重庆大学硕士学位论文

立体图像对的生成算法研究

硕士研究生：支丽欧

指 导 教 师 ：朱庆生　教　授

学科、专业：计算机软件与理论

重庆大学计算机学院

二 OO 七年十月

Master Degree Dissertation of Chongqing University

Study on the Algorithm of Stereo Image Pair Generation

Master Degree Candidate: Zhi Liou

Supervisor: Prof. Zhu Qingsheng

Major: Computer Software and Theory

College of Computer Science

Chongqing University

October 2007

摘　　要

双目立体成像是计算机智能视觉的重要分支，是指对于同一场景中的两幅立体图像对，当观察者经过匹配和理解后，能感知到具有立体感的景象。该技术在虚拟现实、多媒体教学、数字娱乐、产品外观设计、雕刻与建筑等领域有着广泛的应用。

论文首先对双目立体成像相关理论及技术进行研究，给出了一个基于监视器的双目立体成像模型，详细讨论了这个模型所具有的成像性质，以及双目视差、会聚角和 Panum 融合区限制等对立体感知的影响；然后针对平面图像和视频序列帧各自的特性，分别研究其不同的立体转化算法。

利用心理深度暗示和生理深度暗示相结合的原理，对单幅平面图像进行立体转换。构造了均匀分布、分段均匀分布、正态分布、三角分布和拟合灰度值分布等五种分布，通过实验详细讨论当图像子块在水平方向上的随机位移分别服从各种分布时的成像情况，并采用交叉熵和均方根误差两个定量指标对转换效果进行评价；此外还研究了图像子块个数对立体效果的影响。实验结果表明，当随机变量服从正态分布时立体效果较好，且子块个数对结果无明显影响。

由基于时空插值的立体转化理论可知，当追踪一个运动物体时，在每一时刻，双眼会看到视场中不同的部分，但却有可能在不同的时刻看到相同的场景，因此针对单目视频，提出了一种通过在帧间选择立体图像对的方式将其转换成双目立体版：首先采用特征点跟踪算法，获得特征点在相邻帧中的相对位移信息，然后估计帧间视差，并据此为每帧在序列中寻找合适的右眼视图。实验显示，采用帧间视差在-5 到-9 个像素之间的两帧构成立体图像对时，能得到较为稳定的位于头部和监视器屏幕之间的立体虚像；且当摄像机或者场景中的对象具有近似水平方向的运动时，该方法能获得较好的性能。

关键词：立体成像，双目视差，立体转换，深度感知

ABSTRACT

Binocular stereo imaging is a main branch of computer intelligent vision, which means the observer can perceive three dimension scenes through matching and understanding two captured stereo images pair of the same scene. This technique has an abroad application in the field of virtual reality, multimedia education, digital entertainment, appearance design of industrial products, sculpture and architecture etc.

In this dissertation, the theories and techniques related to binocular stereo imaging are firstly studied, and a binocular stereo imaging model based on monitor is proposed. Then the properties of this model as well as the impact of binocular disparity, assemble angle and Panum's fusional area limit to the stereo perception are thoroughly discussed. Then according to the different character of a single planar image and video frames， we work over their stereo conversion algorithm respectively.

According to the principle of psychological and physiological depth cues, we convert the single planar image into stereo. Five distributions such as uniform distribution, piecewise uniform distribution, normal distribution, triangular distribution and fitting gray value distribution are constructed firstly, then we discuss the stereo effect when the horizontal movement of sub-images subject to each of them and two quantitative criterions, cross-entropy and root-mean-square error, are used to evaluate the stereo effect; furthermore the impact of number of sub-blocks on stereo effect has been also discussed. The experimental results show that: the stereo effect is preferable when random variables subject to normal distribution and the number of sub-blocks has little influence to the outcome.

Known from the stereo conversion theory based on spatio-temporal interpolation, when tracking a moving object each eye sees different portion of the visual field at each instant, but may see the same visual field at different time, therefore aiming at monocular video, we present a method to generate its stereo version by selecting stereo pair among the frames. Primarily the displacement information of the feature points in adjacent frames is acquired via feature track algorithm; then the disparity between frames is estimated, and according this the suitable right-eye view is selected among the sequences. Experimental results shows that, two frames with the binocular disparity between -5 and -9 pixels observers can get suitable stereoscopic perception, and good

performance is acquired when the camera or the object in the scene has an approximately horizontal movement.

Keywords: Stereo Imaging, Binocular Disparity, Stereo Conversion, Depth Perception.

目　　录

1 绪 论

1.1 计算机立体视觉概述

视觉信息是人的主要感觉来源，俗话说“百闻不如一见”，人类认识外在世界的信息 80%来自视觉，人类通过眼睛与大脑来获取、处理和理解视觉信息。周围环境中的物体在可见光的照射下，在人眼的视网膜上形成图像，由感光细胞转换成神经脉冲信号，再经神经纤维传入大脑皮层进行处理与理解。视觉不仅指对光信号的感受，它包括了对视觉信息的获取、传输、处理、存储与理解的全过程。信号处理科学与计算机技术出现以后，人们试图用摄像机获取景物图像并转换成数字信号，用计算机实现对视觉信息处理的全过程，从而逐渐形成了一门新兴的学科，即计算机视觉[1, 2]。

因此计算机视觉就是用各种成像系统代替视觉器官作为输入敏感手段，由计算机来代替大脑完成处理和解释。目前，计算机视觉已在遥感图像分析、文字识别、医学图像处理、多媒体技术、图像数据库、工业在线检测与军事上的目标自动识别跟踪等方面和领域取得了广泛应用。

计算机立体视觉是计算机视觉的一个重要分支，它直接模拟了人类视觉处理景物的方式。由于客观世界在空间上是三维的，所以计算机视觉的研究和应用从根本上来说应该是三维的。

计算机立体视觉主要研究如何借助(多图像)成像技术从(多幅)图像里获取场景中物体的距离(深度)信息。立体视觉的基本方法是从两个或多个视点去观察同一场景，获得在不同视角下的一组图像，然后通过图像处理技术和三角测量原理获得不同图像中对应像素间的视差，并进而推断场景中目标的空间位置等。立体视觉的工作过程与人类视觉系统的感知过程有许多类似之处，事实上，人类视觉系统就是一个天然的立体视觉系统。计算机视觉可以在模拟人眼视觉的基础上，扩展、丰富人眼的视觉范围。计算机视觉的研究对提高机器的自动化和智能水平、对智能机器人和智能系统的发展都有很大的促进作用[3]。

一个完整的传统立体视觉系统通常可分为图像获取、摄像机定标、特征提取、立体匹配、深度确定及内插等 6 个大部分[4]。立体图像获取的方式很多，主要取决于应用的场合和目的。摄像机标定是为了确定摄像机的位置、属性参数和建立成像模型，以便确定空间标系中物体点同它在图像平面上像点之间的对应关系。特征提取是为了得到匹配赖以进行的图像特征，由于目前尚没有一种普遍适用的理论可运用于图像特征的提取，从而导致了立体视觉研究中匹配特征的多样性。目前常用的匹配特征主要有点状特征、线状特征和区域特征等。

立体匹配是立体视觉中最重要也是最困难的问题。当空间二维场景被投影为一维图像时，同一景物在不同视角下的图像会有很大不同。场景中的诸多因素：如光照条件、景物几何形状和物理特性、噪声干扰和畸变以及摄像机特性等，都被综合成单一的图像中的灰度值。因此要准确地对包含了如此之多不利因素的图像进行无歧义的匹配是十分困难的。对于任何一种立体匹配方法，其有效性有赖于 3 个问题的解决即：选择正确的匹配特征，寻找特征间的本质属性及建立能正确匹配所选特征的稳定算法。

已知立体成像模型和匹配视差后，三维距离的恢复是很容易的。影响距离测量精度的因素主要有摄像机标定误差、数字量化效应、特征检测与匹配定位精度等。一般来讲，距离的测量精度与匹配定位精度成正比，与摄像机基线长度成反比。增大基线长度可改善距离测量精度，但同时会增大图像间的差异，增加匹配的困难[4]。

立体视觉的最终目的是为了恢复景物可视表面的完整信息，特征匹配方法只能得到离散特征点的视差，需要进行较多的内插处理。对于一个完整的立体视觉系统来讲，不能断然地将匹配与内插重建过程分为两个不相关的独立模块，它们之间应该存在着很多的信息反馈，匹配结果约束内插重建，重建结果引导正确匹配。

虽然立体视觉经过 20 多年的研究，已经有了很大的发展。但无论是从视觉生理的角度，还是从实际应用方面来看，现有的立体视觉技术还处在十分不成熟的阶段。这不仅仅涉及到技术上的原因，更多地在于人类对自身视觉机理还不十分了解。人类视觉系统具有惊人的分析理解能力，但人类是如何精选、获取和分析理解视觉知识的，至今还未充分搞清楚。立体视觉作为一门多学科的交叉科学，正吸引着大批包括视觉生理、心理、物理、数学以及计算视觉等多种学科的研究人员，运用不同的技术手段对其进行深入的研究，它不但具有重要的实用价值，而且对促进人类视觉机理的研究，揭开人类视神经系统的奥秘具有非常重要的意义。

1.2 双目立体成像技术发展现状

计算机立体视觉技术有多种，目前研究较多的是**双目立体视觉**(Binocular Stereo Vision)技术。人们通常总是用双眼同时观看物体。由于两只眼睛**视轴**间距(**瞳距**)的存在，左眼和右眼在观看一定距离的物体时，所接收到的视觉图像是不同的。大脑通过眼球的运动、调整，综合了这两幅图像的信息，感知到生理深度暗示，从而产生立体感，这便是双目立体视觉的基本原理[5, 6]。利用双目立体视觉技术人们可以实现**双目立体成像**(Binocular Stereo Imaging)。简单地说，双目立体成像就

是根据双目立体视觉的基本原理，获取、产生和传输一个空间的**场景**(Scene)，并将这个场景展现出具有**立体感**(Stereoscopic Perception)的景象[7]。

利用计算机立体视觉技术进行**立体成像**(Stereo Imaging)最早是由 Waters 等人提出来的。随着计算机技术的不断发展，双目立体成像逐渐成为最具应用前景的立体成像技术之一，也是人们研究最多的热点之一[8]。当前双目立体成像技术主要分为以下两大类：

第一类方式是使用立体相机拍摄场景得到立体图像，这类方式获得的立体图像效果好。然而需要至少两台经过光学特性、机械特性和电子特性严格校准的摄像机，并且这两台摄像机的聚焦系统、变焦系统、几何失真、增益控制、光圈控制、会聚控制和视差控制等都要求非常精确的一致，这不仅增加了拍摄立体图像的难度而且后期处理的技术要求高[9]；因此有研究提出了单摄像机模型，2005 年，郑继贵等利用光学成像，把单摄像机镜像为一对虚拟摄像机，在 CCD 像面上采集到同一物体存在视差的两幅图像[10]。

第二类方式是通过计算机绘制得到立体图像对。自从胶片发明以来，历史上积累了大量优秀的二维图像和视频，如果能通过图像处理技术将其转化为立体的，具有广泛的应用前景。国内外不少学者对这方面进行了研究，典型的基于图像绘制的方法主要有：

① 利用图像序列对场景进行三维建模：重建景物的三维形状，恢复物体的空间位置信息[11]。主要分为匹配，摄像机标定和三维重建三部分。计算机视觉的问题本质上都是逆问题 (inverse problem)，输入图像的灰度受物体的几何特性，材料表面性质，颜色，环境光照及摄像机参数等许多因素的影响。由灰度反推以上各种参数是一个逆过程，往往都是非线性的，问题的解不具有唯一性，而且对噪声或离散量化引起的误差极其敏感，所以计算机视觉本身存在一定的病态性。如何得到问题的鲁棒解成为三维重建过程的难点所在。虽然许多三维重建的实施方案和数学模型在理论上是比较完善的，但往往受到现场条件的许多限制，严重地影响了其在工程中的应用。例如：有些需要测量出一些与拍摄有关的现场参数或需要放置复杂标定装置；有些对相机的拍摄过程有相当严格的要求，如相机需要有一定的俯角，有相当的高度或有较大的物距，甚至需要有严格位置关系的双相机拍摄。另外，有些方案虽然可以得到较为满意的重建结果，但其复杂的处理过程也会影响到其在许多场合的应用。这些复杂过程有些需要较多的已知参数，有些则是要求一次解出较多的未知参数，大量的运算影响着实时性的要求。

② 单视图形变法，1969年，L. Weiner首先提出了倾斜投影屏立体成像法，后来Shanks用一个向后倾斜的马鞍型曲面代替了Weiner的倾斜投影屏。1997年，Cassandra **T.** Swain利用单目线索如阴影，遮挡，亮度等加强立体效果[12]。天津大

学的侯春萍等人提出了一种基于心理深度暗示和生理深度暗示原理的平面图像立体化方法。该方法将一幅图像划分成$M \times N$个图像子块，然后使每一个子块在水平方向上随机的偏离它们原来所处的位置，这样得到的图像和原图像一起构成立体图像对[9]，本文第四章将会对该方法进行深入研究和分析。2005年Yamada提出了利用冷暖色调等加强立体效果的方法[13]。但根据仅有的一幅图像生成另一幅图像，可利用的信息较少，因此难度较大。

③ 基于深度信息估计的图像新视角生成，Matsumoto 等利用运动视差原理，通过相邻两帧确定摄像机的运动参数，由此得到对象的深度信息，产生立体图像对[14]。但是对相邻帧所有像素点进行匹配，时间复杂度很高，且在无明显纹理的地方匹配准确度较差。在文献[15]中，Harman 等采用机器学习算法产生关键帧的深度图，该方法需要手动输入大量已知点，作为“机器学习”的训练数据。

除了基于图像绘制的方法，我们还可以从一段平滑的视频序列中提取立体图像对，本文第五章着重论述了我们在这方面的研究工作。

当前，立体图像对的获取正成为双目立体成像的一个研究热点。其中第二类方式，即通过计算机绘制得到立体图像对的方式在虚拟植物可视化展现、计算机游戏、电视广告、电影特技等领域有着广阔的应用前景。

1.3 选题背景与研究意义

计算机视觉是一门非常年轻的学科，从 Marr 的视觉计算理论框架形成以来，虽经过了二十年的飞速发展，但还远不能满足工程应用和日常生活等多方面的需要。计算机视觉的应用领域在迅速拓宽，尤其是在普通便携相机成像条件下，如交通导航、现场勘测、自动化生产、虚拟现实等很多领域都在竞相应用这项技术，同时对已有的应用也提出了更高的要求，如提高视觉精度、扩大测量范围、简化操作过程和视觉识别智能化等[16]。其他学科如人工智能、模式识别、神经网络和信号处理的出现和发展，为计算机视觉的研究提供了许多新的方法和工具。新的成像方法和成像仪器的出现也推动着计算机视觉的发展，如图像扫描仪、数码相机、数码摄像机和多种距离传感器等，这些新型的仪器会带来或要求新的图像处理方法。面对相关理论和相关技术的发展，计算机视觉如何在此基础上发展自己的学科，以适应实际应用的需求，这是计算机视觉研究工作者的主要任务，也是本文选取课题的主要依据。

从整体上来看，国内对计算机立体视觉的理论研究和实际运用相对落后。双目立体视觉技术是人们研究较多和最具实用价值的计算机立体视觉技术，它可用于实现双目立体成像。双目立体成像涵盖生理深度暗示的产生、立体效果的评价、

立体图像对的获取等许多方面；涉及的具体技术和方法包括了计算机图形学、摄像机的几何模型的建立、摄像机的几何标定、模式识别、基于图像的绘制，等等。

由上可知，研究、分析、探讨双目立体成像技术，对于提高国内在计算机立体视觉领域的学术水平，促进双目立体成像在虚拟现实、机器视觉、多媒体教学、数字娱乐、产品外观设计、雕刻与建筑等领域的应用，都有着重要的学术和实用意义。

立体图像对的获得是双目立体成像的关键技术，是虚拟植物可视化展现的重要组成部分，对于虚拟植物可视化展现具有重要意义。2006 年重庆大学计算机学院成功申报国家高技术研究发展计划(863 计划)：虚拟作物可视化引擎关键技术研究(项目编号：2006AA10Z233)；2007 年申报成功重庆大学研究生科技创新基金：基于计算机立体视觉的双目立体成像研究(项目编号：200701Y1A0080194)。本学位论文的研究工作是这些项目的重要组成部分。本论文的研究目的，就是希望在对基于计算机立体视觉的双目立体成像技术进行全面、系统研究的基础上，重点探讨获取立体图像对的一些关键问题，在此基础上给出我们课题组的一些研究结论，最终将双目立体成像技术应用于虚拟植物可视化立体展现中。

1.4 本文的主要工作

双目立体成像技术涉及计算机视觉、模式识别、人工智能、认知心理学、计算机图形学等领域中许多具有挑战性的研究论题。已有一些文献对相关问题开展了研究。主要存在的问题是所成的立体图像仍不够逼真和自然，人们对人眼的功能以及双目立体成像模型了解还不够彻底，立体图像对的获取还有一些难题没有得到较好的解决。本文在对双目立体成像相关技术及理论进行研究的基础上，针对基于监视器的双目立体成像模型，深入研究了两种立体图像对获取的方法。根据上述研究内容，本论文各章节安排如下：

第二章介绍了双目立体成像相关技术及理论。首先简要介绍了双目立体成像系统，包括双目立体成像系统的发展现状、分时双目立体成像系统以及立体图像文件类型，等等。然后论述了双目立体成像模型相关理论，包括生理深度暗示和心理深度暗示、会聚角变化值和 Panum 融合区限制。

第三章介绍了相关图像图像处理技术，包括常用图像平滑技术，图像分割原理以及图像特征点提取、跟踪算法和特征相似性度量方法。

第四章研究了对单幅图像进行立体转换的方法，给出了评价转换后立体效果的定量指标——交叉熵和均方根误差，研究了图像子块个数对立体效果的影响，构造了均匀分布、分段均匀分布、正态分布、三角分布和拟合灰度值分布等五种分布，通过实验详细讨论了各种分布对立体效果的影响。

第五章研究了基于单目视频序列的立体图像对的生成方法。在研究双目视差对深度感知影响的基础上，根据特征点提取跟踪的方式获得帧间视差，然后通过在帧间选择立体图像对的方式将单目视频转换成双目立体视频，并通过实验对该方法进行了讨论。

第六章为总结和展望，对全文的研究工作进行了总结，归纳了论文中的创新点，并对下一步的研究工作进行了展望。

2　双目立体成像相关理论及技术

2.1 双目立体成像原理

2.1.1 深度暗示

立体图像是在人的大脑视区中形成的，它与人类的视觉系统是密不可分。人类的立体视觉系统利用多种深度暗示理解三维场景中物体的大小和相对位置，总体上，这些深度暗示可以分为两类：**生理深度暗示**(Physiological Depth Cue)和**心理深度暗示**(Psychological Depth Cue)[17,18]。

生理深度暗示包括人眼**晶状体调节**(Accommodation)、**双眼会聚**(Convergence)和**双目视差**(Binocular disparity)**和运动视差**(Motion parallax)等因素。

① 在人眼观看不同的物体时，眼球晶状体的焦距会发生变化。同时，物体的远近差异还会改变眼球瞳孔的直径。这种现象称为人眼的自适应效应，即晶状体调节。

② 人的左右眼观看同一物体时，两眼视轴相交会形成一个角度，该角度被称为会聚角。要形成这一会聚角，人的双眼便需回转一定的角度，此时人眼纤毛体肌肉便需作一定的功。人的感觉器官能比较出纤毛体肌肉用力的强度，便会感知出物体实际存在的远近。左右眼在观看远近不同的两点时，产生出的会聚角会不一样，眼球转动的程度也不一样。

③ 在双目视觉中，左右眼视网膜上所成的像亦略有不同，这是因为两眼相距约 65mm，当人观察一个立体物体时是从不同角度来观察的，具体的说，左眼看到物体的左边多点，右眼看到物体的右边多点，这就是双目视差。两只眼睛把各自所接收到的视觉信息传递到大脑皮层的视觉中枢，在这里经过一定的整合，产生一个单一的具有深度感的视觉映象。人处在正常身体姿态时，两眼视差是沿水平方向的，称为水平视差；沿视网膜上下方向上的视差为纵向视差，它在生活中很少见。

④ 当场景或者观察者运动的时候，由于视线方向的连续变化，视网膜成像也不断发生变化，这就是运动视差。

其中，双目视差是人眼最强烈的生理深度暗示因素。Bela Julesz 利用随机点图已经证明，在排除一切心理深度暗示之后，一组完全无意义的视觉刺激，只要具备视差条件，就能经双眼产生深度上的感觉[19]。这说明双目视差可以与任何视觉经验无关。由于双目视差对立体视觉的贡献最大，是最强烈的生理深度暗示，所以也是双目立体成像考虑的主要因素。

心理深度暗示主要由视觉经验和视觉记忆构成。人们在观看一张平面彩色图

片时，可以根据图片上的内容判断其中物体、人物之间的距离关系，而这种判断通常十分准确。这说明平面图像中尽管不存在能用人的双目视差等生理深度暗示因素来识别的深度信息，却存在着其它的深度暗示。这些暗示信息是人类对自然景物长期观察而得到的一种立体的视觉记忆和视觉经验，依靠这种视觉记忆和经验，观察者能够从平面图像中准确地提取出物体间的相对位置和深度信息，这种深度暗示被称为心理深度暗示。主要包括以下几种：

① **线性透视**(Linear perspective)，指对象在视网膜上成像大小与物体的距离成反比， 当景物中含有如走廊和轨道等平行线分量时，看起来这些平行线将会聚于视线远方的某一点。

② **光及阴影**(Shading and shadowing)，利用物体上光亮部分和阴影部分的适当分布可增强立体感。

③ **空气透视**(Aerial perspective)，当观看远处弥漫着烟雾的风景物时，会令人有一种强烈的深度感。这是由于空气中的微粒对光的散射及吸收使景物的对比度随着距离的增大而下降所产生的。

④ **重叠遮挡效应**(Interposition)，当两个图像遮挡，重叠、轮廓线相交时，我们认为被遮挡的物体较远。

⑤ **视网膜成像的相对大小**(Retinal image size)，同样大小的物体，当观看距离不同时，在视网膜上成像的大小也不一样。距离越远，视网膜成像便越小。由此通过比较视网膜成像的相对大小来判断出物体的前后关系。其余的心理暗示还有颜色，纹理梯度(Texture gradient)等。

人类视觉系统利用所有这些深度暗示决定场景的相对深度，被观察的景物是同一个景物，其深度关系是同一个深度关系，因此生理深度暗示和心理深度暗示所反映的被观察景物的深度关系应该是一致的，可相加的，暗示越多，观察者能更好地决定深度。这种一致性作为一种恒常的视觉经验被记忆，以至于任何违反这种恒常性的深度信息都会受到强烈的抑制。因此，人类的立体视觉是综合作用的结果，生理深度暗示和心理深度暗示在不同的情况下有相互增强或相互抑制减弱的作用[9]。

图 2.1 是为说明两者之间的关系所设计的一个试验的结果。图中的横坐标 $\boldsymbol{X}$ 表示屏幕的水平方向，纵坐标 $\boldsymbol{E}$ 表示显示在屏幕上的图像给观察者造成的深度感知程度。“●”表示观察者对屏幕上一个图像单元的心理深度感知，“■”表示观察者对屏幕上一个图像单元的生理深度感知，“△”表示屏幕上一个已经含有心理深度信息的图像单元，在加入了生理深度暗示之后给观察者造成的综合的深度感知[9]。

例如图 2.1 中的第一个图像单元给观察者造成的心理深度感知是 Ex_1，随机加入生理深度暗示给观察者造成的生理深度感知是 Es_1，而在这个单元中加入的生理

深度暗示与原有的心理深度暗示不矛盾，生理深度感知和心理深度感知共同形成的综合深度感知为 Ez_1，这个图像单元的深度感知得到加强。

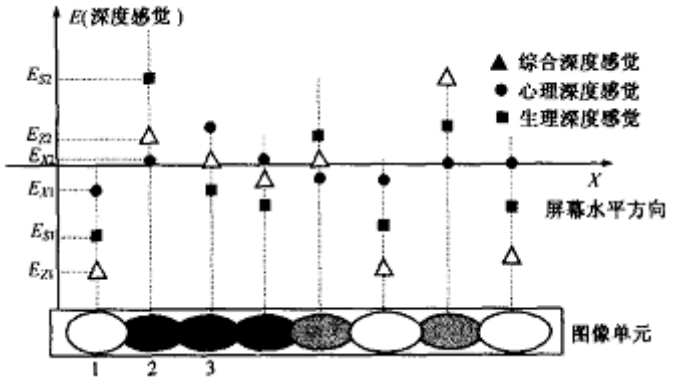


图 2.1 生理深度暗示和心理深度暗示的关系

Fig 2.1 Relationship between physiological depth cue and psychological depth cue

再看图 2.1 中的第二个图像单元，其心理深度感知是 Ex_2，随机加入生理深度暗示给观察者造成的生理深度感知是 Es_2，由于心理立体视觉告诉观察者图像单元 2 的深度不可能比单元 3 的更远，形成的综合深度感知 Ez_2，这时心理深度暗示对生理深度暗示产生了抑制减弱的作用。用同样的方法可以分析得出所有图像单元的综合深度感知，如图 2.1 所示。这样得到的综合深度感知符合心理深度暗示深度的变化规律。由于心理深度暗示反映的也是客观景物的深度变化，因此从这样构成的立体图像中，观察者可以获得符合客观景物深度变化规律的含有双目视差的深度感知。

由此得出结论：当生理深度暗示与心理深度暗示一致时，综合的深度感知增强；而当生理深度暗示与心理深度暗示矛盾时，生理深度暗示会被心理深度暗示抑制减弱[9]。产生这一立体视觉现象的原因是，心理深度暗示尽管不能提供最强的深度暗示，但对违反视觉经验的深度暗示将产生极强的抑制作用，这是单眼视觉模式识别的重要特征，也是人类视觉提供给自身的一种重要的自我保护功能。

2.1.2 双目立体成像模型

本文给出的基于监视器的双目立体成像模型如图2.2所示。图中采用的坐标系是**世界坐标系**。其中，$\boldsymbol{\alpha}$(监视器屏幕)是平行于***XOY***面的投影平面；$\boldsymbol{\beta}$是平行于***XOY***面的像平面，$\boldsymbol{W}$为$\boldsymbol{\beta}$中的一点。设$\boldsymbol{A}_1(h, 0, 0)$、$\boldsymbol{A}_2(-h, 0, 0)(h>0)$分别是右眼和左眼的坐标，因此原点处在**中央眼**的位置，***XOY***面即为观察者双眼所在平面。设像平面与***XOY***面的距离为$o_d(o_d>0)$，投影平面与***XOY***面的距离为$p_d(p_d>0)$。***MO'N***为监视器屏幕的**图像坐标系**，***O'M*//*OX***，***O'N*//*OY***，***O'***在坐标系***XYZ***中的坐标为(x_0, y_0, p_d)。

通过 $\boldsymbol{A}_1$ 和 $\boldsymbol{A}_2$ 观看 $\boldsymbol{W}$，我们可以分别得到 $\boldsymbol{W}$ 在 $\boldsymbol{\alpha}$ 内的**立体图像对** $\boldsymbol{I}_1$ 和 $\boldsymbol{I}_2$。当

A_1 和 A_2 处分别被放置成两台相同摄像机的像平面时，便是**立体摄像机几何模型**。

设 α 为监视器屏幕。在图 2.2 的模型中如果用计算机来生成立体图像对 $I_1^{'}$ 和 $I_2^{'}$，则应使 $I_1^{'}$ 和 $I_2^{'}$ 处在屏幕的同一水平位置，才能保证不会在图像中人为地产生**垂直视差**。

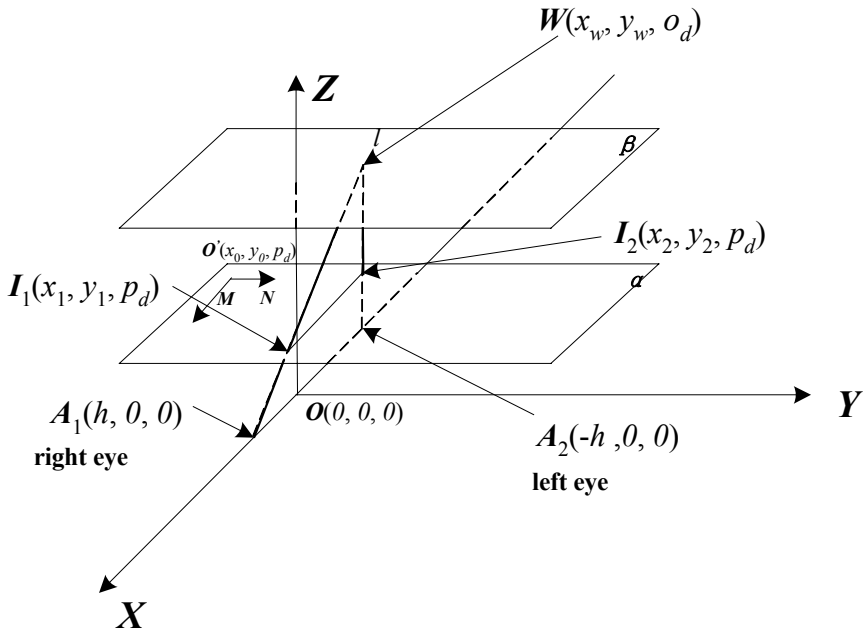


图 2.2 基于监视器的双目立体成像模型

Fig 2.2 Binocular stereo imaging model based on monitor

那么水平视差 D 与 o_d 和 p_d 具有如下关系：

$$D = x_1 - x_2 = x_1^{'} - x_2^{'} = 2h\frac{o_d - p_d}{o_d} \tag{2.1}$$

$2h$ 为瞳距，一般为 65mm。由式 2.1 表明，水平视差只与 p_d 和 o_d 相关，只要平面 α、β 一定，则 β 中任一点所对应的水平视差都相等。今后除非特别声明，视差是指水平视差。

由式 2.1 可知：

① $\lim\limits_{o_d \to +\infty} D = 2h$

这表明对于远处的物体 W，水平视差趋于定值。因此，双目立体视觉模型对近距离(5m 以内)的物体特别有效，对于远处的物体双眼难以分辨出它们之间的深度信息，因而立体感较弱[20]。

② 设 α、β 固定，则 $o_d - p_d$ 为定值 d_p。当观察者远离屏幕时，有

$$\lim_{p_d \to +\infty} D = \frac{2hd_p}{p_d + d_p} = 0 \tag{2.2}$$

即水平视差会减小，反之增加。这表明当观察者远离屏幕时，如果保持成像位置不变，那么立体感会减弱。

2.1.3 会聚角变化值

经验表明，人们习惯于调整焦点和转动双眼来观察不同深度的物体。在进行双目立体成像时，虽然人们感兴趣的像点的深度不同于屏幕，但是仍要求我们聚焦于屏幕。在这种情况下，Valyus发现大多数人能容忍的会聚角变化值为1.6°[20]。如果这个值过大，会引起“调节/汇聚”矛盾，如图2.3所示。

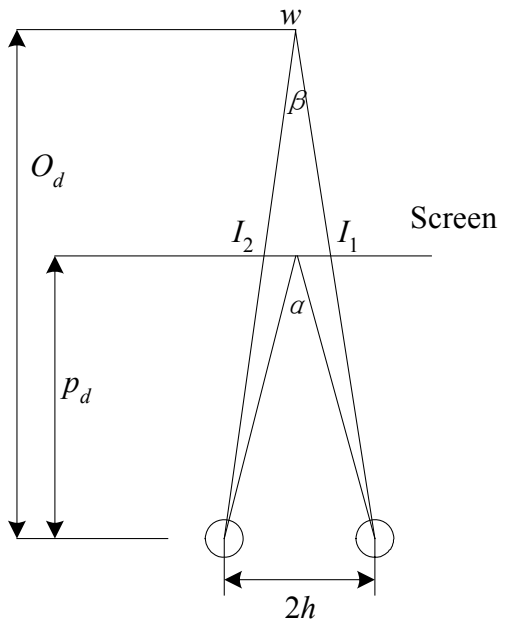


图 2.3 调节与汇聚

Fig 2.3 Accommodation and convergence

Valyus指出：

$$D_{\max} = 0.03 p_d \tag{2.3}$$

其中 $D_{\max}$ 为最大允许视差。Valyus 的理论与后面一节中提到的 **Panum 融合区** (Panum’s fusional area)的理论本质上是一致的，通常取两者中更为严格的限制条件。

公式 (2.3) 限制了立体图像的**景深**(在摄影机镜头或其他成像器前沿着能够取得清晰图像的成像器轴线所测定的物体距离范围)。从图 2.3 可得：

$$\alpha = 2arctg\frac{h}{p_d} \tag{2.4}$$

$$\beta = 2arctg\frac{h}{o_d} \tag{2.5}$$

因此会聚角变化值为：

$$\Delta\theta = \alpha - \beta = 2(arctg\frac{h}{p_d} - arctg\frac{h}{o_d}) \tag{2.6}$$

由于当 $\Delta\theta$ 很小时，有 $tg\theta \approx \theta$ (θ 用弧度表示)，则由公式(2.4)、(2.5)、(2.6)可得：

$$o_d = \frac{h}{tg\frac{\beta}{2}} \approx \frac{h}{\frac{\beta}{2}} = \frac{h}{\frac{\alpha}{2} - \frac{\Delta\theta}{2}} \approx \frac{h}{tg\frac{\alpha}{2} - \frac{\Delta\theta}{2}} = \frac{2h}{\frac{2h}{p_d} - \Delta\theta} \tag{2.7}$$

因为 $-\frac{1.6\pi}{180} \le \Delta\theta \le \frac{1.6\pi}{180}$，所以可以根据公式(2.7)计算出立体图像的景深。由此可见，生成的立体图像对的景深不是无限制的，我们不能将虚像成像在任意的位置。

2.1.4 Panum 融合区

另一个在立体成像中经常遇到的问题是 Panum 融合区问题，是 Panum 于 1858 年发现并提出的，主要解决双像单视问题。

当人眼的双眼注视一个外界物体时，另外一个远于或近于双眼单视界的物体，虽然没有刺激到左、右眼视网膜的对应点上，但只要刺激在两个视网膜对应点附近的一定范围内，也可以在大脑中产生单一的立体视觉。视网膜对应点附近这个很小的范围就叫做**Panum融合区**(Panum's Fsional Area)。如果几个物体刺激两个视网膜形成的像的视差过大，便会出现**复视**，即不能把该物体在左、右眼视网膜上形成的像在大脑中融合成单一的立体视觉。Panum融合区说明并不是所有在两个视网膜上形成含有双目视差的图像都能在大脑中形成单一的立体视觉，只有满足一定双目视差的立体图像对才能被融合成单一的立体图像。如果立体图像对的左、右眼视图的视差过大，那么在一定的距离下观看立体图像时，人的眼睛和大脑就无法协调晶状体调节、双眼会聚与双目视差等多种生理立体视觉因素所提供的深度信息之间的关系，将左、右眼视图融合成单一的立体图像，其结果会给观察者造成很大的不舒适感。图2.4说明了晶状体调节、双眼会聚与双目视差是如何相互影响的。

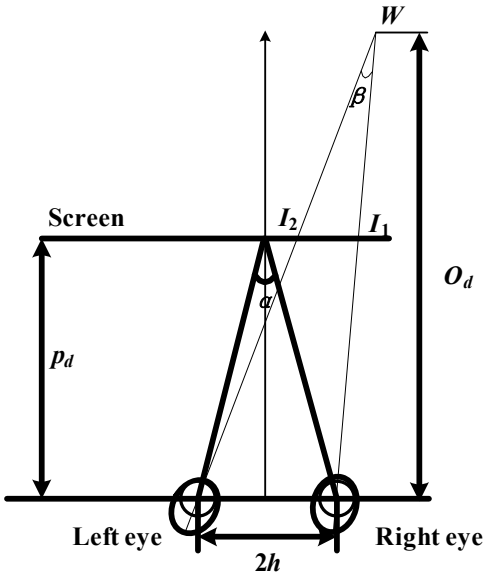


图 2.4 视差与 Panum 融合区

Fig 2.4 Parallax and panum's fusional area

观察者在观看立体图像时，设p_d为观察者与屏幕之间的距离，W是在观察者大脑中形成的双眼单视的像点，当p_d与W的深度之间的差超过一定限制时，晶状体调节所提供的深度暗示将起主导作用，而双目视差所提供的深度暗示无法被融合成单一的立体图像。超出Panum融合区限制的立体图像，会使观察者感到疲劳、头痛、甚至引起复视。实验表明，o_d不能超过的限制为[22]：

$$\frac{Ep_d}{E+\eta p_d} < o_d < \frac{Ep_d}{E-\eta p_d} \tag{2.8}$$

其中E是人眼的瞳孔直径，典型值为0.4cm，η为视锐度 (Visual Acuity)，即通常所指的视力，它表示人通过视觉器官辨认外界物体的敏锐程度，辨认物体细节的能力。一般情况下，我们将能分辨1′视角的视力定为正常视力标准，即$\eta = 1' \approx 2.907 \times 10^{-4}$ rad。为了获得舒适的立体图像，监示器屏幕前后的像点深度值都不应超过公式(2.8)的限制。落入公式(2.8)范围内像点集合所覆盖的区域称为**舒适Panum融合区。**一种成功的双目立体成像系统展示给观察者的立体图像，应该是不使观察者产生视疲劳的图像。因此水平视差的取值范围不能超出舒适Panum融合区的限制。

转换成像素表现形式，即要求立体图像对左右视图中的每一个像素对应的水平视差 D [9]满足不等式(2.9)：

$$-\frac{2h\eta p_d}{EP_\sigma} < D < \frac{2h\eta p_d}{EP_\sigma} \tag{2.9}$$

其中 $2h$ 是瞳距，一般为 6.5 cm；监视器屏幕与观察者的距离为 p_d(p_d>0)，这里取 100 cm，P_σ 是监视器的像素间隔，这里取 $P_\sigma = 3.53\times10^{-2}$ cm[3]；因此可得，$-13 < D < 13$。

2.2 双目立体成像系统

2.2.1 双目立体成像系统

如前所述，由于人类两只眼睛存在瞳距，我们在观看一定距离的物体时，左眼和右眼所接收到的视觉图像是不同的。大脑能够很巧妙地将两眼细微的差别融合，产生有空间感的立体景物。如果我们能够为左右眼分别提供同一场景的不同照片(立体图像对)，就能够利用双目立体视觉的基本原理看到立体图像。双目立体成像系统就是采用一定的策略让人的左眼只能看到左眼视图，右眼只能看到右眼视图。

双目立体成像系统主要由**显示芯片**、**驱动程序**和**观察显示设备**三部分组成。其中观察显示设备有多种，不同的观察显示设备采用的立体图像对显示方式也不同。目前常用的观察显示设备有**双色眼镜**、**液晶光阀眼镜**(LCD shutter glasses)、**头盔显示器**和**自由立体**(Autostereo)**显示器**等[22,23,24]。

① 双色眼镜

客观世界中人们观察到的各种颜色可由红、绿、蓝三种基本颜色(**加色基色**)按不同比例混合而成，计算机正是通过这种方法来产生不同的色彩，这就是所谓的**加色法**；同样，一种颜色的光通过另一种颜色的镜片的过滤，其中一些波长的光会被吸收，变成第三种颜色的光，这就是**减色法**。减色法的基本颜色是青色、洋红色及黄色。它们被称为**减色基色**，分别用于吸收过滤红色、绿色及蓝色的光。根据减色法的原理，可采用如下方法使左眼只能看到左眼视图、右眼只能看到右眼视图：将屏幕底色设计成黑色，左眼视图和右眼视图分别采用红、蓝两种颜色显示；观察者戴上双色眼镜，左眼用黄色镜片，右眼用青色镜片。这样，由于镜片的过滤作用，左眼只能看到红色的视图，右眼只能看到蓝色的视图。左右双眼看到的图像通过大脑的综合，就产生了深度感。

使用双色眼镜来观看立体图像是一种经济快速的方法。该方法成本低，不需要在计算机上安装专门的立体驱动，甚至可以用此方法观看纸质图片。但用双色眼镜观看黑白图片的效果较好，观看彩色图片的效果要差。目前这种观察设备有被淘汰的趋势。

② 液晶光阀眼镜

液晶光阀眼镜所采用的显示方式是**页交换**(Page Flipped)[22]。

在液晶上加一定的电压会改变其分子排列，从而可以控制液晶镜片的开关状态，允许或阻止光线通过液晶镜片。液晶光阀眼镜的页交换的显示方式是指：在某一时刻，关闭右眼液晶镜片，打开左眼液晶镜片，同时让计算机 CRT 显示器显示左眼视图；在下一时刻，关闭左眼液晶镜片，打开右眼液晶镜片，同时让计算机 CRT 显示器显示右眼视图。在此工作方式中改变液晶的开关状态，使左右眼镜片及左、右眼视图以较快的速度切换，大脑会认为左右眼同时看到了各自的图像，并综合左、右眼视图产生深度感知，这就达到了“左眼看左图像，右眼看右图像”的目的。

计算机的 CRT 显示器存在一定的刷新率，必须采用外部辅助电路连接 CRT 显示器与液晶镜片，使 CRT 显示器的刷新率与液晶镜片的开关速率同步，才能达到以上的目的。由于在刷新率不变的情况下左、右眼视图需要交替显示，因此每只眼睛看到的图像刷新率实际上是 CRT 显示器刷新率的一半。若 CRT 显示器刷新速率低的话，眼睛看到的图像刷新率更低，会出现画面闪烁感。可见，CRT 显示器的刷新率越高越好。

液晶光阀眼镜的价格适中，并可用于观看彩色图像。它虽然对 CRT 显示器刷新率要求较高，但立体效果(色彩、**沉浸**(Immersion)感等)较双色眼镜好[21]。因此，它在游戏娱乐、虚拟现实等领域里有广泛的应用。本文实验采用的分时双目立体成像系统就使用了液晶光阀眼镜。

③ 头盔显示器

头盔显示器(Head-Mounted Display，HMD)是 3D 显示技术中起源最早、发展得最完善、应用也较广泛的技术，是专为用户提供虚拟现实中景物彩色立体显示的显示器。它通常固定在用户的头部，用两个 LCD 或 CRT 显示器分别向两只眼睛显示计算机立体驱动生成的立体图像对。大脑将综合左、右眼视图像产生深度感知，头盔显示器的种类较多，根据需求的不同，有**全投入式**的和**半投入式**的，等等[23]。

头部位置跟踪设备是头盔显示器上的主要部件。通过跟踪头部位置，虚拟现实用户的运动感觉和视觉系统能够得以重新匹配跟踪，计算机随时可以知道用户头部的位置及运动方向。因此，计算机就可以随着用户头部的运动，相应的改变呈现在用户视野中的图像，从而提高了用户对虚拟系统知觉的可信度。头部位置跟踪还可以增加双眼视差和运动视差，这些视觉线索能改善用户的深度感知。

头盔显示器的分辨率较低，屏幕成像小，必须放大以达到和人的视野相一致，失真大；人眼在近距离内聚焦会感到疲劳，佩戴它观察无法得到舒适自然的体验；

而且头盔显示器的造价较高。但头盔显示器具有较好的临场感和沉浸感，在许多特定的场合具备特殊的优势，因此被广泛应用于军事、CAD/CAM、工业生产、模拟和训练、3D 显示与电子游戏、显微技术和医疗等领域。

④ 自由立体显示器

自由立体显示器是立体显示技术研究的一个重要方向。利用自由立体显示器，无需佩戴任何眼镜就能产生立体效果[24,28]。美国、日本、德国、加拿大等发达国家早在 20 世纪 60 年代末就开始进行这个领域的研究，并相继提出了一些实现原理和方法。TFT LCD 技术成熟后，掀起了自由立体显示器研究的又一轮热潮。

自由立体显示按按实现方法分主要有**透镜法**和**光栅法**两种[26]。两种方法都通过控制活动的 LCD 彩色阵列，显示一种合成的图像。该图像包含竖直交替排列的图像条纹，这些条纹构成了具有双目视差的左、右眼视图像。左、右眼视图以一定的速率切换，分别显示在奇数列和偶数列上，交替的并行显示使得每只眼睛只看到其对应列，亦即其对应的左眼视图或右眼视图，因而产生了深度感。

2.2.2 分时双目立体成像系统

分时双目立体成像系统是模拟人眼观察物体的方式，将左右两幅视图交替显示来产生立体图像的一种常用的系统[26]。本文实验采用的是基于 NVIDIA 显卡的分时双目立体成像系统。目前市面上显卡所采用的显示卡芯片中，支持立体显示功能且提供立体驱动程序的主要是 NVIDIA 公司生产的 TNT 系列、MX 系列、Geforce 系列的显示卡芯片，而其他公司生产的大部分显示卡芯片尚不能支持立体显示功能。因此，本论文中使用的显示卡芯片是 NVIDIA 公司生产的 Geforce 系列显示卡芯片，其中芯片类型为 GeForce4 MX 440 with AGP8X。CRT 监示器分辨率为 800×600 时刷新率最大能达到 100Hz。观察设备是液晶光阀眼镜，采用的显示方式为**页交换**(Page Flipped)显示方式。图像信号输出部分的硬件线路连接示意图如图 2.5 所示。

在图 2.6 所示的硬件线路连接示意图中，**同步信号线接头**的作用是连接显示卡输出接口和 CRT 显示器的信号线接头，并且引出一条显示信号的同步信号线，经过控制盒，最终与液晶光阀眼镜相连接，这样同步信号就能控制左右液晶光阀分时、交替开启和关闭，以保证左眼只能看到左画面，右眼只能看到右画面。

此外，CRT 显示器的信号线接头为 VGA 插头，控制盒和液晶光阀眼镜的插头为 3.5mm 的立体声耳机插头。

在图像输出过程中，很有可能出现左、右眼视图反转(即左眼看到了右画面，右眼看到了左画面)的情况，这样将无法使观看者产生正确的生理深度暗示。因此控制盒上装有一个左右眼交换开关。当左、右眼视图反转的时候，按一下这个开关就会使这两幅图像再次反转，从而恢复正常。

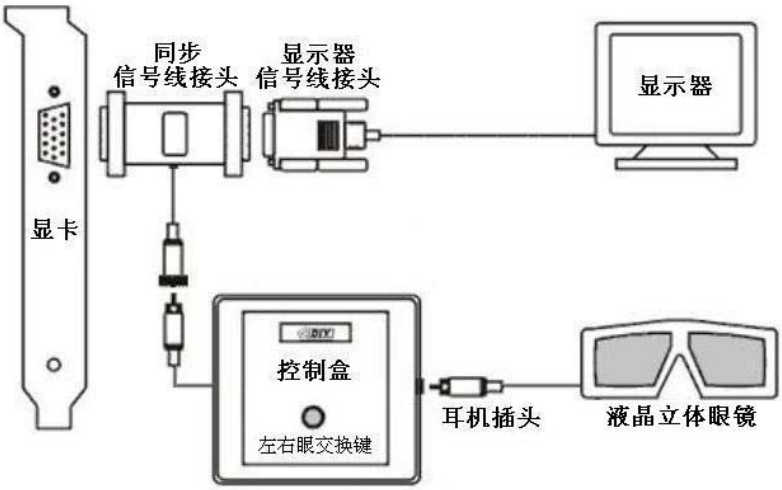


图 2.5 分时双目立体成像系统的硬件

Fig 2.5 Hardware of the time-sharing binocular stereo imaging system

该分时双目立体成像系统的工作原理是：计算机将左眼视图和右眼视图交替迅速地显示在监视器上。当左眼视图显示时，液晶光阀眼镜使右眼的液晶镜片遮断，使得只有左眼才能看到监视器；同理，当右眼视图显示时，液晶光阀眼镜使左眼的液晶镜片遮断，使得只有右眼才能看到监视器，这样便产生了立体感。NVIDIA 系列显卡能够将**JPS 文件**中的左右眼视图(立体图像对)交替显示在监视器上[26]。因此，只要生成含有立体图像对的 JPS 文件，就能通过这个系统看到立体图像。

常用的立体图像文件类型有 H3D(.h3d)文件类型和 JPS(.jps)文件类型等等[26]。本论文采用了 JPS 文件格式存储立体图像。JPS(JPEG Stereo)文件格式的立体图像是由同一空间场景中两幅同为 JPG 格式的左右眼视图拼凑成的，因此 JPS 格式与 JPG 格式没有本质上的区别，把后缀名为 jps 的图像文件改为 jpg 就可以作为 JPG 文件使用。

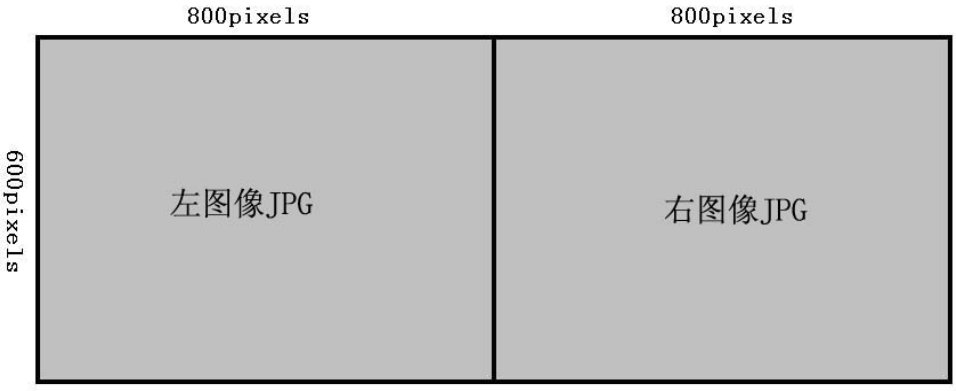


图 2.6 JPS 文件类型

Fig 2.6 JPS file type

标准的 JPS 文件是由左眼视图与右眼视图交错拼凑成的，图 2.6 是 JPS 文件中图像的一种排列情况。打开 JPS 文件进行观看时，立体驱动程序会自动将文件中的图像一分为二，作为左右眼视图分别进行加载。

图 2.7　立体图像

Fig 2.7 stereo image

图 2.7 是用立体相机的两个摄像头模拟人的双眼对物体从两个不同角度获取的立体图像。

2.3 本章小结

本章主要论述了双目立体成像相关技术及理论。

首先介绍了深度暗示对深度感知的影响，包括生理深度暗示和心理深度暗示；当生理深度暗示与心理深度暗示一致时，综合的深度感觉增强；而当生理深度暗示与心理深度暗示矛盾时，生理深度暗示会被心理深度暗示抑制减弱。然后简要介绍会聚角变化值和 Panum 融合区的限制。

最后介绍了现有的成像系统：当前流行的双目立体成像系统主要由**显示芯片**、**驱动程序**和**观察显示设备**三部分组成。其中观察显示设备有多种，不同的观察显示设备采用的立体图像对显示方式也不同。目前常用的观察显示设备有**双色眼镜**、**液晶光阀眼镜**(LCD shutter glasses)、**头盔显示器**、**自由立体**(Autostereo)**显示器**等。分时双目立体成像系统是模拟人眼观察物体的方式，将左右两幅视图交替显示来产生立体图像的一种常用的系统。

上述双目立体成像相关理论和技术，构成了本论文的研究基础。

3　相关图像处理技术

3.1 图像颜色空间

一般的图像(即模拟图像)是不能直接用数字计算机来处理的。为了使图像能够被数字计算机处理，首先必须将模拟图像转换为数字图像，即需要对图像进行数字化，把模拟图像分割成如图 3.1 所示的像素矩阵，每个像素的亮度、色度值都用一个整数来表示。

一幅数字化后的图像的数据量为：M(每列象素数)× N(每行像素数) ×b(灰度量化所占用位数)bit。抽样点数越多，量化级数越多，图像数字化的质量越高，但是该图像的数据量也就越大。为了使一幅图像既能得到满意的视觉效果，又能使其数据量最小，一般需要针对图像的具体内容来确定相应的M, N和b的值。一幅M×N个像素的数字图像，其像素灰度值可用 M 行 N 列的矩阵 F 来表示，如图 3.1 所示。这样，对数字图像的各种处理就可以变成对矩阵 F 的各种运算。

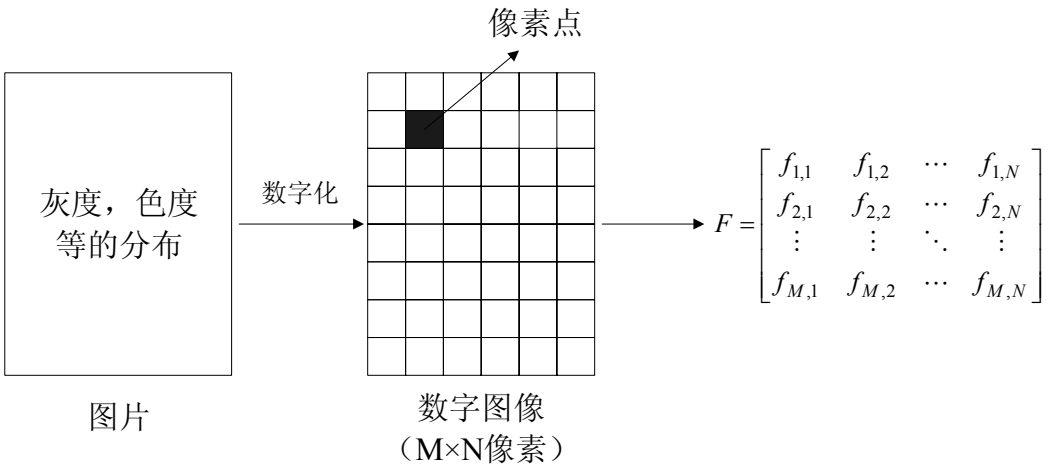


图 3.1 数字图像的表示

Fig 3.1 Representation of digital image

颜色是人脑对于外界光刺激的一种反应，在物理学角度来说是人眼对于不同波长的光线的一种映像，而在其他领域需要具体对颜色进行描述或者使用的时候，颜色通常用三个相对独立的属性来描述。三个独立变量综合作用所构成的三维立体结构就是一个颜色空间，即有如式 3.1 的表达：

$$G=(x)X+(y)Y+(z)Z \tag{3.1}$$

X, Y, Z 是三基色颜色量，x, y, z 是比例系数并且需要满足以下条件：

1．x>0，y>0，z>0；

2．y 的数值等于彩色光的亮度；

3．当 x=y=z 时表示标准白光；

式 3.1 说明颜色可以从不同的角度，用三个一组的不同属性加以描述，就会产生不同的颜色空间，但被描述的颜色对象本身是客观的，不同颜色空间只是从不同的角度去衡量同一个对象。

在不同的颜色空间里对同一幅图像进行处理会产生不同的结果，一些颜色空间例如：RGB，HIS，CIE，虽然都可以被用于彩色图像分割，但是没有任何一个可以被应用于所有的图像目标彩色空间，因此选择一个合适的颜色空间是图像处理技术的基本要求和首要步骤。人们为了适应不同的应用场合的需要已经构架了各种各样的颜色空间[29]。在这里将这几个颜色空间按照常见的方式分为三类：

① RGB 颜色空间

这类颜色空间是比较常见的，特点是与设备相关，即基于 RGB 颜色系统的不同的扫描仪对同一幅图像会得到不同的结果。这类颜色空间主要包 RGB，YUV 以及 $I_1I_2I_3$ 等。红色(R)、绿色(G)和蓝色(B)被称作三种基本的颜色，人类感知到的颜色就是这三种基本颜色联合产生的。利用这三种基本颜色，通过它们的线性或非线性表示能得到其它彩色空间的表示。在实际应用中，RGB 颜色模型用于磷粉屏幕的颜色生成，是一个由黑到白的过程，称为增色处理[30]；CMY 颜色模型主要用于描述绘图和打印彩色输出的颜色，是一个由白到黑的过程，成为减色过程。

② CIE 颜色空间

这类颜色空间是由国际照明委员会 CIE (Commission Internationale del Elarirage-the International Commission on Illumination)定义的颜色空间，通常作为国际性的颜色空间标准，用作颜色的基本度量方法。颜色空间包括 CIEXYZ, Lab, Luv 等颜色空间等，都是由 RGB 颜色空间经过非线性变换得到的。

CIEXYZ 颜色空间是一种非均匀颜色空间，主要用于彩色电视系统中，它与 RGB 颜色空间的转换公式如下：

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = A \begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 0.607 & 0.174 & 0.200 \\ 0.299 & 0.587 & 0.144 \\ 0.000 & 0.066 & 1.116 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{3.2}$$

转换矩阵 A 决定于显示设备所采用的三种荧光粉的色度坐标和标准光源，当采用 NTSC 制下的转换关系式时，NTSC 制采用的标准光源是白光 C 白，因此，转换矩阵为如 3.2 所示。

CIE1976 L*u*v*颜色空间实在 CIEXYZ 的基础上得到：

$$\begin{cases} L^* = 116 f(\dfrac{Y}{Y_n}) - 16 \\ u^* = 13L^*(u' - u_n) \\ v^* = 13L^*(v' - v_n) \end{cases} \tag{3.3}$$

其中：$f(x) = \begin{cases} x^{\frac{1}{3}} & x > 0.008856 \\ 7.787x + \dfrac{16}{116} & x \le 0.008856 \end{cases}$，$u' = \dfrac{4X}{X + 15Y + 5Z}$，$v' = \dfrac{9Y}{X + 15Y + 3Z}$，

$u_n = \dfrac{4X_n}{X_n + 15Y_n + 5Z_n}$，$v_n = \dfrac{9Y_n}{X_n + 15Y_n + 3Z_n}$，$X_n$、$Y_n$、$Z_n$是标准白光的三刺激值。

③ 孟塞尔颜色空间

孟塞尔颜色系统（Munsell colour system），用立体模型表示出物体表面的亮度、色调和饱和度作为颜色的分类和标定的体系方法[31]。这是一个根据颜色的视觉特点所制定的颜色分类和标定系统。Munsell 颜色系统的颜色卡片在视觉上的差异是均匀的。其色调、明度值和彩度反映了物体颜色的心理规律，它们可以分别代表颜色的色调、明度和彩度的色知觉特性。

总的来说，对于每一种颜色空间都有其应用领域，不能单纯的认为某一颜色空间的好坏。

3.2 图像预处理

一般情况下，任何一幅未经处理的原始图像，都存在着一定程度的噪声干扰。噪声恶化了图像质量，使图像模糊，甚至淹没特征，给分析带来困难。因此图像预处理的目的就在于：(1)改善图像的视觉效果，提高图像的清晰度；(2)将图像转换成一种更适合于人或机器分析处理的形式，并不旨在还原图像的真实性，而是通过处理突出那些感兴趣的信息。计算机视觉预处理就是检测前的预处理过程，它包括噪声的滤除、边缘的增强、对比度的改善和边缘和感兴趣区域的有效提取等。

消除图像噪声的工作称之为图像平滑或滤波，平滑的目的在于消除混杂在图像中的干扰，改善图像质量，强化图像表现特征。由于噪声源众多（如光栅扫描、底片颗粒、机械元件、通信传输等），噪声种类复杂（如加性噪声、乘性噪声、量化噪声等），所以平滑方法也多种多样。这里只简要介绍三种常用的方法：邻域平均法、高斯滤波法和中值滤波法[30]。

① 邻域平均法

令被讨论像素的灰度值为 $f(i, j)$，以其为中心，窗口像素组成的点集用 A 表示，共有 N 个像素。经邻域平均法滤波后，像素 $f(i, j)$对应的输出为：

$$g(i,j)=\frac{1}{N}\sum_{(x,y)\in A}f(x,y) \tag{3.4}$$

即用窗口像素的平均值取代 $f(i,j)$ 原来的灰度值。

邻域平均法有力地抑制了噪声，同时也出现了因平均作用后而引起的模糊现象，模糊程度与邻域半径成正比。

②高斯滤波法

为了克服简单局部平均的弊病，出现了加权平均法：根据参与平均像素的特点赋予不同权值，高斯滤波就是其中一种最常见的形式。窗口中像素 $f(x,y)$的权重由该点到中心点的距离决定：

$$w(x,y)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{d^2}{2\sigma^2}} \tag{3.5}$$

其中：$d=\sqrt{(x-i)^2+(y-j)^2}$ 。因此像素 $f(i,j)$对应的滤波后为：

$$g(i,j)=\sum_{(x,y)\in A}w(x,y)\cdot f(x,y) \tag{3.6}$$

高斯函数具有可分离性，二维高斯可通过将图像先与一维高斯函数进行卷积，然后将结果与垂直的相同函数卷积得到。当 $\sigma^2=2$ 时，一个 7×7 离散高斯模板如下所示：

1	4	7	10	7	4	1
4	12	26	33	26	12	4
7	26	55	71	55	26	7
10	33	71	91	71	33	10
7	26	55	71	55	26	7
4	12	26	33	26	12	4
1	4	7	10	7	4	1

图 3.2 7×7 离散高斯模板

Fig 3.2 7×7 discrete gauss templet

③ 中值滤波法

中值滤波是一种非线性滤波，在一定条件下，对于脉冲干扰和颗粒噪声有良好抑制作用，而且对图像边缘能较好地保持。

中值滤波是采用一个含有奇数个点的滑动窗口，用窗口中各点灰度值的中值 $g(i,j)$来代替窗口中心点像素 $f(i,j)$的灰度值。

$$g(i,j)=\underset{(x,y\in A)}{Med}(f(x,y)) \tag{3.7}$$

二维中值滤波的窗口形状有多种，如线形、方形、十字形、圆形、菱形等。不同的窗口形状产生不同的滤波效果，使用中应根据图像的内容和不同的要求加以选择。

3.3 图像分割

在图像的研究和应用中，人们往往只对图像中的某部分感兴趣，这些部分常称为目标或前景（其他部分称为背景），它们一般对应图像中特定的、具有独特性质的区域。为了辨识和分析目标，需要将它们分离提取出来，在此基础上才有可能对目标进一步利用。图像分割就是把图像分成各具特性的区域并提取出感兴趣目标的技术和过程。这里特性可以是像素的灰度、颜色、纹理等，预先定义的目标可以对应单个区域，也可以对应多个区域。

多年来人们对图像分割提出了不同的解释和表达，借助集合概念对图像分割可给出如下比较正式的定义[32]：

令集合 R 代表整个图像区域，对 R 的分割可看做将 R 分成 N 个满足以下五个条件的非空子集（子区域）R_1，R_2，…，R_N：

① $\bigcup_{i=1}^{N} R_i = R$；

② 对所有的 i 和 j，$i \neq j$，有 $R_i \bigcap R_j = \Phi$；

③ 对 $i = 1,2,\cdots, N$，有 $P(R_i) = TRUE$；

④ 对 $i \neq j$，有 $P(R_i \bigcap R_j) = FALSE$；

⑤ 对 $i = 1,2,\cdots, N$，R_i 是连通的区域。

其中 $P(R_i)$是对所有在集合 R_i 中元素的逻辑谓词，Φ 代表空集。

对图像的分割可基于相邻像素在像素值方面的两个性质：不连续性和相似性。区域内部的像素一般具有某种相似性，而在区域之间的边界上一般具有某种不连续性。所以分割算法可据此分为利用区域间特性不连续性的基于边界的算法和利用区域内特性相似性的基于区域的算法。

实际应用中图像分割不仅要把一幅图像分成满足上面五个条件的各具特性的区域而且需要把其中感兴趣的目标区域提取出来。只有这样才算真正完成了图像分割的任务。图像分割技术的发展与许多其他学科和领域，例如，数学，物理学，电子学，计算机科学等学科密切相关，常用的典型分割方法有：阈值分割，边缘检测和区域增长等[33]。

3.3.1 直方图阈值法

直方阈值法是灰度图像广泛使用的一种分割方法，它基于对灰度图像的这样一种假设：目标或背景内部的相邻像素间的灰度值是相似的，但不同目标或背景上的像素灰度差异较大，其反映在直方图上，就是不同目标或背景对应不同的峰。分割时，选取的阈值应位于直方图两个不同峰之间的谷上，以便将各个峰分开[3,34,35]。

由于彩色图像不仅只有灰度这一个属性，所以使用直方图阈值法会出现很大的不同。大多数方法都是对彩色图像的每个分量(属性)分别采用直方图阈值法。由于彩色信息通常由 R，G，B 或它们的线性/非线性组合来表示，所以用三维数组来表示彩色图像的直方图并在其中选出合适的阈值，并不是一件轻松的工作，另一方面，确定图像中目标的数目计算量也很大。方法之一就是把三维空间投影到一个维数较低的空间，在维数较低的空间选择阈值对图像进行分割。

如果阈值化操作只是在单个颜色分量上进行，则忽略了 3 个颜色分量间的相关性，不能同时考虑 3 个颜色分量的信息。若能够找到一条直线，使得投影在其上的 3 维空间的点能够很好地分开，这样既能对颜色空间进行降维处理又可以同时利用 3 个颜色分量的信息。

阈值分割的优点是实现简单，对于不同类的物体灰度值或其它特征值相差很大时，它能很有效地对图像进行分割。但是对于彩色图像，如何使用颜色信息获得对分割有效的直方图是一个需要仔细思考的问题。彩色图像的阈值分割存在以下缺点：①单独基于颜色信息，而不考虑图像的空间信息得到的区域可能很不完整，对灰度不均匀很敏感；②对于图像中不存在明显色差或颜色范围有较大重叠的图像分割问题难以得到准确的结果。 所以该方法还是经常和其它方法结合起来运用。

3.3.2 基于边缘检测的方法

边界检测的方法广泛应用于灰度图像的分割，这类方法主要基于图像灰度级的不连续性，它通过检测不同区域之间的边界来实现图像的分割，这与认知视觉过程有些相似。依据执行方式的不同，这类方法又分为串行边缘检测技术和并行边缘检测技术。在串行边缘检测技术中，当前像素点是否属于待检测的边缘，取决于先前像素的验证结果；而在并行边缘检测技术中，一个像素点是否属于待检测的边缘，取决于当前正在检测的像素点以及该像素点的一些相邻像素点，这样该模型可以同时用于检测图像中的所有像素点，因而称之为并行边缘检测技术[32]。

串行边界查找方法通常是查找高梯度值的像素，然后将他们连接起来形成曲线表示对象的边缘。串行边界查找方法在很大程度上受起始点的影响，以前检测像素的结果对下一像素的判断也有较大影响。

并行边缘检测主要是指并行微分算子方法。并行微分算子法对图像中灰度的变化进行检测，通过求一阶导数极值点或二阶导数过零点来检测边缘。常用的一阶导数算子有梯度算子、Prewitt 算子和 Sobel 算子，二阶导数算子有 Laplacian 算子，还有 Kirsch 算子和 Wallis 算于等非线性算子。这类算法存在的问题主要是：梯度算子不仅对边缘信息敏感，而且对图像噪声也很敏感，因此对于灰度渐变的图像不能找到正确的边界。

对于灰度图像，灰度边缘是灰度值不连续(或突变)的结果，灰度值不同取决于亮度值(brightness)的变化。然而在彩色图像中，边界所包含的信息远比灰度图像丰富的多。例如，亮度(brightness)相同、色度(hue)不同的目标间的边界也是应当被检测出来的。对应地，彩色图像边界定义为三维色彩空间里的不连续处。对“不连续性”的处理大致可以分为 3 类：①在色彩空间定义“距离”，用色彩之间的距离作为是否连续的量度；②彩色空间由三个灰度图像构成，对每个灰度图像分别进行边缘检测，最后用某种规则综合边缘检测结果；③对彩色图像的三个分量各自形成的边界设定“一致性约束条件”，同时保证三个分量形成的边界具有最大不相关性，对三个分量的边界综合分析得出彩色图像的边界[36]。

值得注意的是：边缘检测不能得到最终的分割结果，因为单独的边缘检测只能产生边缘点，而不是一个完整意义上的图像分割过程，这样边缘点信息需要后续处理或与其他分割算法结合起来，才能完成分割任务。

3.3.3 基于区域的方法

基于区域的分割方法，主要包括区域生长、区域分裂合并和以上方法的组合。区域生长的基本思想是将具有相似性质的像素集合起来构成区域。先对每个需要分割的区域找一个种子像素作为生长的起点，然后将种子像素周围邻域中与种子像素有相同或相似性质的像素(根据某种事先确定的生长或相似准则来判定)合并到种子像素所在区域中。将这些新像素当作新的种子像素继续进行上面的过程，直到再没有满足条件的像素可被包括进来，该区域就长成了。在实际应用此方法时需要解决三个问题：①选择或确定一组能正确代表所需区域的种子像素；②确定在生长过程中能将相邻像素包括进来的准则；③制定让生长过程停止的条件或规则。种子像素的选取常可借助具体问题的特点进行。

区域生长算法的优点是易于实现、计算简单。与阈值分割类似，区域增长也很少单独使用，往往是与其它分割方法一起使用。区域生长的缺点是：①它需要人工交互以获得种子点，这样使用者必须在每个需要抽取出的区域中植入一个种子点；②区域增长方式也对噪声敏感，导致抽取出的区域有空洞或者在局部体效应的情况下将分开的区域连接起来。

区域分裂合并的思想是可以从整幅图像开始通过不断分裂得到各个区域。实

际应用中常先把图像分成任意大小且不重叠的区域，然后再合并或分裂这些区域以满足分割的要求。在这类方法中，常需要根据图像的统计特性设定图像区域属性的一致性测度，其中最常用的测度多基于灰度统计特征，另外也可借助区域的边缘信息来决定是否对区域进行合并或分裂。分裂合并方法不需要预先指定种子点，它的研究重点是分裂和合并规则的设计。但是，分裂合并技术可能会使分割区域的边界被破坏。

对于存在明显相似准则的图像来说，区域生长，区域分裂合并能取得较好的分割结果，因为在分割过程中不仅考虑了颜色信息而且考虑了空间关联信息；同时，由于准则的统计特性，可以有效消除孤立噪声的干扰，具有很强的鲁棒性。然而以上方法的缺点是：生长和分裂合并只能串行进行；分割结果受初始种子点的选择以及生长或分裂合并的过程顺序的影响。

3.3.4 特征空间聚类

特征空间聚类算法不需要训练样本。是一种无监督的统计方法，它是通过迭代地执行分类算法来提取各类的特征值。其中 K-均值、模糊 C 均值(Fuzzy C-mean, FCM) 等是最常用的分类方法。对于彩色图像，颜色空间本身就是一种特征空间，颜色空间聚类方法用于彩色图像分割具有直观易于实现的特点，并且能同时利用 3 个分量的颜色信息。

基于聚类技术的基本原理是将一幅彩色图像聚为特征空间中的几簇，每一簇都对应着图像中的目标。聚类分割方法通常按如下步骤进行：首先是把彩色图像中的像素按照某种聚类规则对应到特征空间中的簇去，当所有的像素都处理完后，再把特征域中的簇映射回空间域，实现图像的分割。

聚类方法存在的主要问题有：①聚类分析不需要训练集，所以聚类时需要提供初始参数，初始参数(特别是聚类数目)的设定对最终分类结果影响较大：②聚类也没有考虑空间关联信息，因此也对噪声敏感。

近年来，随着各学科许多新理论和方法的提出，人们也提出了许多结合一些特定理论，方法和工具的分割技术，如基于数学形态学，统计模式识别理论，人工神经网络，信息论，模糊集合和逻辑，小波变换，遗传算法等等。

虽然存在多种多样的彩色图像分割方法，但是迄今为止，不存在一个可以解决任何分割问题的算法。同时大量实验表明，没有一种分割方法对所有的颜色特征都是有效的，同样也没有任何一种颜色坐标对所有的分割方法都有效。所以，根据需要解决分割问题的特点，选择恰当的分割方法和最佳的彩色空间，设计出有效、可行的针对某类问题的分割算法是解决分割问题的有效途径[36]。

3.4 特征点跟踪

特征点跟踪是计算机视觉中最基础的操作之一，并且是最方便的方式从图像序列中提取运动信息。尽管如此，现存的特征跟踪方法相对来说还是比较粗糙的，大概分为两类：基于对应点匹配的方法和基于纹理相关的方法。

基于对应点的方法是指从每帧图像中提取一系列特征（一般为角点特征），然后将这两组特征进行匹配。该方法不同时要求帧间能够可靠和稳定地检测到相同的特征点。因此一个比较大的缺点就是匹配错误较大。

基于纹理相关性的方法指从第一帧提取一组特征点，该特征点在接下来帧的位置是通过在一个合适大小的窗口中进行全局搜索得到，使得和第一帧中的特征点周围的纹理相关性最好。该方法的缺点是特征点会出现偏移现象，此外不能很好地处理后来帧中出现旋转缩放和斜视的情况。

3.4.1 特征点提取

在各种特征中，特征点是一种稳定的、旋转不变、能克服灰度反转的有效特征。特征点或者感兴趣点是图像中易于确定的特殊点，一般为二维图像亮度变化剧烈的点或图像边缘曲线上具有曲率极大值的点，如角点、直线交叉点、T 型交汇点、高曲率点以及特定区域的中心、重心等等。由于其信息含量很高，可以对视觉处理提供足够的约束，能极大地提高计算速度，使得实时处理成为可能。在图像之间进行可靠的对应点匹配，使得特征点检测在光流计算、目标跟踪、三维场景重构和运动估计等机器视觉方面起着十分重要的作用。

目前的特征点检测算法可以说各种各样，但大致分为两类。第一类为基于梯度的方法：通过提取边缘信息，寻找具有最大曲率或者边缘交叉点，特征点计算值的大小不仅与边缘强度有关，而且与边缘方向的变化率有关；第二类为基于模板的方法：考虑像素邻域点的灰度变化，即图像亮度的变化，将与邻点亮度对比足够大的点定义为特征点。下面介绍 4 种基于模板的检测方法：Kitchen-Rosenfeld 检测算法，Harris 检测方法，Kanade-Lucas-Tomasi 检测算法和 SUSAN 特征点检测算法。

较早的直接基于灰度图像特征点检测是 Kitchen-Rosenfeld 算法，通过模板窗口局部梯度幅值和梯度方向的变换率来计算特征点度量值：

$$C = \frac{I_{xx} I_y^2 + I_{yy} I_x^2 - 2 I_{xy} I_x I_y}{I_x^2 + I_y^2} \tag{3.8}$$

其中 I 为灰度值，I_x 是 I 的一阶偏导，I_{xx} 是 I 的二阶偏导。然后根据 C 与给定的阈值大小关系来判定该点是否是角点。该检测算子常用作其他算子的基准。

Harris 检测方法[37]考虑的是用一个矩形窗在图像上移动，由模板窗口取得原图

像衍生出 2×2 的局部结构矩阵：

$$M=\sum_{x,y}w(x,y)\begin{bmatrix}<Ix^2> & <IxIy>\\ <IxIy> & <Iy^2>\end{bmatrix} \tag{3.9}$$

$w(x,y)$ 为窗口函数，$<Ix^2>,<Iy^2>$ 和 $<IxIy>$ 分别为对 Ix^2，Iy^2 和 $IxIy$ 进行高斯滤波后的值。对该矩阵 M 求取特征值 λ_1 和 λ_2，建立度量函数，$R=\det(M)-k(trace(M))^2$，其中 $\det(M)=\lambda_1\lambda_2$，$trace(M)=\lambda_1+\lambda_2$，根据 R 是否大于 0，即可判断该点是否是角点。

该算法是依赖空间一阶偏导，需要图像平滑来提高性能，但提高稳定性的同时，也降低了定位的精确性，因此模板窗口是由高斯平滑核的大小决定的。该方法具有旋转不变性，在检测 L 形角点时具有较高的可靠性和稳定性，但检测的角点有较大的冗余，需要根据实际经验来确定 R 的阈值。

Kanade-Lucas-Tomasi (KLT) [38]特征点检测算法也是对基于一个计算窗口模板 W×W 下的图像计算局部结构矩阵，计其特征值 λ_1 和 λ_2，根据给定阈值 λ，按照式 $\min(\lambda_1,\lambda_2)>\lambda$ 来判定其是否为角点。这里的关键是阈值 λ 和窗口 W 的大小的确定，W 的大小一般为 2~10，太大的窗口会引起角点偏移，窗口太小则会丢失相距较近的角点。

近年来 SUSAN 特征点检测算法得到越来越多的关注，同值分割吸收核（Univalue Segment Assimilating Nucleus，SUSAN)算法[39]，是基于像素邻域半径为 k 的圆形模板，对每个像素基于其模板邻域的图像灰度计算于角点响应函数(CFR)值，如果大于某一阈值且为局部极大值，则认为该点为角点,一般 k 取 1 或 2。

有不少文献对这四种算法进行了评价[39]，一般情况下 KLT 算法比 Harris 算法检测特征点的质量高，但 KLT 算法适用于特征点数目不多且光源简单的情况，Harris 适用于特征点数目较多且光源复杂的情况。除了对单幅图像能进行特征点检测以外，KLT 算法比 Harris 算法对图像序列的特征点检测效果更好。Kitchen-Rosenfeld 算法和 SUSAN 算法一般来说不适合序列图像的特征点跟踪。在本论文的基于视频序列的立体图像对获取方法中就采用彩色 KLT 特征点检测和跟踪算法，第五章会对此进行详细地描述。

3.4.2 相似性度量

根据匹配基元和方式的不同，目前的特征点匹配算法基本上可分为三类：即基于区域相关(Area-based)的匹配、基于特征(Feature-based)的匹配和基于相位(Phase-based)的匹配。这三类算法的匹配基元不同，因此它们判断对应点匹配的理论依据也有所不同，同时匹配基元的稳定性、致密性和歧义性程度直接决定了各个算法的基本特性。同时，各类匹配算法中不乏一些共有的约束条件，比如 Marr

立体视觉计算理论中提出的唯一性、连续性和外极线约束等。

① 基于区域相关的匹配算法：是把一幅图像中的某一点的灰度邻域作为模板，在另一幅图像中搜索具有相同(或相似)灰度值分布的对应点邻域。该方法可以得到致密的视差场，但是缺点就是计算量大、速度慢，对畸变、旋转等比较敏感和在深度间断处出现容易出现误匹配，并且易受噪声干扰，鲁棒性差。

② 基于特征的匹配算法：不是直接利用图像灰度，而是通过灰度导出的符号特征来实现匹配，因此，具有较强的抗干扰性，相对于基于区域的匹配计算量小得多，速度快，但是匹配结果受特征检测精度的影响，得到的匹配点的数目较少，不能得到浓密的视差图。

③ 基于相位的算法：是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，视差精度可到亚像素级，视差场密集，对各种噪声干扰鲁棒性高，对高频噪声和畸变有很好的抵制作用，而且由于它与人类视觉感知过程的相似性，是当前立体匹配研究的新热点，但其计算比较复杂，计算量也很大。另一方面，对于不同的匹配基元，相似性测度的算法模型可以是通用的。

设$f_l(x,y)$和$f_r(x,y)$分别为左右两幅图，(x_l,y_l)和(x_r,y_r)分别为这两幅图中某特征点的中心，以该点为中心 m×n 大小的领域为基源，分别记为 T 和 S，他们之间的差别度量为：

$$D(T,S)=\sum_{i=1}^{m}\sum_{j=1}^{n}\left[S(i,j)-T(i,j)\right]^2 \tag{3.10}$$

当 D(T, S)达到最小时，两者最匹配，若它们的差异为零，则 T 和 S 完全相同。归一化该式：

$$\mathrm{C}(T,S)=\frac{\sum_{i=1}^{m}\sum_{j=1}^{n}\left[T(i,j)-S(i,j)\right]^2}{\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}T(i,j)^2\cdot\sum_{i=1}^{m}\sum_{j=1}^{n}S(i,j)^2}} \tag{3.11}$$

该方法是以窗口为单位进行匹配计算的，因此选择适当窗口的大小成为匹配准确性的关键。窗口选择的过小时，由于不能包含足够的亮度变化，使得亮度变化与图像噪声的比率很小，从而只能得到一个含有很大误差的视察估计值；同样，若窗口定义的过大，当所定义的窗口区域中点的深度值发生突变时，采用上述计算准则得到的值将不能表示正确的匹配，此时，窗口定义的越大，匹配效果越差，而且窗口越大运算量越大，运算的时间越长。为避免窗口过大所出现的问题，可以采用事先检测图像边界的方法来定义窗口的大小，但边界的检测又是一个很棘手的问题。现有的基于区域的立体匹配算法中很多都是使用固定的窗口大小来进行立体匹配计算的[40]。

3.5 本章小结

本章主要介绍了后续实验中用到的图像处理相关技术。

首先介绍了数字图像的采集，彩色颜色空间模型，以及三种常用的图像滤波平滑方法：邻域平均法、高斯滤波法和中值滤波法；然后对基于直方图阈值法，基于边缘检测，基于区域和基于特征聚类等四大类图像分割方法进行阐述。

最后简要解释了特征点跟踪原理，对四种基于模板检测的特征点提取算法进行比较，列出了基本的相似性度量方法，为第五章中计算帧间视差提供理论基础。

4　基于单视图的立体图像对生成

4.1 引言

人类在漫长的历史长河中积累了大量珍贵的绘画和摄影作品，这些作品本身是二维的平面图像。如果能利用图像处理技术将这些平面图像转换成立体图像，则会给人们带来更加真实的视觉享受。已有一些文献对相关的问题进行了探讨[9,41,42]。这些文献大都采用双目立体成像的基本原理实现平面图像立体化。

如我们在第二章中讨论过的，目前存在的立体图像大多是通过双摄影机或立体摄像机拍摄的，这种方式成本高，拍摄难度大；或者是通过对坐标系中的物体进行透视投影得到，但该方法对复杂的户外实景失去了作用；我们知道平面图像中蕴藏着丰富的三维信息，侯春萍等人提出了一种平面图像立体化方法[9]。该方法利用了心理深度暗示(Psychological Depth Cue)和生理深度暗示(Physiological Depth Cue)的原理[9, 22]，将一幅图像划分成如图 4.1 所示的 $M \times N$ 个图像子块，然后使每一个子块在水平方向上随机的偏离它们原来所处的位置，得到的图像和原图像一起构成立体图像对。处理的流程如图 4.2 所示，具体过程及参数的意义参见文献[9]。

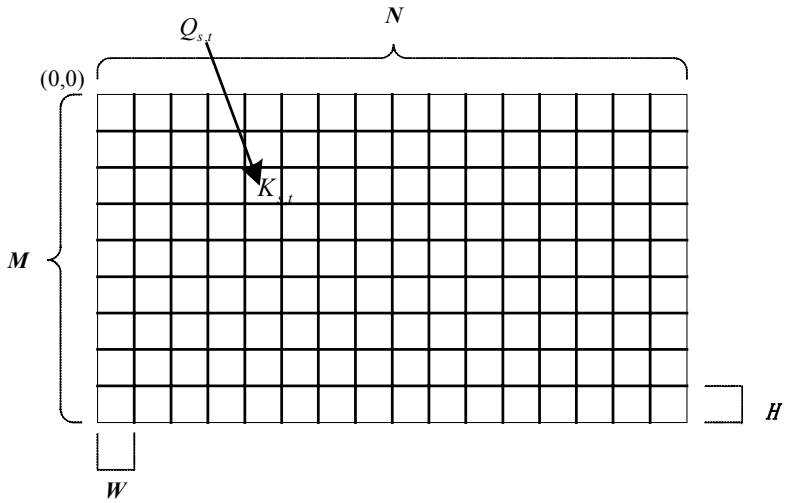


图 4.1 图像分成 M×N 个子块

Fig 4.1 Image divide into M×N block

然而 Hou 等人没有对平面图像立体化的立体效果给出定量的评价指标，也没有深入讨论立体化的实现、随机变量和图像子块的个数对转换后的立体效果的影响。本章以 Hou 方法为基础，在标准的计算机监视器上对普通的平面图像进行了立体化，并较为深入地讨论了这些问题。

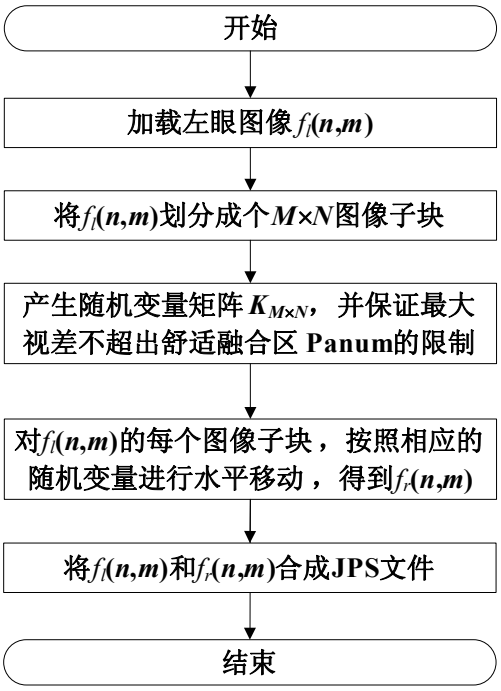


图 4.2 平面图像立体化流程图

Fig 4.2 Procedure of image conversion from planar to stereo

4.2 立体化效果的评价指标

由于每个观察者眼睛本身以及观察方式和习惯的不同，对同一幅图可能会有不同的评价。为了便于比较，我们采用交叉熵和均方根误差这两个定量指标来评价转换效果的好坏。设立体相机拍摄的或根据双目立体视觉的基本原理由软件生成的左右眼视图分别为 *l*Img 和 *r*Img，对 *l*Img 进行立体化，得到大小与 *r*Img 相同的左眼视图 *r*Img'。

① 交叉熵(Cross-entropy)

设 $\boldsymbol{X}$ 是具有 n 个有限状态值的随机变量，$p_i = P\{\boldsymbol{X} = x_i\ , i = 0, 1, \ldots , n\text{-}1\}$。交叉熵 $I_{[P, Q]}$用来度量两种概率分布 $P = \{p_0, p_1, \ldots, p_{n\text{-}1}\}$和 $Q = \{q_0, q_2, \ldots, q_{n\text{-}1}\}$之间的信息量的差异：

$$I_{[P,Q]} = \sum_{i=0}^{n-1} p_i \ln \frac{p_i}{q_i} (q_i \neq 0) \qquad (4.1)$$

这里 p_i 是对灰度图像 P 中灰度值为 i 的像素概率统计，$n = 256$ 表示灰度级数。交叉熵直接反映了两幅图像像素的信息量差异，是评价两幅图像差别的关键指标。一般情况下，交叉熵越小，两幅图越接近。

交叉熵反映的只是整体上一幅图像与另一幅图像之间的信息量差异，有可能

出现这种情况：两幅图像 $\boldsymbol{P}$ 和 $\boldsymbol{Q}$ 内容差别很大，但是交叉熵 $\boldsymbol{I}_{[\boldsymbol{P},\boldsymbol{Q}]}$却比较小，因此有必要参考另外一个指标——均方根误差。

② 均方根误差

设图像的高和宽分别为 $\boldsymbol{M}$ 和 $\boldsymbol{N}$（单位为像素），$r\mathrm{Img}(i, j)$表示图像 $r\mathrm{Img}$ 在像素点(i, j)处的灰度值，则均方根误差：

$$E_{RMS}(r\,\mathrm{Im}\,g', r\,\mathrm{Im}\,g) = \{\frac{1}{mn}[\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[r\,\mathrm{Im}\,g'(i,j) - r\,\mathrm{Im}\,g(i,j)]^2]\}^{1/2} \tag{4.2}$$

如果转换后得到的右眼视图 $r\mathrm{Img}'$与真正拍摄得到的右眼视图 $r\mathrm{Img}$ 之间的交叉熵和均方根误差比较小，则说明两视图较接近，即转换后的立体效果较好。大量的实验和统计结果表明事实的确如此。

4.3 随机变量

在图像的立体化过程中，为了使图像子块在水平方向上的移动具有随机性，我们对每一个图像子块 $\boldsymbol{Q}_{s,t}$ 都设置了一个随机变量 $\boldsymbol{K}_{s,t}$，这样整幅图像就对应一个随机变量矩阵 $\boldsymbol{K}_{M\times N}$ 。当随机变量 $\boldsymbol{K}_{s,t}$ 的均值为 1 时，转换后图像大小总体上不变，且水平视差的均值为 0。图像子块在水平位置上随机的左右移动，由此产生水平方向的**正视差**或**负视差**[9]。

图 4.3 原始图像（左眼视图 lImg + 右眼视图 rImg）

Fig 4.3 Original stereo image (left –eye view lImg+ right-eye view rImg)

下面讨论随机变量服从不同的分布时对立体效果的影响。实验使用的原始图像如图 4.3(每幅图像的大小为 800×600)所示，程序对左眼视图进行立体化。为了便于比较，实验中固定取 $\boldsymbol{M}=150$，$\boldsymbol{N}=200$。

4.3.1 均匀分布

当随机变量 $\boldsymbol{K}_{s,t}$ 服从[0.5, 1.5]上的均匀分布时，均值为 1。

正如我们在第二章中讨论的，无论服从何种分布的随机变量 $\boldsymbol{K}_{s,t}$，还应该使每

个图像子块 $\boldsymbol{Q}_{s,t}$ 在水平方向上的最大位移不超出舒适 Panum 融合区(Panum's fusional area)的限制，即要求转换后原图像的每一个像素(i, j)对应的水平视差 Δq_{ij} [9,22]满足不等式：$-13 < \Delta q_{ij} < 13$。

在该实验中，产生一组符合要求的随机变量的算法如下：

算法：ProduceKmn：产生一个符合要求的随机变量矩阵 $\boldsymbol{K}_{\boldsymbol{M}\times\boldsymbol{N}}$。

$\boldsymbol{K}_{\boldsymbol{M}\times\boldsymbol{N}}$：随机变量矩阵，共 $\boldsymbol{M}$ 行 $\boldsymbol{N}$ 列($\boldsymbol{M}\geq 1, \boldsymbol{N}\geq 1$)。矩阵中每一个元素 $\boldsymbol{K}_{s,t}$ 与 $\boldsymbol{Q}_{s,t}$ 对应。

输入：未填值的随机变量矩阵 $\boldsymbol{K}_{\boldsymbol{M}\times\boldsymbol{N}}$。

输出：填好随机数的符合要求的随机变量矩阵 $\boldsymbol{K}_{\boldsymbol{M}\times\boldsymbol{N}}$。

步骤：

for each row $\boldsymbol{K}_s$ in $\boldsymbol{K}_{M\times N}$ do
{
  repeat
  {
    for each element $\boldsymbol{K}_{[s,\,t]}$ in row $\boldsymbol{K}_s$ produce a random number $\boldsymbol{K}_{s,t} \in [0.5, 1.5]$, α=1;
    if $\boldsymbol{K}_{s,t}$ satisfies the constraint of *Panum's Fusional Area* then
      $\boldsymbol{K}_{[s,\,t]} := \boldsymbol{K}_{s,t}$;
    else
      repeat producing $\boldsymbol{K}_{s,t}$; if $\boldsymbol{K}_{s,t}$ still not satisfies the constraint after three times, clear all elements in the front of $\boldsymbol{K}_{[s,\,t]}$
  }
  until all elements in row $\boldsymbol{K}_s$ have random numbers which satisfy the constraint;
};

将一幅图像划分成$\boldsymbol{M}\times\boldsymbol{N}$个图像子块之后，行与行之间的图像子块对应的随机数没有关系，而每一行内部各图像子块之间对应的随机数是有依赖关系的。因此，为了降低时间复杂度，算法ProduceKmn采取的方法是：每产生一个随机数，就立刻检查它是否符合要求；如果某一图像子块对应的随机数不符合要求，则再重复产生。如果重复三次还是不符合要求，那么这一行对应的随机数就全部重新产生。

表 4.1 中的"均匀分布"列是一组符合要求的服从均匀分布的随机变量取值。根据这组取值得到的右眼视图如图 4.4 (a) 所示。将其和原始图像中的右眼视图进行对比，可得到相应的交叉熵和均方根误差。重复进行实验 10 次，得到实验数据如表 4.2 中"均匀分布"列所示。

表 4.1　服从不同分布的随机变量集取值

Table 4.1 Data sets of random numbers that have different distributions

随机变量	取值					
	均匀分布	分段均匀 (β=0.85)	分段均匀 (β=0.82)	正态分布	三角分布	拟合灰度分布
$\boldsymbol{K}_{1,1}$	0.6065	0.7338	0.9663	0.8622	1.0772	1.2585
$\boldsymbol{K}_{1,2}$	1.2955	1.1746	1.24303	0.8765	1.0549	1.1235
$\boldsymbol{K}_{1,3}$	1.3059	0.6820	0.7338	0.6049	0.7257	0.8461
…	…	…	…	…	…	…
$\boldsymbol{K}_{150,200}$	1.1794	1.0191	0.7107	1.2648	1.0003	0.9798

4.3.2 分段均匀分布

我们知道，在同一时刻，左眼看到物体的左边多点，右眼看到物体的右边多点，从图 4.3 也可以看出，左眼看到的场景成像相对右眼要向右偏一点，因此，当根据左眼视图 *l*Img 生成左眼视图 *r*Img' 时，可将左眼视图划分成均等的左右两部分，左半部分整体上上多压缩一些，右半部分整体上多拉伸一些，以减少 *r*Img 和 *r*Img' 的不同。根据这个思想，我们给出分段均分分布的定义。

设图像对应的随机变量矩阵 $\boldsymbol{K}_{M\times N}$ 的左半部分的随机变量服从[0.5, 0.5 + β]上的均匀分布，右半部分服从[1.5 – β, 1.5]上的均匀分布，其中 $\beta \in (0, 1]$。我们称这种均匀为分段均匀分布(Piecewise Uniform Distribution)。

设每个图像字块的水平宽度为 W，垂直宽度为 H，则像素点(i, j)所在的图像字块 $\boldsymbol{K}_{s,t}$ 为：$s = [i/\mathrm{H}]+1$，$t = [j/\mathrm{W}]+1$，该点在子块内的序号为 $k = j - \mathrm{W}\cdot(t-1)$，$k$ = 1, 2, …, W。如果 q_{ij}^* 为该点经过变化后的水平坐标，那么水平图像视差的数学期望 $\mathrm{E}(\Delta q_{ij})$ 为：

$$\begin{aligned} E(\Delta q_{ij}) &= E(q_{ij}^* - q_{ij}) \\ &= E(W\times[(K_{s,1}+K_{s,2}+\cdots+K_{s,t-1})-(t-1)]+k(K_{s,t}-1)) \\ &= \begin{cases} (\beta-1)[W(t-1)+k]/2, & 1<t\le N/2 \\ (1-\beta)[W(t-1-N)+k]/2, & t>N/2 \end{cases} \end{aligned} \tag{4.3}$$

从公式(4.3)可以看出，当 $t < \boldsymbol{N}$ 时，$\boldsymbol{E}(\Delta q_{i,j}) < 0$；当 $t = \boldsymbol{N}$，$k = \boldsymbol{W}$ 时，$\boldsymbol{E}(\Delta q_{i,j}) = 0$。这说明，左眼视图最后一列像素对应的水平视差的均值为 0，生成的右眼视图没有丢失左眼视图的内容。由这两幅图像构成立体图像对是负视差的情形。

当 β = 0.85，表 4.1 中的第二列是一组服从分段均匀分布且满足限制条件的随机变量取值。根据这组取值得到的右眼视图 4.4 (b)所示。相应的 $I_{[r\mathrm{Img}',\ r\mathrm{Img}]}$和

E_{RMS}(*r*Img', *r*Img)可以根据图 4.4(b)和图 4.3 中的右眼视图得到。重复实验 10 次，将实验得到的数据填入表 4.2 相应的栏中，并根据这些数据在图 4.7 中画出相应的曲线。从数据上显示分段均匀效果较均匀分布要好，通过在显示器上观察证实了该结果。

我们分别取 β 为 0.80，0.81，0.82，…，0.90 等值进行实验，结果显示当 β<0.82 时，产生随机数的算法 ProduceKmn 不收敛；当 β≥0.82 时，β 越小，立体效果越好，但耗时也越长，且可以发现 β 为 0.82 和 0.83 时立体结果无明显区别。

(a) 随机变量服从均匀分布

(b) 随机变量服从分段均匀分布(β=0.85)

(c) 随机变量服从分段均匀分布(β=0.82)

(d) 随机变量服从正态分布

(e) 随机变量服从三角分布

(f) 随机变量服从拟合灰度值分布

图 4.4 右眼视图 *r*Img'

Fig 4.4 The right-eye view *r*Img'

4.3.3 正态分布

当随机变量 $\boldsymbol{K}_{s,t}$ 服从均值 $\alpha = 1$ 的正态分布时，为了便于比较，限制 $\boldsymbol{K}_{s,t} \in [0.5, 1.5]$。根据“$3\sigma$”原则，有 $P\{0.5 \le \boldsymbol{K}_{s,t} < 1.5\} = P\{\alpha - 3\sigma \le \boldsymbol{K}_{s,t} < \alpha + 3\sigma\} = 0.9974$，从中求得 $\sigma = 1/6$，$\boldsymbol{D}(\boldsymbol{K}_{s,t}) = \sigma^2 = 1 / 36$ ($\sigma > 0$)。即当产生一个随机数时，就立即检查它是否落在区间[0.5, 1.5]，如果落在这个区间就保留，否则舍弃。

与 4.3.1 节相似，采用 ProduceKmn 算法，产生随机变量矩阵 $\boldsymbol{K}_{\boldsymbol{M}\times\boldsymbol{N}}$，将实验得到的数据填入表 4.1 和表 4.2 相应的栏中，重复实验 10 次。与均匀分布和分段均匀相比，当随机变量服从正态分布时，得到的右眼视图 *r*Img'与原始右眼视图 *r*Img 之间的交叉熵和均方根误差较小，立体化效果较好。

4.3.4 三角分布

为了进行对比我们通过程序模拟了服从如下分布的随机变量 $\boldsymbol{K}_{s,t}$，即

$$f(K_{s,t}) = \begin{cases} -\dfrac{\pi}{2}\cos(\pi K_{s,t}) & 0.5 \le K_{s,t} \le 1.5 \\ 0 & \text{else} \end{cases} \tag{4.5}$$

此时 $E(K_{s,t}) = \int_{-\infty}^{+\infty} K_{s,t} f(K_{s,t}) dK_{s,t} = \int_{0.5}^{1.5} K_{s,t} \cdot (-\dfrac{\pi}{2}) \cos(\pi K_{s,t}) dK_{s,t} = 1$，我们称之为三角分布。

同样，我们采用该分布重复实验 10 次，将实验得到的数据分别填入表 4.1 和表 4.2 相应的栏中。结果表明，当随机变量服从三角分布时，立体化效果介于均匀分布和正态分布之间。

4.3.5 拟合灰度值的分布

我们可以对原始图像 *l*Img 的灰度值进行统计，拟合灰度直方图来得到一个函数作为随机变量 $\boldsymbol{K}_{s,t}$ 的概率密度函数，通过程序模拟产生服从该分布的随机数，并最终可求得 *r*Img'。

*l*Img 的灰度直方图如图 4.5 所示。我们将其进行归一化，其中横坐标为 $k/255$，纵坐标为 $n_k / \max(n_k)$，n_k 是图像中灰度值为 k 的像素点个数，k=0, 1, 2 ,…,255，得到的归一化灰度值分布如图 4.6 所示，采用最小二乘法对图 4.6 中的离散点进行线性拟合。设 $p(k) = p_1 * k + p_2$，根据 *l*Img 的灰度值的分布可求得：$p(k) \approx$-0.7117* k + 0.7539。它所对应的曲线如图 4.6 所示。

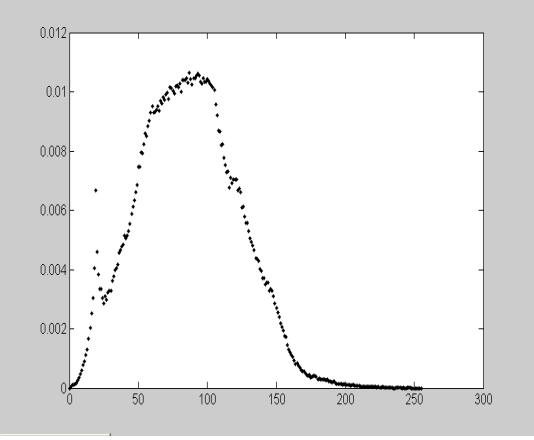


图 4.5 *I*Img 的灰度值分布

Fig 4.5 Distribution of the gray value of *I*Img

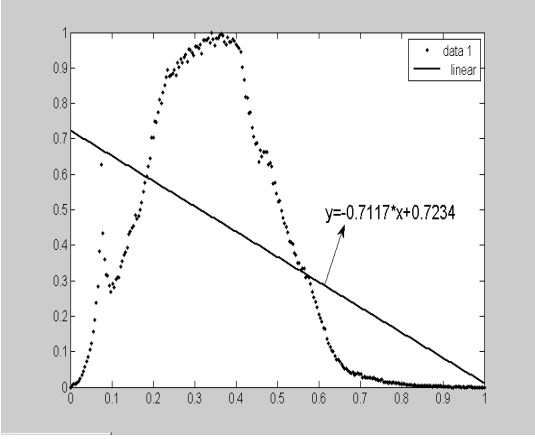


图 4.6 归一化灰度值分布及对应的拟合曲线

Fig 4.6 The normalized distribution of the gray value and the curve of $p(k)$

我们希望转换后的图像大小不变，通过将该曲线向右平移 k_0，向纵轴上移 c_0，使得随机变量 k 的数学期望为 1。设调整后的曲线为 $f(k)=p(k-k_0)+c_0$，那么 $f(k)$ 必须同时满足：

$$\int_{k_0}^{k_0+1} kf(k)=\int_{k_0}^{k_0+1} k[p(k-k_0)+c_0]dk=1$$

$$\int_{k_0}^{k_0+1} f(k)=\int_{k_0}^{k_0+1} [p(k-k_0)+c_0]dk=1$$

则根据以上两个方程可求得：k_0=0.5593，c_0=0.6324，因此：

$$f(k)=\begin{cases}-7117\,k+1.7539 & 0.5593\le k\le 1.5593\\ 0 & \text{else}\end{cases} \tag{4.6}$$

$f(k)$即为随机变量 $\boldsymbol{K}_{s,t}$ 的概率密度函数。然后可用如下方法构造服从概率密度函数为 $f(k)$的随机数：

① 求 $f(k)$的分布函数 $\boldsymbol{F}(k)$：

$$\begin{aligned}F(k)&=\int_{-\infty}^{k} f(k)dk=\int_{0.5593}^{k}(-0.7117\,k+1.7539)dk\\&=-0.3559k^2+1.539k-0.8696\end{aligned} \tag{4.7}$$

② 因为 0<$F(k)$<1，令 r= $F(k)$，求 $\boldsymbol{F}(k)$的反函数 $\boldsymbol{F}^1(r)$：

$$k=F^{-1}(r)=2.4640-\sqrt{3.6281-2.8096r} \tag{4.8}$$

先产生[0, 1)区间上服从均匀分布的随机数，再代入式 4.8 就可计算得到服从概率密度函数为 $f(k)$的随机数。

模拟产生的随机数同样也必须不超出舒适 Panum 融合区的限制，重复实验 10 次，将实验得到的数据分别填入表 4.1 和表 4.2 相应的栏中。根据第一组数据所得到的图像见图 4.4(f)。

表 4.2　随机变量服从不同分布时的交叉熵 $I_{[r\mathrm{Img}', r\mathrm{Img}]}$ 和均方根误差 $E_{RMS}(r\mathrm{Img}', r\mathrm{Img})$

Table 4.2 Cross-entropy and root-mean-square errors of gray scale

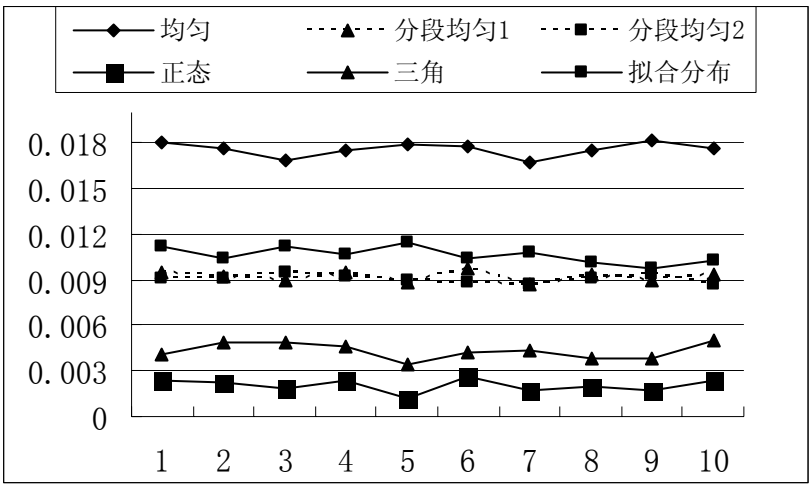
实验次数	均匀分布		分段均匀 1(β=0.85)		分段均匀 2(β=0.82)	
	$I_{[r\mathrm{Img}', r\mathrm{Img}]}$	$E_{RMS}(r\mathrm{Img}', r\mathrm{Img})$	$I_{[r\mathrm{Img}', r\mathrm{Img}]}$	$E_{RMS}(r\mathrm{Img}', r\mathrm{Img})$	$I_{[r\mathrm{Img}', r\mathrm{Img}]}$	$E_{RMS}(r\mathrm{Img}', r\mathrm{Img})$
1	0.01806	39.9229	0.009473	38.5195	0.009123	38.3765
2	0.01758	40.0173	0.009202	38.2196	0.009067	38.4596
3	0.01686	40.1146	0.00890	38.4959	0.009461	38.6081
4	0.01747	39.7917	0.009482	38.4949	0.009275	38.4786
5	0.01784	40.0173	0.008789	38.5870	0.008927	38.1726
6	0.01776	40.0588	0.009764	38.5349	0.008777	38.3618
7	0.01670	39.8302	0.008628	38.5813	0.008668	38.6903
8	0.01747	40.1110	0.009407	38.6177	0.009127	38.464
9	0.01810	39.8957	0.008922	38.4466	0.009369	38.2474
10	0.01763	40.1807	0.009320	38.3381	0.008670	38.4685
平均值	0.0175	39.9940	0.009190	38.4840	0.009050	38.433
标准差	0.0004	0.1228	0.00035	0.1160	0.00027	0.1457

表4.2 随机变量服从不同分布时的交叉熵$I_{[r\mathrm{Img}', r\mathrm{Img}]}$和均方根误差$E_{RMS}(r\mathrm{Img}', r\mathrm{Img})$ (续表)

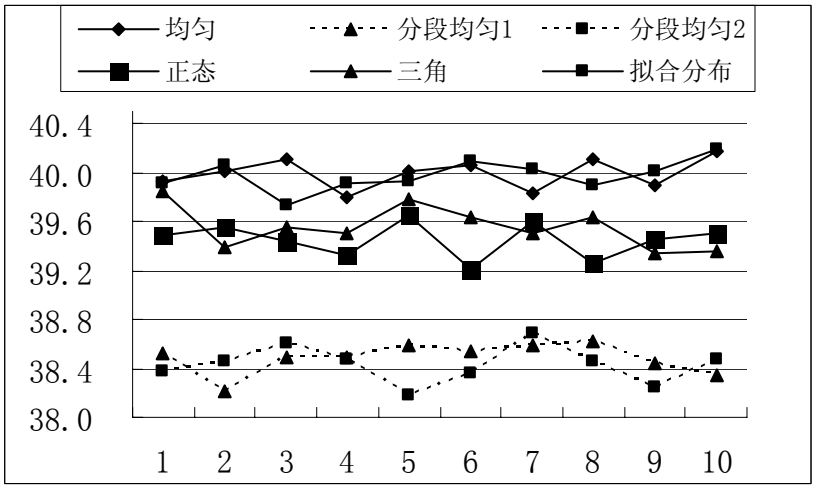
Table 4.2 Cross-entropy and root-mean-square errors (continued)

实验次数	正态分布		三角分布		拟合灰度分布	
	$I_{[r\mathrm{Img}', r\mathrm{Img}]}$	$E_{RMS}(r\mathrm{Img}', r\mathrm{Img})$	$I_{[r\mathrm{Img}', r\mathrm{Img}]}$	$E_{RMS}(r\mathrm{Img}', r\mathrm{Img})$	$I_{[r\mathrm{Img}', r\mathrm{Img}]}$	$E_{RMS}(r\mathrm{Img}', r\mathrm{Img})$
1	0.002408	39.4920	0.004144	39.8466	0.01112	39.9130
2	0.002276	39.5462	0.004910	39.3905	0.01035	40.0656
3	0.001904	39.4358	0.004846	39.5574	0.01115	39.7282
4	0.002342	39.322	0.004541	39.5108	0.01065	39.9146
5	0.001170	39.6516	0.003477	39.7859	0.01146	39.9326
6	0.002691	39.2035	0.004266	39.6348	0.01035	40.0989
7	0.001704	39.5933	0.004313	39.5107	0.01084	40.0188
8	0.002004	39.2538	0.003826	39.6402	0.01015	39.8901
9	0.001717	39.4462	0.003831	39.3317	0.00968	40.0053
10	0.002372	39.4959	0.005024	39.3523	0.01028	40.1939
平均值	0.002060	39.4440	0.004320	39.5560	0.01060	39.9760
标准差	0.00043	0.1378	0.00049	0.1655	0.00050	0.1232

4.3.6 实验结果分析



(a) 交叉熵 (Cross-entropy) $I_{[r\text{Img}', r\text{Img}]}$



(b) 均方根误差（root-mean-square errors）$E_{RMS}(r\text{Img}', r\text{Img})$

图 4.7　根据表 4.2 中的数据得到的曲线图

Fig 4.7 the graph according to the result of table 4.2

从图 4.7 可以看出，分段均匀分布比均匀分布的交叉熵和均方根误差都小，但若 β 取值较小则转换速度减慢，且对于不同的图像 β 的最佳取值会有所不同，需通过多次试验得到；拟合灰度值的分布较均匀分布效果好，但没有三角分布和正态分布优秀； 分段均匀分布的均方根误差是五种分布中最小的；然而正态分布的交叉熵和均方根误差相对其他来说都比较理想。实际观察显示，正态分布的效果要比分段均匀分布的效果更稳定一些，并且多数情况下比其他分布要好。

理论上，$I_{[r\text{Img}', r\text{Img}]}$和 $E_{RMS}(r\text{Img}', r\text{Img})$的最佳值应该都是 0，这时表明 rImg' 和 rImg 几乎没有差别。但从表 4.3 和图 4.5 可以看出，不管随机变量服从何种分布，在每次实验中随机变量取不同的值，但 $I_{[r\text{Img}', r\text{Img}]}$和 $E_{RMS}(r\text{Img}', r\text{Img})$ 都没有衰减为 0 的趋势；相反地，它们基本上都稳定在平均值附近。重复更多次实验我们发

现，$I_{[r\mathrm{Img}',\ r\mathrm{Img}]}$和 $E_{RMS}(r\mathrm{Img}',\ r\mathrm{Img})$ 本身也是随机变量，并且相互是独立的。这表明当随机变量服从同一种分布时，随机变量的取值对立体效果没有多大的影响。当随机变量服从不同分布时，不同分布的交叉熵和均方根误差会有所不同，这表明不同的分布会对立体效果产生一定的影响。一般说来，正态分布的立体效果较好，实际观察也证实了这一点。

4.4 图像子块数目

我们知道该立体化转化方法需要将一副原始图像划分成 $M\times N$ 个图像子块，因此有必要讨论参数 M 和 N 对立体效果的影响。实验使用的原始图像同图 4.3，随机变量 $K_{s,t}$ 服从均匀分布，且满足 $0.5\le K_{s,t}\le 1.5$。程序对左眼视图进行立体化，实验过程和评价指标的计算与第 4.3 节相同，改变 M 和 N 的值进行多次实验，实验结果在表 4.3 中显示。

表 4.3 M 和 N 取不同值时的交叉熵和均方根误差

Table 4.3 Cross-entropy and root-mean-square error of gray scale (M and N are set different pairs)

实验次数	M	N	$I_{[lImg',\ lImg]}$	$E_{RMS}(lImg',\ lImg)$
1	1	4	0.006896	39.6582
2	1	8	0.01433	41.4028
3	1	16	0.02279	38.3179
4	1	25	0.03524	40.332
5	1	32	0.06348	40.1543
6	1	80	0.04234	39.6801
7	1	100	0.03670	40.4305
8	1	160	0.02622	39.5668
9	1	200	0.01949	39.0941
10	1	400	0.00293	38.8545
11	1	800	0.00391	38.2515
12	3	4	0.00415	40.3318
13	6	8	0.01094	39.0847
14	12	16	0.02687	40.0367
15	24	32	0.04080	40.2996
16	60	80	0.03884	40.4347
17	120	160	0.02488	40.2022
18	150	200	0.01806	39.9229
19	300	400	0.001095	39.5908
20	600	800	0.005147	38.9223
…	…	…	…	…

从表 4.3 可以看出：

从交叉熵和均方根误差这两个评价指标来看，M 取 1(对原始图像只做垂直划分)和 M 取其它值之间的差别不是很大。并且，两个指标也没有衰减为 0 的趋势。从理论上分析这种现象是可以理解的：因为随机变量 $K_{s,t}$ 服从均匀分布，并且都是对同一副图像 *l*Img 立体化。可以计算出 $I_{[l\text{Img},\ r\text{Img}]} = 0.005149$，对应的 $E_{RMS}(l\text{Img}, r\text{Img})=38.92347$。根据上述立体化方法可知，$E_{RMS}(r\text{Img}', r\text{Img})$ 不应偏离 $E_{RMS}(l\text{Img}, r\text{Img})$太远，$I_{[r\text{Img}',r\text{Img}]}$也类似。因此这两个指标没有衰减为 0 的趋势。

实验中发现，当 M = 1 时，相当于对原始图像只做垂直划分，这样得到的图像较为平滑，所以从实验的效果来看，看到的立体图像要平滑一些，不过立体感要弱一些。在 M = 1 的情况下，当 N 趋于 *r*Img 的宽度时，看到的立体图像变得越来越平滑；但是立体感变得越来越弱，直到最后没有立体感。从理论上分析这种现象是可以理解的：当 N 趋于 *r*Img 的宽度时，差不多几个像素就构成了一个图像子块，由于像素的坐标需要取整数值，这就造成不少像素没有移动，因此得到的 *r*Img'与 *l*Img 相差不大，立体感就会减弱。

我们可以得出结论：在 p_d 为一个合适的值的前提下，图像子块的个数取不同的值对立体化效果的综合影响不大，只对立体图像的平滑和立体感产生微弱的影响。当图像变得平滑时立体感就变弱，反之，立体感就变强。平滑与立体感是互相矛盾的，人们只能取一个折中值。

当随机变量 $K_{s,t}$ 服从其他分布时，也有类似的结论。

4.5 本章小结

通过计算机技术将历史上优秀的平面图像和绘画作品转换成立体版，是一件很有价值的事情，获取同一场景的立体图像对是实现平面图像立体化的一个关键问题。

本章基于 Hou 方法在计算机监视器上对普通的平面图像进行了立体化，给出了评价转换后的立体效果的定量指标(以交叉熵和均方根误差两个定量指标对转换效果进行评价)，并讨论了各参数对立体效果的影响。

我们分析比较了随机变量服从五种不同分布的情况，从实验结果可知，在大多数情况下，当图像子块在水平方向上的偏移量服从均值为 1 的正态分布时，具有最优的转换效果，其余从优到差依次为三角分布，分段均匀分布，拟合灰度值分布和均匀分布。且当随机变量矩阵中的每一个随机变量都服从同一种分布时，随机变量的取值对立体效果没有多大的影响。

当监视器屏幕与观察者的距离为一个合适的值(例如 1 m)时，图像子块的个数取不同的值对立体效果的综合影响不大，只对立体图像的平滑和立体感产生微弱

的影响。平滑与立体感是互相矛盾的，人们只能取一个折中值。

平面图像立体化的转换效果的评价指标——交叉熵反映的只是整体上一幅图像与另一幅图像之间的信息量差异，实验表明，单独使用交叉熵来评价转换效果是不准确的。我们首先可以考察交叉熵，在交叉熵比较小的情况下，再考察灰度均方根误差，一般地，交叉熵和灰度均方根误差的值越小效果越好。可以用这个指标控制平面图像立体化的过程。

在讨论随机变量对转换效果的影响时，我们注意到算法 ProduceKmn 需要对产生的随机数进行检查，因此这个算法在很坏的情况下有可能变得不稳定。在最好的情况下，算法 ProduceKmn 的时间复杂度是 O(M×N)。在一定的情况下，特别是 $\Delta q_{t,k}$ 的取值范围变得较小时，计算量急剧增大，甚至不收敛。因此，需要进一步研究如何保证和加快算法的收敛速度。

5　基于视频序列的立体转换

5.1 引言

第4章介绍的基于侯方法的立体转换主要用于单幅静态图像的立体化。随着立体成像显示设备的不断研发成功，三维视频将会成为下一代多媒体的主流显示方式，比如，立体地展现植物的生长过程。因此，有必要研究视频的立体化。

如绪论中所述，自从第一张胶片发明以来，人类历史上积累了大量优秀的二维视频，因此一个比较可行的方式是将已有的单目二维视频转换成三维立体版，已有一些文献对相关的问题进行了探讨。Matsumoto等利用运动视差原理，通过相邻两帧确定摄像机的运动参数，由此得到对象的深度信息，产生立体图像对[43]。但是对相邻帧中所有像素点进行匹配，时间复杂度很高，且在无明显纹理的地方匹配准确度较差。在文献[44]中，Harman等采用机器学习算法产生关键帧的深度图，该方法需要手动输入大量已知点，作为“机器学习”的训练数据。

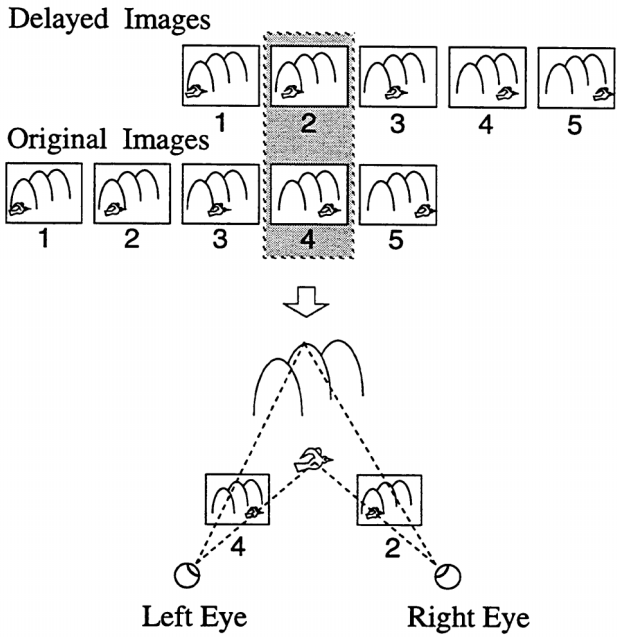


图 5.1 基于单目视频图像序列的立体图像对生成原理

Fig 5.1 Principle of stereo pairs creation based on video frequency image sequence

人类视觉具有很强的空间和时间感知能力。当追踪一个运动物体时，在每一时刻，双眼会看到视场中不同的部分，但却有可能在不同的时刻看到相同的场景。基于文献[45]中时空插值（spatio-temporal interpolation）立体转化方法的理论基础，本章提出了一种通过在帧间选择立体图像对的方式将单目视频转换成双目立体视

频：首先通过特征点跟踪算法获得某一帧中的特征点在后续帧中的位移信息，并据此计算两帧间视差；然后在相邻帧中为原始帧选择合适的立体图像对，使之具有最适宜的视差。图5.1是对该方法可行性一种粗略定性解释。

5.2 视差对深度感知的影响

虽然我们视网膜上成像是二维平面的，但感知到的世界却是立体三维的。多种深度暗示如视差、遮挡、阴影以及明亮等在晶状体的调节和收缩作用下，共同影响着我们对三维立体空间中对象大小和位置的理解。但是起决定作用是双目视差，Bela Julesz演示的随机点立体图像对就是对其很好的证明[19]。

5.2.1 视差对深度感知的影响

在帧间选择立体图像对的重要选择依据是视差，它影响着人们的**深度感知**(depth perception)。因此，必须先讨论视差对深度感知的影响。

如图 5.2 所示，图中所标参数的具体含义参见 3.1 基于监视器的双目立体成像模型，由该模型可知，水平视差 $D = x_r - x_l = 2h\dfrac{o_d - p_d}{o_d}$。

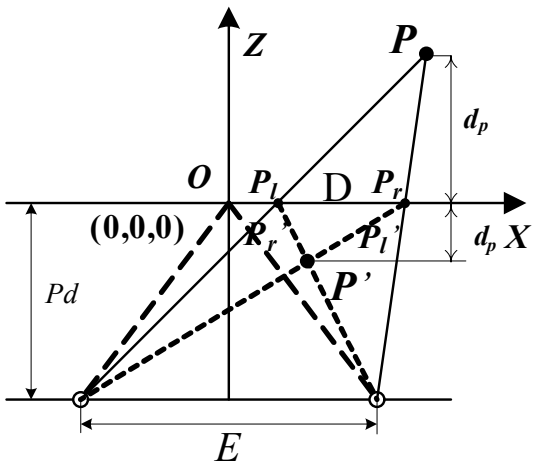


图 5.2 立体成像系统

Fig 5.2 Stereo imaging system

设 $\boldsymbol{E}$ 是瞳距，则 $\boldsymbol{E} = 2h$。设 d_p 是人眼感觉到的像相对于屏幕的深度：$o_d - p_d = d_p$，我们称之为**视觉距离**，它是深度感知的一种体现。

因此，感知到的深度和视差之间的关系[46,47]可描述为：

$$\frac{D}{E} = \frac{d_p}{p_d + d_p} \tag{5.1}$$

如果将$\boldsymbol{d_p}$表示成D的函数形式，则：

$$d_p = p_d \frac{D}{E - D} \tag{5.2}$$

因此对于左右眼视图中的两个对应点 $P_l(x_l, y_l)$和 $P_r(x_r, y_r)$ ，如果能被融合成一个虚拟三维像点(x, y, z)，那么其坐标为：

$$\begin{cases} x = x' \dfrac{E}{E - D} \\ y = y' \dfrac{E}{E - D} \\ z = p_d \dfrac{D}{E - D} \end{cases} \tag{5.3}$$

其中 $x' = (x_l + x_r)/2$， $y' = (y_l + y_r)/2 = y_l = y_r$。

p_d和E 是常量，因此成像点坐标仅受视差变化影响。式5.2表明如果想要成像点位于头部和监视器之间，视差必须小于零，此时被称为交叉视差或者负视差。当$\boldsymbol{D} < 0$时，有$d_p<0$，在这种情况下，我们将式5.2改写成视差绝对值的形式：

$$| d_p | = p_d \frac{| D|}{E + |D|} \tag{5.4}$$

视差的绝对值越大，成像离监视器越远，反之亦然。图 5.3 给出了$|D|$ 和 $|d_p|$理论上的关系曲线图：当视差较小时，曲率很大，立体感对视差很敏感；当视差变大时，曲线变得平滑，此时视差的变化对立体感的影响变小。

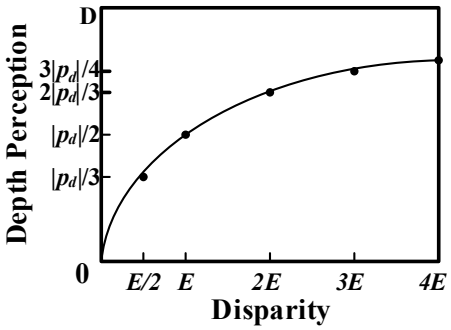


图 5.3 视差和深度感知之间的关系

Fig 5.3 Relationship between disparity and depth perception

5.2.2 舒适视差限制

我们已经知道立体图像对之间的视差不能超出视网膜的视差范围(Panum融合区)，否则观察者不能将其融合，产生重影，并会感到疲劳，甚至头晕。当采用双目时分立体显示系统时，交叉视差的融合限制被认为大约在27'，非交叉视差为24'，当用像素单位度量时：-13≤ D≤ 13。

通过对真实立体图像对的分析，我们得到了同上面分析一致的结论。如图 5.4 所示，每幅图像大小为 800×600, 颜色分辨率为 RGB 256×256×256. 我们采用结合了彩色边缘检测和种子区域增长的半自动对象提取算法[48]，将图中的棱形框分别提取出来，结果如图 5.5 所示。 表 5.1 给出了每个棱形框的中心点坐标以及对应框间的目视差。

图 5.4 原始立体图像对

Fig 5.4 Original stereo image pairs

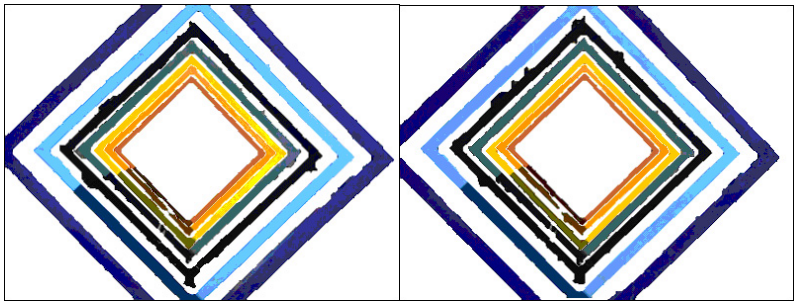


图 5.5 图像对中提取出的对象

Fig 5.5 Objects extracted from the stereo pair

表 5.1 每个矩形框的中心以及对应的双目视差(单位：像素)

Table 5.1 Center point of each square and their corresponding disparity（in pixel）

矩形框	左眼视图		右眼视图		视差	
	X_c	Y_c	X_c	Y_c	水平	垂直
1	383.3	300.4	379.0	302.3	-4.3	-1.9
2	381.6	300.9	377.2	299.6	-4.4	1.3
3	378.8	301.2	373.7	302.4	-5.1	-1.2
4	377.2	302.6	371.4	304.6	-5.8	-2.0
5	373.8	301.0	366.6	299.4	-7.2	1.6
6	378.6	297.0	367.2	296.2	-11.4	0.8

表 5.1 显示，从里到外对应棱形框间的视差越来越大，成像离屏幕也越远。但

是存在垂直方向的视差，这是由于摄像机标定和对象提取算法的精确度导致。Sobel通过实验得出：如果垂直视差在一定范围内，对立体深度感知起到非常小的影响[49]。通过对其他类型的立体图像进行分析后，得到相似的结果，当视差在-5 到 -9个像素之间时，观察者可感到舒适的立体感。因此在接下来的实验中，我们选择具有帧间视差在这范围的两帧组成立体对。

5.3 立体图像对选择算法

考虑一段通过摄像机运动拍摄到的静止场景的普通单目视频序列，特别是当摄像机只有水平运动的时候，可以通过在序列中选择不同的帧组成立体图像对，关键问题是决定合适的延时方向和延时时间，以得到舒适的立体感。

因为在静止场景的视频序列帧中，没有唯一具体的对象，且是在没有任何先验知识的情况下，现有的对象自动提取算法对提取复杂场景下的复杂对象，效果都不尽人意，因此，利用帧中特征点在相邻帧中的加权平均位移，来作为我们选择的依据，该加权平均位移被我们称为帧间视差。权重由特征点在相邻两帧中的跟踪性能决定。非相邻帧间视差是由相邻帧间视差累加得到。

对于给定的一段单目视频序列，我们将其作为左眼视频，通过以下两个步骤得到每帧对应的右眼视频序列帧：首先提取序列第一帧中的特征点，通过跟踪算法获得其在后续帧中的位置信息；然后计算帧间视差，选择与该帧具有最适宜视差的两帧作为立体帧。

5.3.1 特征点跟踪算法

特征点跟踪是计算机视觉中最基础的操作之一，并且是最方便的方式从图像序列中提取运动信息。在本实验中采用彩色 KLT 特征点检测和跟踪算法[50]。

设彩色图像序列中每个像素的色彩表示为时变向量函数 $I(X,t)=\left[Ir(X,t)\quad Ig(X,t)\quad Ib(X,t)\right]^T$，其中 $X=[x,y]^T$ 为像素坐标，$Ir(X,t)$，$Ig(X,t)$，$Ib(X,t)$ 分别表示 t 时刻 **X** 处像素在 RGB 色彩空间 3 个分量的强度值。如果相邻两幅图像之间的时间间隔很短即采样频率很高，那么可以认为以点 X 为中心的较小图像窗口 W 经过了某种几何变换后在 $t+\tau$ 时刻色彩强度保持不变，即

$$I(X,t)=I(\delta(X),t+\tau)\quad X\in W \tag{5.5}$$

式中 $\delta(X)$ 表示 X 处因运动而产生的几何变换。在图像序列的高频采样假设下，可以近似认为每个小窗口作纯平移运动，也就是 $\delta(X)=X+d$，其中 $d=[dx\quad dy]^T$ 是 t 时刻该窗口内所有像素的平移向量。由此，特征跟踪的任务就是在图像序列的每对相邻两帧图像中为一组自动选取的特征点计算其相应的平

移向量 d 值。

为避免噪声干扰，采用相邻图像中每个特征点所在窗口内像素的方差和（SSD）作为跟踪残差，即

$$\varepsilon = \sum_{w} \left\| I(X+d, t+\tau) - I(X,t) \right\|^2 \tag{5.6}$$

将上式使用泰勒级数展开：$I(X+d,t+\tau) \approx I(X,t) + \nabla I(X,t)^T d + I_t(X,t)\tau$

因此整理式 5.6 得到：

$$\varepsilon \approx \sum_{W} \left\| \nabla I(X,t)^T d + I_t(X,t)\tau \right\|^2 \tag{5.7}$$

其中：

$$\nabla I(X,t) = g = \begin{bmatrix} \partial I/\partial x \\ \partial I/\partial y \end{bmatrix} = \begin{bmatrix} \partial Ir/\partial x & \partial Ig/\partial x & \partial Ib/\partial x \\ \partial Ir/\partial y & \partial Ig/\partial y & \partial Ib/\partial y \end{bmatrix} = \begin{bmatrix} Ir_x & Ig_x & Ib_x \\ Ir_y & Ig_y & Ib_y \end{bmatrix},$$

$I_t(X,t) = I_t = \partial I/\partial t = \begin{bmatrix} \partial Ir/\partial t & \partial Ig/\partial t & \partial Ib/\partial t \end{bmatrix}^T$。

显然，能使式(5.7)最小化的 d 值就是最优的平移向量值，因此对 ε 求 d 的偏导，并令其为 0，则有：$\dfrac{\partial \varepsilon}{\partial d} = \sum_{W} 2(gg^T d + \tau g I_t) = 0$。

设

$$e = -\tau \sum_{W} g I_t \tag{5.8}$$

$$G = \sum_{\mathrm{W}} gg^T = \begin{bmatrix} \mathrm{A} & \mathrm{C} \\ \mathrm{C} & \mathrm{B} \end{bmatrix} \tag{5.9}$$

其中：$\begin{cases} A = \sum_{W} Ir_x^2 + Ig_x^2 + Ib_x^2 \\ B = \sum_{W} Ir_y^2 + Ig_y^2 + Ig_y^2 \\ C = \sum_{W} Ir_x Ir_y + Ig_x Ig_y + Ib_x Ib_y \end{cases}$。

最终得到：

$$Gd = e \tag{5.10}$$

若 G 可逆，则平移向量的估计为：$d = G^{-1}e$。由于像素坐标都位于整数网格上，为了在跟踪过程中获得的特征点坐标达到子像素精度，采用 *Newton-Raphson* 优化方法对平移量求精，并且在迭代过程中使用双线性插值方法获得小数坐标位

置处的色彩强度。设 $d_k = G^{-1}e$ 为第 k 次迭代时估计的平移量，那么最小化残差 ε 的迭代算法是

$$\begin{cases} d_0 = 0 \\ d_{k+1} = d_k + G^{-1}\sum_W \left[g(I(X,t) - I(X + d_k, t+1)) \right] \end{cases} \tag{5.11}$$

其中 $I(X + d_k, t+1)$ 表示在 $t+1$ 时刻 $X + d_k$ 坐标处的色彩强度向量。如果超过 20 次迭代还不收敛，则放弃该点，标记跟踪失败。

5.3.2 特征点选择准则

在跟踪过程中，并不是图像中所有的部分包含了完整的运动信息，比如在水平的色彩强度边缘只能确定运动的垂直分量，这就是所谓的孔径问题（aperture problem）。在有图像噪声和区域变形的情况下，可以考虑图像上多方向色彩强度变化为一种稳定的结构，因此可以选用在图像中高纹理区域的角点作为需要跟踪的特征点[39,51]。

实际上，2×2 系数矩阵 G 代表了很好的度量。也就是说如果 G 不受图像噪声影响并且具有很好的制约，我们就可以在连续帧中跟踪一个窗口。噪声要求暗示 G 的两个特征值都必须很大，制约要求意味着数量上的几阶也不会有变化。

如果 λ_1 和 λ_2 为 G 的两个特征值，并且 $\lambda_1 < \lambda_2$,那么：

$$\lambda_{1,2} = \frac{A + C \pm \sqrt{(A-C)^2 + 4B^2}}{2} \tag{5.12}$$

实际上，同特征值 $\lambda_{1,2}$ 关联的特征向量 $h_{1,2}$ 相互垂直，且 $\lambda_{1,2}$ 分别对应窗口滑动时 SSD 最大和最小的两个方向。通过分析 $\lambda_{1,2}$ 的大小，图像中每个点可以有以下分类结果：

(1) $\lambda_1 \approx \lambda_2 \approx 0$，表明特征窗口向任意方向作单位移动其 SSD 很小，那么该点处没有明显的结构特征；

(2) $\lambda_1 \approx 0, \lambda_2 >> 0$，表明特征窗口仅在某一方向(边缘的法向)作单位移动时 SSD 较大，可见该点位于色彩边缘上；

(3) λ_1 和 λ_2 都很大而且不同，说明特征窗口向任何方向作单位运动会产生较大的 SSD，这表明该点为角点或其他纹理性较强的点。

因此可选择 $\lambda = \min(\lambda_1, \lambda_2) > \lambda_{th}$ 的点作为特征点。但是由于图像纹理的不均匀性，若 λ_{th} 过小，则可能会在某局部（如树叶，草丛等）检测到大量密集的特征点；而若 λ_{th} 过大，则不能检测到足够的特征点。基于以上分析，对于某点如果其 λ 值大于某个较小门限值 λ_{th}，并且是局部最大值，可作为特征点进行跟踪，即特征点 P_i 的集合 F 为：

$$F = \{P_i \mid \lambda_i > \lambda_{\text{th}} \ \& \ \lambda_i = \max_{j \in window}(\lambda_j), \ \ i = 1,2...,n\} \tag{5.13}$$

基于统计原理，在实验中我们根据图像内容自适应获取 λ_{th}。对于前一帧中每个像素点，计算其 7×7 窗口矩阵 G，然后得到其最小特征值 λ，并按 n 个等级，将其进行直方图统计，对应概率分布为 $P = \{p_0, p_1, \ldots, p_{\text{n-1}}\}$；我们取 $\lambda_{\text{th}} = \lambda_k$，$k$ 为使得 $\sum_{i=0}^{k} p_i \geq percent$ 的最小值，我们知道一幅图像中，角点和边缘线占少数，大部分区域为均匀区域，因此实验中这个百分比取 80% 到 95%之间。

5.3.3 整体视差估计

在正常情况下，残差应该呈高斯分布(μ, σ)，如果某个特征点的跟踪残差超过一定范围，则被认为跟踪错误，即 outliers。采用 Median Absolute Deviation (MAD) 作为跟踪性能评估主要依据[52]：

$$MAD = \underset{i}{\text{med}}\{|\varepsilon_{\text{i}} - \underset{j}{\text{med}}\varepsilon_{\text{j}}|\} \tag{5.14}$$

基于 X84 拒绝准则，因此我们将满足$|\varepsilon_{\text{i}} - \text{med}\varepsilon_{\text{j}}| > kMAD$ 的认为是 outliers。由于 $MAD = \Phi^{-1}(3/4)\ \sigma \approx 0.6745\sigma$，并且在区间 $[u-3\sigma, u+3\sigma]$ 上包含了大于 99.7%的分布，所以可得到 $k = 4.4$。

根据特征点在相邻两帧中跟踪的残差，我们采用下式作为跟踪性能评价：

$$w_i = \begin{cases} 1 - \dfrac{\varepsilon_i}{kMAD + med\varepsilon_j} & |\varepsilon_{\text{i}} - \text{med}\varepsilon_{\text{j}}| < kMAD \\ 0 & else \end{cases} \tag{5.15}$$

然后帧间视差就是通过对相邻帧中所有特征点位移的加权均值得到，权重由跟踪性能决定：

$$\begin{cases} dx = \dfrac{\sum_{i=1}^{N} w_i \cdot dx_i}{\sum_{i=1}^{N} w_i} \\ dy = \dfrac{\sum_{i=1}^{N} w_i \cdot dy_i}{\sum_{i=1}^{N} w_i} \end{cases}, \tag{5.16}$$

但是考虑户外视频序列，特别是当有远处的天空和山时的情况，空间跨度很大，从远景中提取出来的特征点在相邻帧中几乎不动，因此我们添加粗糙的垂直信

息来改进视差：

$$\begin{cases} dx = \dfrac{\sum_{i=1}^{N} \dfrac{y_i}{height} \cdot w_i \cdot dx_i}{\sum_{i=1}^{N} \dfrac{y_i}{height} \cdot w_i} \\ dy = \dfrac{\sum_{i=1}^{N} \dfrac{y_i}{height} \cdot w_i \cdot dy_i}{\sum_{i=1}^{N} \dfrac{y_i}{height} \cdot w_i} \end{cases} \tag{5.17}$$

对于非相邻帧间的视差是通过相邻帧间视差累积得到，即第 k 与 k+2 帧间的视差为第 k 与 k+1 帧间视差加上第 k+1 与 k+2 帧间视差获得。

5.3.4 算法描述

我们对上述所有描述进行总结，给出彩色视频立体转化算法的详细步骤：

① 根据式5.9和式5.12，计算第k帧中每个像素点的特征值，记录较小的特征值，选择满足式5.13的点作为特征点；

② 利用式5.11计算得到每个特征点在后续帧即第k+1帧中的位移；

③ 评估每个特征点的跟踪性能，根据式5.16或者5.17计算帧间视差；

④ 每一帧选择合适的右眼视图帧，使得其帧间视差大约在-5 到-9 个像素之间。

5.4 实验结果

表 5.2 帧间视差与对应右眼视图帧 (单位：像素)

Table 5.2 Inter-frame disparity and corresponding right-eye view frame（in pixel）

帧序号		**1**	**2**	**3**	**4**	**5**
Tsukuba	相邻帧间视差	-5.1,0.08	-7.1,-0.07	-4.3,-0.3	-3.6,0.09	-3.9,0.1
	右眼视图帧号	2	3	5	6	7
	对应帧间视差	-5.1	-7.1	-7.9	-7.5	-7.6
Flower Garden	相邻帧间视差	-6.1,0.03	-4.2,0.07	-4.0,0.2	-4.0,0.4	-3.8,-0.01
	右眼视图帧号	2	4	5	6	7
	对应帧间视差	-6.1	-8.2	-8.0	-7.8	-7.5
Chinese Garden	相邻帧间视差	-3.6,0.2	-3.4,-0.1	-2.6,-0.1	-2.4,-0.1	-4.4,0.2
	右眼视图帧号	3	4	5	6	7
	对应帧间视差	-7.0	-6.0	-5.0	-6.8	-7.3

实验中，我们测试了各种视频源，三段从第 1 帧到第 5 帧的序列段的转换效果在表 5.2 中给出。首先 tsukuba 图像序列通过我们的方法被转换成双目版，以第

2 帧和第 3 帧为例，如图 5.6 所示，第 2 帧中检测到 26 个特征点，每个特征点在第 3 帧中的位移分别为-4.7, -5.1, -5.0, -5.2, -6.2, …, -8.0, -5.7, -11.3, -10.5, -5.5 个像素。其中最大和最小位移分别为-11.3 和-4.6 个像素。由式 5.15 和 5.16 计算得加权平均位移，即水平帧间视差为-7.1 个像素，因此第 3 帧被选择作为第 2 帧的右眼视图。通过在显示系统中观察，可感觉到一定程度的稳定舒适的深度。其他帧对，如第 3 帧和第 5 帧，第 4 帧和第 6 帧有相似的结果。

图 5.6 Tsukuba 序列中第 2 帧和第 3 帧

Fig. 5.6 Tsukuba sequence frame 2nd and frame 3th

然后我们测试了著名的花园视频序列。第 2 帧中检测到 53 个特征，最大和最小位移分别为-10.1 和-2.0 个像素，与第 3 帧平均水平视差为-4.2 像素；第 3 帧和第 4 帧间的视差为-4.0 像素，因此我们选择第 4 帧与第 2 帧组成立体图像对，其间的水平视差为-8.2 像素。

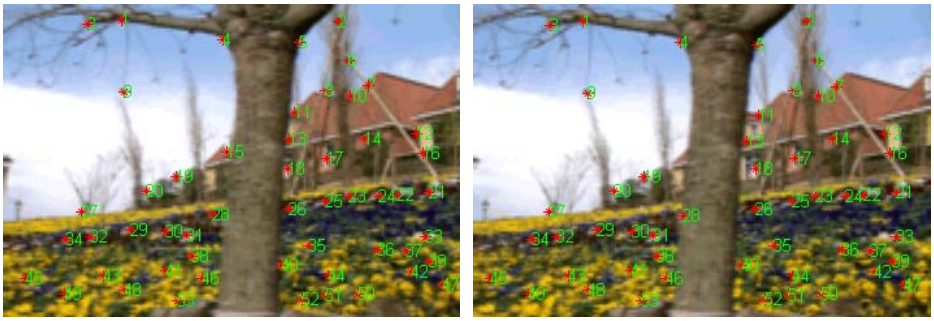


图 5.7 花园序列第 2 帧和第 3 帧

Fig. 5.7 Flower garden frame 2nd and frame 3th

测试了一段关于中国传统园林的视频，结果在表 5.2 中给出。图 5.8 显示的是其中一对立体对第 1 帧和第 3 帧，具有帧间水平视差-7.0 个像素。

图 5.8 中国传统园林序列第 1 帧和第 3 帧
Fig. 5.8 Chinese traditional garden frame 1st and frame 3rd

当然，对其他类型的视频我们也进行了测试，结果表明当摄像机或者镜头中的物体做近似水平方向的运动，则能获得较好的立体图像对；但当运动只存在垂直方向时，结果不是很准确。

5.5 本章小结

本章首先解释了视差与深度感知的关系，介绍了特征点提取和跟踪算法的原理，然后提出了一种转换现有单目视频序列的立体图像生成算法。该方法根据特征点在帧中的相对位置，计算相对帧间视差，为每帧在序列中寻找合适的右(左)眼视图。通过对户外和室内视频序列的测试，实验和分析表明，多数情况下，采用帧间视差在-5 到-9(像素)之间的两帧构成立体图像对，能得到较为稳定的位于头部和监视器屏幕之间的虚像；且当摄像机或者场景中的对象具有近似水平方向的运动时，该方法能获得较好的性能。

当只存在垂直方向的运动时，或视频序列中有太多的关键帧时，立体图像对不容易找到，或者找到了但是立体效果差。如何消除垂直方向运动限制是一件不太容易的事。一种可行的方法是采用**视图变形**(View Morphing)技术将某一帧图像作为原始图像 $\boldsymbol{I_0}$，其它帧 $\boldsymbol{I_i}$ 通过前置变换方法变换到同一个平行平面上，得到图像 $I_0^{'}$ 和 $I_i^{'}$，且满足扫描线特性。即在不改变摄像机光学中心的前提下，将两幅图像进行对齐。此外，还可以把深度和视差结合到立体化方法中。这些是我们下一步研究工作的重点。

6 总结与展望

6.1 工作总结

论文完成的主要工作有：

① 研究了基于计算机立体视觉的双目立体成像系统和模型，详细讨论了成像相关技术及理论。简要总结了现有各种成像系统，对实验中所用的显示方式及硬件设备进行了详细介绍。

② 学习了图像处理中的各种基本概念，深入理解了图像彩色空间模型，图像滤波和图像分割等技术与理论；并重点分析了图像特征点提取和跟踪原理。

③ 针对侯春萍提出的基于单幅图像的立体图像对生成方法，给出了评价转换后的立体效果的定量指标(以交叉熵和均方根误差)，详细讨论了各种参数对立体效果的影响，并编程实现。该方法无需对图像进行高级处理，只需将图像子块进行随机平移，且该过程与图像内容无关，是一种快速获得一定立体感的有效方法。

④ 根据时空视差原理，提出了一种通过在帧间选择立体图像对的方式将单目普通二维视频转换成双目立体视频：首先通过KLT彩色跟踪算法提取第一帧中的特征点，得到这些特征点在相邻帧中的位移信息，并据此计算帧间视差；然后根据帧间视差，为每原始帧选择合适的立体图像对。该方法在充分利用现有资源的基础上，通过对图像特征点的提取，跟踪和分析，得到序列帧间视差关系，无需三维建模，实现容易。

以上工作涉及到计算机视觉、模式识别、人工智能、认知心理学、计算机图形学等领域的研究，是计算机视觉与计算机图形学等多个领域交叉的学科之一，具有较为广泛的理论研究价值和应用前景。

6.2 展望

立体视觉作为一门多学科的交叉科学，用各种成像系统代替视觉器官作为输入敏感手段，由计算机来代替大脑完成处理和解释。目前，计算机视觉已在遥感图像分析、文字识别、医学图像处理、多媒体技术、图像数据库、工业在线检测与军事上的目标自动识别跟踪等方面等领域取得了广泛应用。通过对基于双目立体成像系统和模型的研究，我们认为还有必要在针对以下问题进行深入研究：

① 我们知道，离观察者越近的物体，双眼视差越大；反之，离观察者越远的物体，双眼视差越小，而基于侯方法的单视图立体转化方法中，并未考虑图像内容本身，将前景和背景统一对待。因此，可先通过图像分割提取技术，将其分成不同深度的几部分，各自采用不同的转换参数，可能会提高转换效果。

② 通过在帧间选择立体图像对的方式，只适合于摄像机或场景中的物体做近似水平方向上的运动的情形，如何消除垂直方向运动限制是一件不太容易的事。

把深度和视差结合到该方法中，可能会是一种比较有效的消除垂直方向运动限制的方案，或者通过**视图变形**技术中的前置变换方法，先将不同帧变换到同一个水平平面上。虽然现有一些通过图像对间立体匹配方式，从一段视频序列中得到每帧对应的深度信息的方法，但稠密匹配始终没有很好的解决；因此解决匹配问题是所有问题的根本和关键，这也是我们接下来要开展的研究工作。

③ 由于自由立体显示器无需佩戴任何眼镜就能看到立体效果，它代表了未来立体显示设备发展的方向。本文的研究是建立在以标准的 CRT 监视器作为显示设备的基础之上的，且需要佩戴液晶光阀眼镜。因此，有必要研究基于自由立体显示器的双目立体成像。

致　　谢

在论文完成之际，我要衷心感谢许多老师、同学、朋友和亲人的帮助与鼓励，我今天的成绩是和重庆大学老师们的悉心关怀和精心指导分不开的。

在攻读硕士学位的三年时间里，导师在学业上给予我耐心的指导，且给予我们宽松愉快的研究环境和众多的机会，使得我顺利地完成了研究生阶段的学习。导师渊博的知识、严谨的治学风范、积极的人生态度、勤奋工作和无私的奉献精神使我深受启迪。导师实事求是的科研精神、不断开拓创新的学术思维和高度的责任感使我终身受益。从尊敬的导师身上，我不仅学到了扎实、宽广的专业知识，也学到了做人的道理。在此，谨向朱老师表示我崇高的敬意和衷心的感谢!

感谢软件中心的王茜副教授、付鹤岗副教授，感谢计算机学院所有老师们，在重庆大学七年的学习生活中，他们给了我很多有益的指导和帮助。

此外，还要感谢实验室里关心我的师兄弟、师姐妹们，特别是刘然师兄，在课题选择和论文写作上给于很大的帮助。

最后还要感谢我的父母和朋友，他们在我困难的时候帮助我、关心我、给我无比的信心和勇气，我取得的成绩是和他们的关心和鼓励分不开的。

同时衷心地感谢在百忙之中评阅论文和参加答辩的各位专家、教授!

支丽欧

二 OO 七年十月 于重庆

参考文献

[1] 马颂德，张正友编．计算机视觉．北京：科学出版社，1998．

[2] 吴立德编. 计算机视觉．上海：复旦大学出版社, 1993．

[3] George Sperling．Binocular Vision: A Physical and a Neural Theory． The American Journal of Psychology, 1970, 83(4)：461-534

[4] 游素亚，徐光．立体视觉研究的现状与进展．中国图象图形学报，1997，2(1)：17-24

[5] Larry F Hodges．Tutorial: Time-Multiplexed Stereoscopic Computer Graphs．IEEE Computer Graphics & Application，1993，3：10-20

[6] Jean Hsu, etc.．Issues in the Design of Studies to Test the Effectiveness of Stereo Imaging．IEEE Transactions on System, Man and Cybernetics- Part A: Systems and Human．1996，26(6)：810-819

[7] Myeung-Sook Yoh．The Reality of Virtual Reality．Proceedings of the seventh International Conference on Virtual Systems and Multimedia (VSMM'01)，IEEE Computer Society， 2001，1-9

[8] Frederick P.Brooks, Jr．What's Real About Virtual Reality?．IEEE Computer Graphics and Applications，1999，Nov-Dec，16-27

[9] 侯春萍，俞斯乐．一种平面图像立体化的新方法．电子学报，2002，30(12)：399-402

[10] 邾继贵等．单摄像机虚拟立体视觉测量技术研究．光学学报，2005，25(7)：943-948

[11] 王新宇．基于计算机立体视觉的三维重建：学位论文．湖南：中南大学，2004．

[12] Cassandra T. Swain．Integration of Monocular Cues to Create Depth Effect．IEEE International Conference on Acoustics, Speech, and Signal processing (ICASSP). 1997，4，2745-2748

[13] K Yamada，K Suehiro，H Nakamura．Pseudo 3D Image Generation with Simple Depth Models．International Conference on Consumer Electronics，ICCE Digest of Technical Papers. 2005，277-278

[14] Y. Matsumoto, H. Terasaki, K. Sugimoto, and T. Arakawa．Conversion System of Monocular Image Sequence to Stereo using Motion Parallax．SPIE Photonic West, 1997, 3012：108-115

[15] P Harman, J Flack, S Fox, and M. Dowley．Rapid 2D to 3D Conversion．SPIE Stereoscopic Displays and Reality Systems IX, 2002，4660：78-86

[16] 段华．基于双目立体视觉的计算机三维重建：学位论文．南京：南京航空航天大学，2003．

[17] Display Technology: Stereo & 3D Display Technologies David F. McAllister

[18] 蒋庆全. 说说立体三维显示与立体三维电视．中国电子科技集团公司南京电子工程邮局所，http://www.chinafpd.net/magazine/ts_con.asp?id=5223，2006.10

[19] Bela Julesz. Binocular Depth Perception without Familiarity Cues. Science, New Series, 1964，145(3630)：356-362

[20] 顾郁莲，蔡宣平．计算机立体视图绘制技术．国防科技参考，1998，19(1)：63-70

[21] 周丽萍．虚拟现实中立体视觉的研究．计算机应用，1999，19（4）：24-26

[22] 侯春萍, 阿陆南, 俞斯乐．立体成像系统数学模型和视差控制方法．天津大学学报，2005，38(5)：455-465

[23] 曾芬芳编．虚拟现实技术．上海:上海交通大学出版社, 1997.

[24] 李克彬，李世其．3D 显示技术的最新研究进展．计算机工程， 2003，29（12）

[25] William R.Sherman, Alan B.Craig．Understanding virtual reality: interface, application, and design．1st ed. San Francisco, CA: Morgan Kaufmann Publishers, 2004

[26] NVIDIA Corporation．NVIDIA 3D Stereo User's Guide. 2 ed．2001

[27] Mingyue Ding, Lixia Yang．3D stereoscopic imaging and its application．Acta Electronica Sinica．1995, 23(10)：124-8.

[28] Kebin Li, Shiqi Li．The Newest Research of 3D Display． Computer Engineering. 2003, 29(12)：3-4

[29] 石魏．基于人类视觉特征的彩色图像分割技术研究：学位论文．山东：中国海洋大学，2006．

[30] 贾云得编．机器视觉．北京：科学出版社, 2000．

[31] 颜色 Munsell 系统，http://www.worldlingo.com/wl/services/S1790.5/translation? wl_srclang=PT&wl_trglang=zh_cn&wl_rurl=http%3A%2F%2Fen.wikipedia.org%2Fwiki%2FValue_%2528colorimetry%2529&wl_url=http%3A%2F%2Fen.wikipedia.org%2Fwiki%2FMunsell_color_system

[32] 章毓晋编．图象分割．北京：科学出版社，2001．

[33] 林开颜，吴军辉，徐立鸿．彩色图像分割方法综述．中国图象图形学报，2005，10(1)：1-10

[34] 吴一全，朱兆达．图像处理中阈值选取方法 30 年（1962—1992）的进展（一）．数据采集与处理，1993，8(3)：193-201

[35] 吴一全，朱兆达． 图像处理中阈值选取方法 30 年（1962—1992）的进展（二）．数据采集与处理，1993，8(4)：268-279

[36] 张晓芸．彩色图像分割算法的研究与实现：学位论文．重庆：重庆大学，2005 年．

[37] Chris Harris and Mike Stephens．A Combined Corner and Edge Detection．Proceedings of The Fourth Alvey Vision Conference, Manchester, 1988, 147-151

[38] Jianbo Shi, Carlo Tomasi．Good Feature to track．IEEE Conference on Computer Vision and Pattern Recognition (CVPR'94)．1994, 593-600

[39] P. Tissainayagam and D. Suter．Assessing the performance of Corner Detectors for Point Feature tracking Applications．Image and Vision Computing．2004, 22(8)：663-679

[40] 孟晶晶．基于区域增长的立体匹配算法的研究：学位论文．大连：大连理工大学，2006．

[41] Qingsheng Zhu, Ran Liu, Xiaoyan Xu．Properties of a Binocular Stereo Vision Model．Proceedings of the 11th Joint International Computer Conference．Chongqing, China, World Scientific Press．2005, 831-834

[42] Indri Atmosukarto, Anna Cavender, Chandrika Jayant．Reconstructing Antique Stereo Pairs．http://www.cs.washington.edu /homes/cjayant/finalproject/0-paper.html, 2006, 3

[43] Y. Matsumoto, H. Terasaki, K. Sugimoto, and T. Arakawa．Conversion System of Monocular Image Sequence to Stereo using Motion Parallax．SPIE Photonic West．1997, 3012：108-115

[44] P. Harman, J. Flack, S. Fox, and M. Dowley．Rapid 2D to 3D Conversion．SPIE Stereoscopic Displays and Reality Systems IX. 2002, 4660：78-86

[45] B. J. Garcia．Approaches to Stereoscopic Video Based on Spatio-Temporal Interpolation．SPIE Photonic West. 1990, 2635：85-95

[46] D. Drascic and P. Milgram．Positioning Accuracy of virtual stereographic pointer in a real stereoscopic video world．SPIE. 1991, 1457：58-69

[47] L. Bouguila, M. Ishii and M. Sato．Effect of Coupling Haptics and Stereopsis on Depth Perception in Virtual Environment．Proceedings of the 1st Workshop on Haptic Human Computer Interaction. Glasgow, Scotland. 2000, 54-62

[48] Yu Xiaohan, and Juha Yla-Jaaski． Image segmentation Combining Region Growing and Edge Detection．11th IAPR International Conference on Pattern Recognition, Vol.III. Conference C: Image, Speech and Signal Analysis, Proceedings. 1992, 481-484

[49] Erik C. Sobel, Thomas S. Collett．Does Vertical Disparity Scale the Perception of Stereoscopic Depth?．Proceeding: Biological Sciences.1991,244(1310)：87-90.

[50] 蔡涛，李德华，朱洲等．基于彩色图像序列的特征点检测和跟踪．计算机工程，2005，31(8)：12-14

[51] Carlo Tomasi, Takeo Kanade．Detection and Tracking of Point Features．Technical Report CMU-CS-91-132, Carnegie Mellon University, April 1991

[52] A. Fusiello, E. Trucco, and T. Tommasini, etc．Improving Feature Tracking with Robust Statistics．Pattern Analysis & Applications．1999, 2：312–320

附　　录

A. 作者在攻读硕士学位期间发表的论文目录

[1] 朱庆生，支丽欧，刘然等. 平面图像立体化关键技术研究. 计算机科学. 2007，34(7)：225-228

[2] 刘　然，朱庆生，许小艳，支丽欧等．基于监视器的双目立体视觉的立体效果．同济大学学报(已录用) ．

B. 作者在攻读硕士学位期间参加的科研项目

[1] 高等学校博士学科点专项科研基金（SRFDP），项目编号：20050611027，虚拟作物生长可视化关键技术研究

[2] 重庆大学研究生科技创新基金，项目编号：200701Y1A0080194，基于计算机立体视觉的双目立体成像研究