聚集区域通常是具有固定视差的二维区域。窗口可以使用固定窗口、可变窗口、自适应形状窗口等。

视差最优化计算：这一步根据上一步像素的匹配代价聚集结果来为每一个像素确定其视差值。局部匹配方法在视差计算步骤使用的方法很直接，只是简单的为每一个像素选择使得像素匹配代价最小的视差作为像素的最终视差，即赢者通吃算法(Winner Take All)。但是，这个方法有一个缺陷，那就是目标图像中的像素点可能被多次选作匹配点，因此可能造成很多的误匹配。

视差细化：这一步的目的是为了进一步降低生成的视差图的误匹配率，从而整体提高匹配视差图的质量。常用的一种方法是使用中值滤波，消除图像中的毛刺点，或者使用中间像素视差精炼法，通过曲线来拟合视差变化，从而消除可能的毛刺点。此方法适用于图像中像素灰度平滑变化的区域。

3.2 立体匹配方法的分类和基本过程

立体匹配方法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，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，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，可分为稠密匹配和稀疏匹配。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。

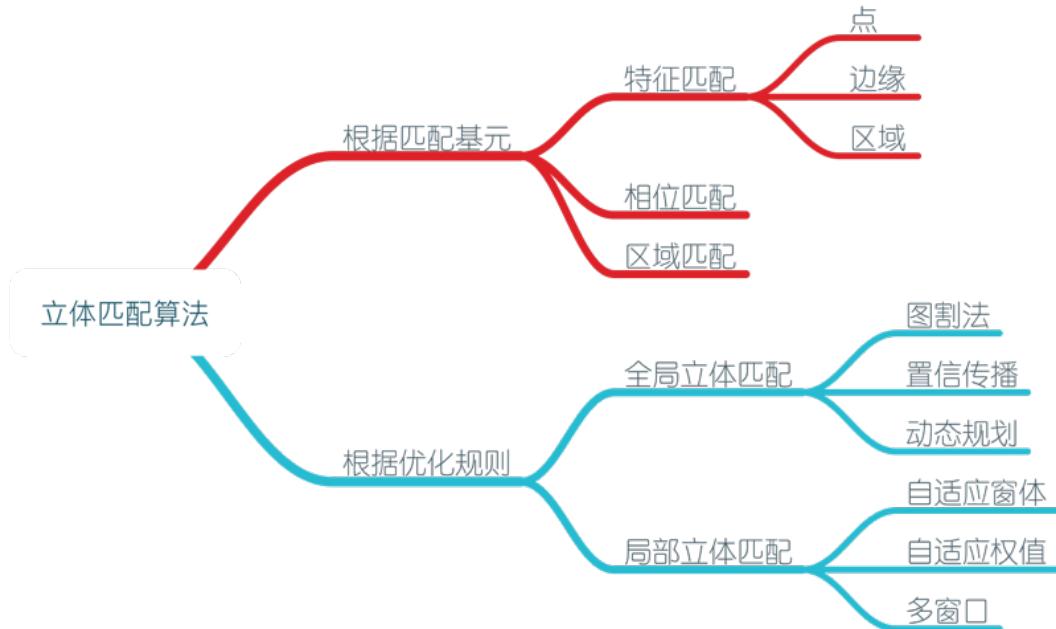


图 3-2 立体匹配常见算法分类

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。特征的匹配算法，主要是基于几何特征信息（边缘、线、轮廓、兴趣点、角点和几何基元等），针对几何特征点进行视差估计，所以先要提取图像的特征点，进而利用这些特征点的视差值信息来重建三维空间场景。

匹配所需要的主要步骤：图像预处理、提取特征、特征点的匹配获取稀疏视差图，如果想得到稠密的视差图，需要采用插值的方法。可获得稀疏的视差图，

经差值估计可获得稠密视差图。可提取点、线、面等局部特征，也可提取多边形和图像结构等全局特征。缺点是特征提取易受遮挡、光线、重复纹理等影响较大；差值估计计算量大，对特征不明显的图像匹配准确度差，而且特征提取的过程也会影响匹配的速度。

3.2.2 全局立体匹配方法

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，采用全局的优化理论方法估计视差，建立全局能量函数，通过最小化全局能量函数得到最优视差值。全局立体匹配算法将待匹配点的匹配代价函数和相邻像素之间由不同视差造成的相互影响关系函数组成。这种平滑函数反映在立体匹配上就是上的立体匹配相关约束条件。

全局匹配算法得到的结果比较准确，但是其运行时间比较长，不适合实时运行，近年来随着计算机计算能力的大幅上升有所改善。主要的算法有图割（graphcuts）、信念传播（belief propagation）、动态规划等算法。

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稀疏视差图，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。

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。除基于窗口的局部匹配算法以外，近年来也出现了一些其他的以基于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。此外还有基于自适应权值的局部立体匹配算法，这一类算法表面采用固定的

匹配窗口进行匹配，实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。

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。

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距离的权值），这样，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家全胜(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：

局部算法复杂度低：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。

局部算法匹配效果越来越接近全局算法：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，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。

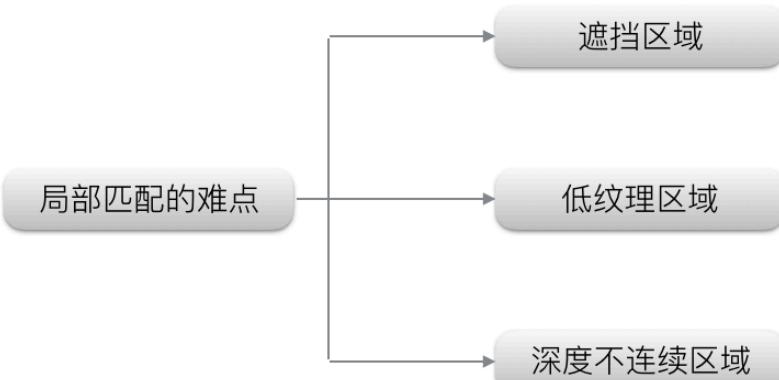


图 3-3 局部匹配的难点

遮挡区域：由于遮挡等因素，某些区域的像素点不可能同时出现在输入的两幅图像中，也就不可能找到对应的匹配点。遮挡是立体匹配中面临的困难之一，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。遮挡像素的检测，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。

低纹理区域：该区域内的像素高度相似，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，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。在匹配的过程如果窗口的过大导致在低纹理区域匹配过程中产生多个匹配候选点，通过对匹配窗口的筛选来改善匹配效果。

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，匹配时很容易产生边界不清晰问题。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，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匹配功能并将优化局部立体匹配算法的匹配结果进行分析比较。

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，接下我们将在实验中实际搭建一个双目视觉系统，并实现一种基础的立体匹配方法，通过不断优化这个算法提高匹配的精确度。

在本实验中，立体匹配方法实现主要是利用了计算机视觉函数库OpenCV，用C语言实现算法逻辑。

4.1 摄像机离线标定

摄像机标定结果将用来纠正输入的左右图像，这对于后面立体匹配的匹配效果影响很大。实验使用OpenCV实现经典的棋盘法标定，棋盘如图4-1所示。

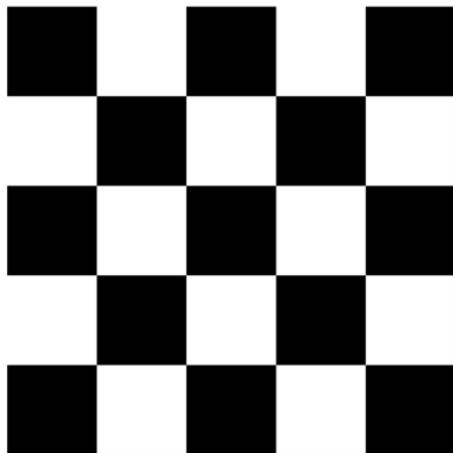


图 4-1 实验棋盘图

实验之前必须明确标定的三个问题：

相机标定的目的：获取摄像机的内参和外参矩阵（同时也会得到每一幅标定图像的选择和平移矩阵），内参和外参数数可以对之后相机拍摄的图像就进行矫正，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：摄像机的内参、外参数数。

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：

1. 准备标定图片
2. 对每一张标定图片，提取角点信息
3. 对每一张标定图片，进一步提取亚像素角点信息
4. 在棋盘标定图上绘制找到的内角点（非必须，仅为了显示）
5. 相机标定
6. 对标定结果进行评价
7. 查看标定效果——利用标定结果对棋盘图进行矫正

双目相机确定好后，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。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)和图4-2(b)所示：

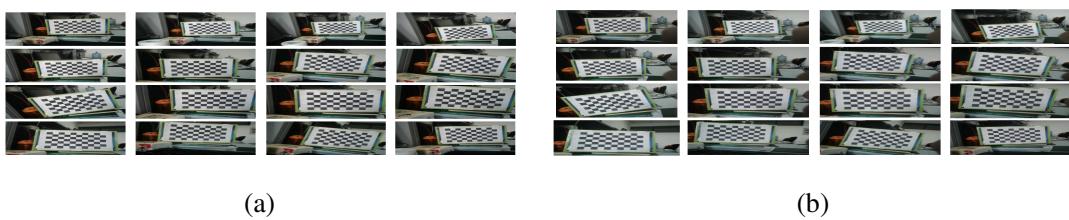


图 4-2 采集的图像(a)左图像;(b)右图像

将图像提取焦点后标定，得出的标定结果如图4-3(a)和4-3(b)所示。

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4-4和图4-5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。

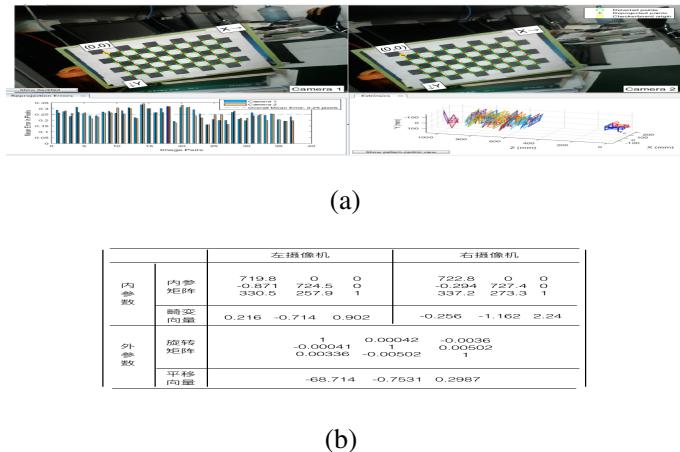


图 4-3 标定结果(a)标定结果;(b)内外参数记录表



图 4-4 标定前左右图像位置

4.2 实验过程匹配效果的评估

在之前的很长一段时间里，立体匹配领域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。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。

常用的度量方法包括 RMS (root-mean-squared) 和 POBMP (Percentage of bad matching pixels)

本实验的评估函数使用 POBMP 度量标准，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：

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



图 4-5 标定前后左右图像位置

其中, $d_c(x, y)$ 是计算出来的视差图中像素 (x, y) 的视差图, $d_t(x, y)$ 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的匹配效果图和实际匹配结果, 计算得出误匹配率。

Daniel 等人不仅给出了立体匹配算法的度量标准, 更给出了用于测试算法的标准测试图像序列, 如图4-6。实验过程中将先采用Middlebury网站提供的标准左右图, 并将匹配结果与网站提供的标准视差图对比评估。

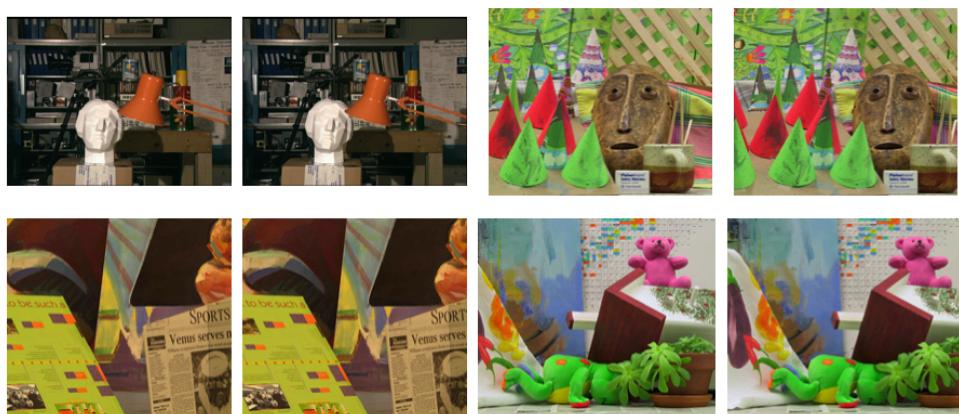


图 4-6 四种标准图像

除了纠正过的标准左右图像, 还有图像的遮挡区域, 视差图等, 以网站提供的Tsukuba图为例, 如图4-7所示, 起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 Tsukuba图像信息

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

本节将以经典的局部立体匹配方法SAD为例，展开介绍一次完整的局部立体匹配过程。

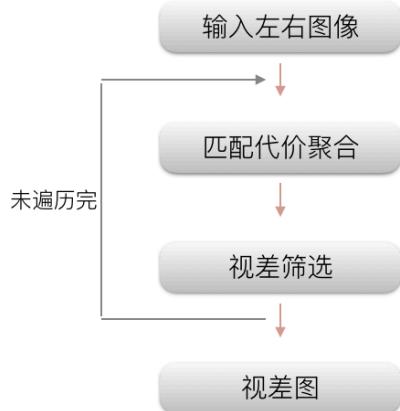


图 4-8 局部匹配算法流程图

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)，相似的还有平方误差和(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算法基本流程：

- 构造一个小窗口，类似与卷积和。
- 用窗口覆盖左图像，选择出覆盖区域内的所有像素点。
- 用窗口覆盖右图像，选择出覆盖区域内的所有像素点。

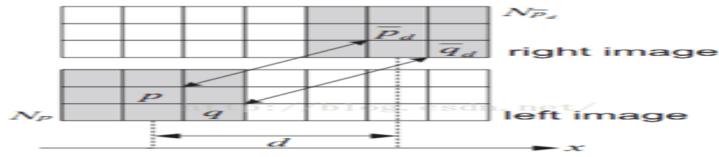


图 4-9 SAD算法过程示意图

- 左覆盖区域减去右覆盖区域，求出所有像素点差的绝对值之和。
- 移动右图像的窗口，判断是否超出搜索范围，否则继续重复上两步。
- 找到这个范围内SAD值最小的窗口，即找到了左边图像的最佳匹配。

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8为例，输出视差图像如图4-10所示。

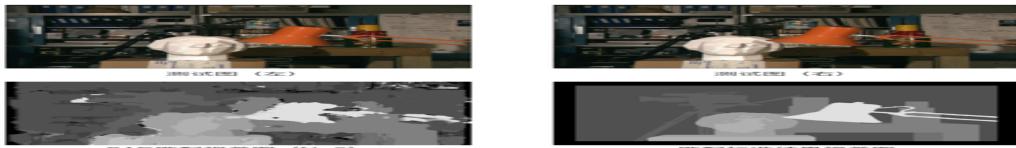


图 4-10 Tsukuba经过实验算法匹配输出视差图

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。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

窗口大小	3	5	8	10	12	15
Tsukuba	12.08	8.84	8.39	9.07	10.02	10.29
Venus	7.22	4.30	3.67	3.68	4.99	5.04
Cones	10.34	5.87	4.88	4.76	4.69	4.58
Teddy	11.56	9.81	9.28	9.25	9.44	9.89

图 4-11 各标准图窗口大小与误匹配率的关系表（单位：%）

窗口大小	3	5	8	10	12	16
Tsukuba	1.34	4.84	12.5	64.2	259	646
Venus	2.22	6.09	17.8	88.2	325	824
Cones	1.78	5.67	15.2	89.3	398	859
Teddy	1.87	6.34	16.6	96.2	377	903

图 4-12 各标准图窗口大小与匹配速度的关系表（单位：s）

同的窗口大小值，得出误匹配率，多次试验可以得出其关系如表4-11所示，匹配算法运行时间如表4-12所示。

在窗口大小为N=8时匹配误匹配率较低，当N大于8之后误匹配率上升或者变换不大，而且算法运行时间开始大幅增加，所以可以得出标准图的整体最佳匹配窗口大小为8

4.5 优化实验算法

针对匹配过程中可能遇到的问题，通过对SAD算法进行优化以提高算法的匹配精准度。

4.5.1 增加Census变换

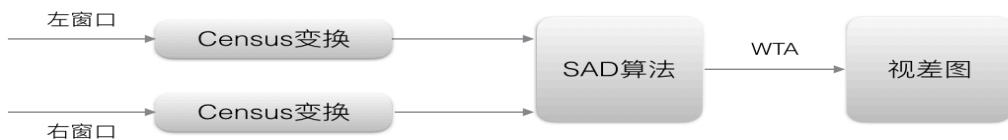


图 4-13 Census 变换示意图1

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, q 为非中心像素, q 为中心像素。 $I(q)$ 为非中心像素的灰度值, $I(p)$ 为中央像素的灰度值。如图 4-14 所示, N_L 和 $CensusN_L$ 分别为变换前后的窗口。

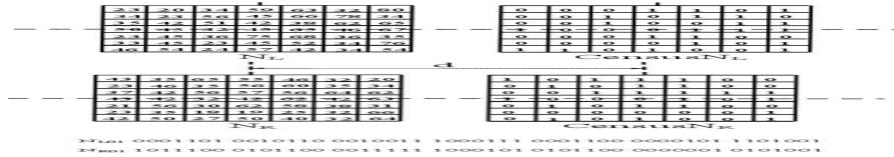


图 4-14 Census 变换示意图2

在分别对参考窗口 L_N 和目标窗口 R_N 进行 Census 变换之后, 就得到了两个仅由 0 和 1 组成的 01 串 L_{01N} 和 R_{01N} , 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 L_{01N} 和 R_{01N} 之间的汉明距离来度量参考窗口 L_N 和目标窗口 R_N 之间的相似度。易知, L_{01N} 和 R_{01N} 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。

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, 不直接采用窗口中心像素的灰度值, 而是采用整个窗口中所有像素灰度值的平均值作为比较对象, 这样就可以减少由于中心像素的不准确而造成的匹配误差。

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 4-15 遮挡区域示意图

以左图为参考图像，右图为目标图像，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。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。以Tsukuba标准图为例，如图4-16所示。

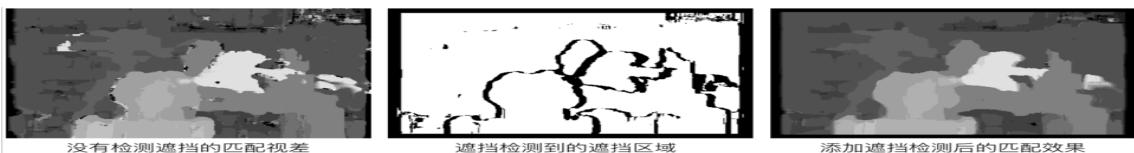


图 4-16 遮挡检测前后匹配效果对比

在图4-16中可以看到，没有添加遮挡检测的视差图噪点较多，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的概率，得到了匹配效果较好的视差图。

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。

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：

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

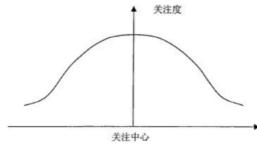


图 4-17 人类双眼关注度与关注中心的相关性示意图

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, s 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。

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, 来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, 乘以几何距离权重保存, 最后将窗口所有像素点统计相加, 得到最佳匹配点, 输出视差图。以Tsukuba为例, 经过几何距离自适应权重之后的匹配效果如图4-18所示, 误匹配率从原来的8.39%下降到了6.77%。

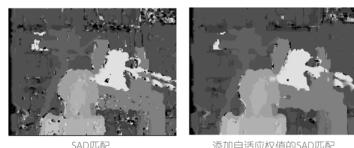


图 4-18 几何距离自适应权重优化效果对比

4.6 实验结论及全文最后总结

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, 实现了基础图像匹配效果, 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] 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
- [2] 胡汉平. 双目立体测距关键技术研究[D]. 中国科学院, 2014
- [3] 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
- [4] Bradski G, Kaehler A. 学习OpenCV[M]. 北京: 清华大学出版社, 2009
- [5] 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
- [6] Richard, Szeliski. 计算机视觉-算法与应用[M]. 北京: 清华大学出版社, 2012
- [7] 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
- [8] 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
- [9] 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

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

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

视觉匹配是一个寻找两幅输入图像之间相似性的问题，它是计算机视觉领域的一个关键难题，拥有广泛的实际应用场景，因此近几年对于视觉匹配的研究也越来越多。

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。在以上四个步骤中，立体匹配是最重要也是最困难的一步。一般地，立体匹配算法分为两大类：局部立体匹配算法和全局立体匹配算法。通过对两种算法的优缺点，本文将重点研究和讨论局部立体匹配算法。

本文在经典局部匹配算法 SAD 的基础上，通过分析算法的不足，进行了算法优化，提出了基于遮挡检测的改进的SAD算法。通过对输入图像进行自左向右和自右向左的双向匹配得到的左图和右图对应的两幅视差图进行遮挡检测，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效果。