



计算机科学

COMPUTER SCIENCE

基于人眼双目视觉的 3D 深度运动感知特性研究

路平, 张地, 肖俊峰, 毕科

引用本文

路平, 张地, 肖俊峰, 毕科. 基于人眼双目视觉的 3D 深度运动感知特性研究[J]. 计算机科学, 2022, 49(10): 176-182.

LU Ping, ZHANG Di, XIAO Jun-feng, BI Ke. [Study on 3D Motion-in-Depth Perception Based on Binocular Vision](#)[J]. Computer Science, 2022, 49(10): 176-182.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于 HVS 的水下图像质量评价](#)

Underwater Image Quality Assessment Based on HVS

计算机科学, 2022, 49(5): 98-104. <https://doi.org/10.11896/jsjcx.210100224>

[物流业几何尺寸智能化测量系统](#)

Intelligent Geometry Size Measurement System for Logistics Industry

计算机科学, 2018, 45(8): 218-222. <https://doi.org/10.11896/j.issn.1002-137X.2018.08.039>

[基于小波子带系数预测的深度图像编码](#)

Depth Map Coding Based on Wavelet Inter-subband Coefficients Prediction

计算机科学, 2014, 41(10): 134-138. <https://doi.org/10.11896/j.issn.1002-137X.2014.10.031>

[基于 MSB 和 HVS 的空域信息隐藏算法的研究](#)

Study of Spatial Information Hiding Algorithm Based on MSB and HVS

计算机科学, 2012, 39(9): 89-93.

[一种基于视觉模型的 DCT 域公开水印算法](#)

DCT Based Public Watermarking Algorithm with Visual Model

计算机科学, 2009, 36(7): 281-283. <https://doi.org/10.11896/j.issn.1002-137X.2009.07.069>

基于人眼双目视觉的 3D 深度运动感知特性研究

路平¹ 张地² 肖俊峰³ 毕科¹

1 北京邮电大学理学院 北京 100876

2 中国传媒大学数据科学与智能媒体学院 北京 100024

3 北京邮电大学电子工程学院 北京 100876

(luping@bupt.edu.cn)

摘要 获取立体视觉信息是人们感知世界的基本能力之一,其中运动立体视觉信息的获取不仅是生物视觉系统在动态世界中生存的关键能力,也是人工视觉系统高效处理立体视频的重要手段。为了设计出贴合人眼视觉特性的 3D 深度运动感知模型,需要明确挖掘人类对立体运动感知的显著特征后再设计主观实验。文中根据单目和双目线索设计了立体运动视觉刺激视频,并采用控制变量的方法,分别探究了目标球体运动速度、参考球体旋转半径及被试正确感知的相互关系,并分析了目标运动方向与被试感知方向之间的关系。首先对实验条件进行探究,结果发现目标与参考的相对距离越小拦截成功率越高,该结果表明目标同参考之间的相对位置关系会影响被试的感知正确性。其次,采用了拦截成功百分比和感知偏差两个行为测量标准对被试感知能力进行分析,结果发现相比横向运动,深度运动正确感知的成功率超出约 42.67%~47.01%。这表明不同运动方位的感知不对称,且深度运动带来的视觉刺激更明显。该研究探索了人类对运动感知的显著特征,并为后续设计 3D 运动感知模型提供了一种新的判断模型感知效果的主观对比标准,使原有立体感知能力的指标更加细化。

关键词: 人类视觉系统;3D 视频;运动感知;双目视差;深度运动

中图法分类号 TP391

Study on 3D Motion-in-Depth Perception Based on Binocular Vision

LU Ping¹, ZHANG Di², XIAO Jun-feng³ and BI Ke¹

1 School of Science, Beijing University of Posts and Telecommunications, Beijing 100876, China

2 School of Data Science and Smart Media, Communication University of China, Beijing 100024, China

3 School of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China

Abstract Obtaining stereoscopic information is one of the basic abilities of human beings to perceive the world. Through stereo vision, we can judge the shape, size, distance, relative position of objects, as well as the direction and speed of changes in object motion information. Among them, the perception information of moving objects plays an important role in stereo vision perception. The acquisition of motion visual information is not only the key ability of biological vision systems to survive in a dynamic world, but also an important means for artificial vision systems to efficiently process stereoscopic video. Therefore, in order to design a 3D depth motion perception model that conforms to the visual characteristics of human eyes, it is necessary to explicitly excavate the salient features of human perception of stereoscopic motion, so as to design experiments to explore. In this paper, motion stereo videos are designed as visual stimuli based on monocular and binocular cues, and subjective experiments are designed using the control variable method. The experiment explores two parts: the influence of the relative distance between the target and the reference sphere on the subjects' perception ability, and the relationship between the actual movement direction of the target and the subjects' perception direction. Experimental data is analyzed by using two behavioral measures: the percentage of successfully intercepted targets and the perceived bias. The conclusion shows that, firstly, the smaller the relative distance between the target and the reference, the higher the interception success rate. The target velocity and the reference's motion radius affect the relative distance of the target and reference spheres. This indicates that the relative positional relationship between the target and the reference plays an important role in the human eye's perception of moving objects. Motion perception has a certain relativity, and motion is easier to perceive at the position with reference point and the position close to the reference point. Second, we find that perceptions elicited by deep motion are more pronounced than those induced by lateral motion. The correct intercept rate of perceived depth direction is 42.67%~47.01% higher than that of lateral motion. This shows that the visual stimulation brought by deep motion is more obvious, and the perception ability of objects moving in different directions is asymmetric. However, when

到稿日期:2022-05-30 返修日期:2022-07-11

通信作者:张地(dizhang@cuc.edu.cn)

there is an interception error in the depth direction, the perceptual deviation is larger, and the deviation is about 0.1583~0.3665. This study explores the salient features of human perception of motion and provides insights into the observer's process of motion perception in 3D environments. This study explores the significant characteristics of human motion perception, and provides a new subjective contrast standard to judge the perception effect of 3D motion perception model for the subsequent design of 3D motion perception model work, which makes the original stereo perception ability index more refined.

Keywords Human visual system, 3D video, Motion perception, Binocular parallax, Motion-in-Depth

1 引言

立体视觉信息的感知是我们获取立体世界信息的基本能力。通过立体视觉我们可以判断物体的一系列特性、物体间相对位置关系、物体运动变化的方向和速度等静态和动态视觉信息。其中运动视觉信息的获取对于生物视觉系统和人工视觉系统都有重要意义,因此引起了人们的广泛关注。为了拟合出生物感知特性的3D视频运动感知模型,我们设计了被试感知立体视频序列中运动物体的主观实验,以探究人眼感知立体深度运动的典型特征;结合双目融合和大脑神经机制分析特征深入挖掘生物视觉的工作机制,为模型提供了贴近人眼感知能力的对比标准;深入拓展了实验的研究角度,以期提出更全面的描述人运动感知能力的指标。

1.1 相关工作

1.1.1 双目视觉信息的加工与处理

人眼视觉系统(Human Visual System, HVS)是人类获取外界视觉信息的生理基础。双目接收到外部世界的影像后由视网膜传入视觉中枢。其在水平方向上的差异会造成视角上的细微差别, Wheatstone把这种眼间差异的作用阐述为大脑通过结合来自双目的图像信息来观察视觉空间^[1]。

视觉信号自视网膜经过外侧膝状体传至大脑皮层的视觉中枢,在层层传递的过程中对视觉信息进行加工处理。大脑皮层中的视觉信息分别由背侧和腹侧两条主要的信息加工皮质通路进行进一步处理,其中参与处理物体空间位置信息、相关运动控制的皮质通路是背侧通路,它包含枕叶到后顶叶的一系列脑区,由初级视觉皮层V1/V2等区域延伸到MT/MST等更高级视觉中枢区域。

不同视皮层区域具有检测和编码各种不同视觉特征的能力。例如目前研究最广泛的初级视皮层,它在运动深度感知中体现出对方位的选择性,当视觉刺激信号匹配到细胞的最优方位时,细胞激活产生强烈响应;刺激方位偏离最优方位时,细胞被抑制,响应变弱;直到刺激方位与最优方位垂直时,细胞的响应最低^[2]。Poggio等对被训练的猴子进行研究,同时记录了视觉皮层初级(V1)次级(V2)区域神经元反应,证实了视觉皮层中的神经元确实对空间中不同深度的物体有选择性地作出反应^[3]。Tanaka等发现在内侧颞上区(MSTd)背侧的神经元对观察者和三维世界间的相对运动产生的光流模式具有选择性^[4]。Xu等提出螺旋空间模型表明MSTd区域的神经元以一种连续的方式调谐基本的光流类型,从而实现现实中运动的响应^[5]。

针对双目视觉信息融合的问题,Hubel等在针对猫的纹状皮层存在双眼驱动细胞的研究中第一次明确了来自双眼的视觉信息在很早的阶段就会发生汇合^[6]。Hyson等通过记录

收敛的眼球运动,全局对齐双目图像,发现图像分离收敛度误差和眼睛收敛度之间的差异高达3°。他们做出假设:大脑感知融合时发生了神经重映射以补偿视网膜的错位^[7]。Ding等提出了以最小错位和最大相关性判定双目匹配对的方法,总结出静态运动中双眼视觉感觉融合机制的概念图式^[8]。此外,对场景相关的物体运动的准确感知需要整合自我运动的视觉和非视觉信息^[9]。目前为止,双眼融合的详细机制尚未通过生理学研究手段直接得到解答。

1.1.2 立体视频运动感知模型

立体视频运动感知模型的设计通常受人眼视觉系统(HVS)观看世界方式的启发,通过传统方式或神经网络的方式建立数学模型拟合HVS各部分的功能特性实现对立体视频中深度运动的感知。

作为视觉特征中占据主导地位的三维运动,最早的研究是将深度感知和运动感知分别进行处理,建立了相对独立的两套简化神经系统模型。直到1979年Gibson提出的光流场概念将二者紧密结合在一起。Julesz提出的随机点立体图(RDS)排除了立体视觉的单眼线索,得出3D运动研究中视差(Changing Disparity, CD)是产生立体视的充分条件的结论^[10]。在后期研究中,研究人员常常将眼间速度差异(IOVD)与视差(CD)两种线索结合考虑。Cormack等提出了一种双光流场,当给定真实场景结构和三维运动时,它能够描述观察者双目捕获视觉信息的几何模型^[11]。在无监督情况下, Kim等考虑了长视频中视差和光流间的时空关系,设计了能够感知视差和运动光流的联合框架^[12]。

1.1.3 被试感知主观实验

被试感知立体视频的主观实验通常从两个方面出发,一种是采用实际物体作为视觉刺激,如Maltz等期望探究熟悉物体的大小是否影响人对物体真实大小和距离的感知,制作了不同尺寸的魔方作为视觉刺激,由被试提供对物体大小和距离的感知情况^[13];另一种是依据射影几何定律分别设计双目场景的刺激,借助外设观看刺激,此类情形更为普遍,如Thompson等采用随机点立体图作为视觉刺激,通过改变随机点在视野中的位置和方向,令其远离或靠近观察者,以此研究双目线索对运动感知的贡献^[14]。Whitner等设计了双目反向的正弦光栅来作为视觉刺激,描述比较物体运动时速度和视差之间的时空整合性^[15]。Peter等设计了基于OpenGL渲染纹理的虚拟房间,被试在房间中移动,改变所处位置,并在移动之前和之后均对球体位置和方向做出判断,探究观察者在自身移动的情况下对距离方向的判断机制^[16]。

1.2 主要工作成果

本文根据单目和双目线索设计了立体运动视觉刺激视频,通过分析视频中物体的运动速度、运动方向、与参考物体

相对位置关系等因素对感知状况的影响,探究人类感知运动的显著特征。实验发现:首先,运动目标同参考物体之间的相对位置关系会影响被试的感知正确性;其次,不同运动方位的感知不对称,与横向运动相比,深度运动引起的感知更明显,同时,深度方向感知错误时引起的感知偏差往往更大。本研究不仅为对比 3D 运动感知模型和被试主观感知提供了良好的数据,而且为后续设计 3D 运动感知模型提供了一种新的判断模型感知效果的主观对比标准,并使原有立体感知能力的指标更为细化。

2 感知模型设计

依据人对外部世界的感知过程将运动感知分为 3 个部分:时间感知(我们每时每刻都在观看世界)、空间感知(双眼提供了同时刻空间上不同的视觉投射)、时空整合(大脑将获取的丰富信息整合,提供世界影像)。Peng 等设计了双光流卷积神经网络模型,实现对人体动作的识别^[17]。本文结合人类双目融合中的部分神经机制提出了 3D 视频运动感知模型框架,如图 1 所示。

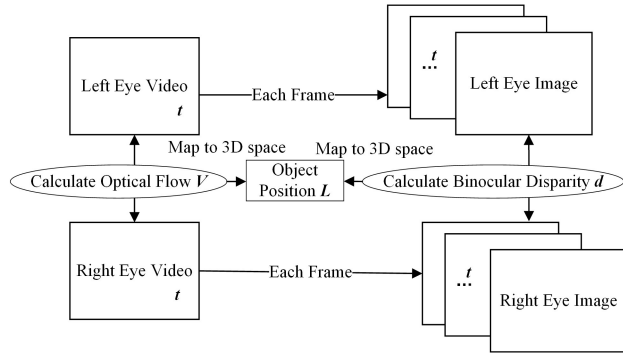


图 1 3D 视频运动感知模型框架

Fig. 1 3D video motion perception model framework

2.1 时间感知模块

时间感知模块考虑了同一空间中对连续视觉信息采集的过程,该模块选择光流法估计 3D 视频中的物体运动。光流法是一种能将物体运动的矢量场投影到摄像机上的三维速度所产生的明显的二维运动的图像流量化的方法,能够提取动态单目场景中可用的信息。Chessa 等^[18]结合生物学特性,对调谐 MT 神经元的速度空间进行采样,并从神经元活动中解码局部速度,设计了能有效处理自然图像场景的模型,光流 V 的计算如式(1)所示:

$$V(u, v, t, \varphi) = \sum_{i=1}^M v_i E^{MT}(u, v, t, \varphi, v_i) \quad (1)$$

其中, $v_i, i=1, 2, \dots, M$ 表示 V1 层的 M 个调谐速度, E^{MT} 表示在 MT 层在方向 φ 的单元响应, u, v, t 分别表示图像中点的横、纵坐标以及时间。结合视觉皮层典型参与运动分析的神经机制,能够获取良好的拟合人观测特征的模型。

此部分对应图 1 左侧的光流估计部分,通过光流估计,分别计算在立体视频时长范围内左右眼视频中运动物体的光流变化,捕捉在二维平面上运动的映射变化,获取对运动物体位置和方向的初步判断。

2.2 空间感知模块

空间感知模块在同一时刻对空间中视觉信息进行采集。

在人的视觉处理系统中,双目即是空间中的信息采集器。Chessa 等^[19]受生物学启发提出了一种求解双目视差的模型,该模型利用灵长类视觉系统的皮层范式估计视差,从神经建模的角度实现了皮质结构与视觉运动通路共享,还模拟了 V1-MT 视觉通路的视差敏感性。其中方向 φ 的视差 d^φ 的计算式如式(2)所示:

$$d^\varphi(u, v, \varphi) = \sum_{i=1}^N d_i E^{MT}(u, v, \varphi, d_i) \quad (2)$$

其中, $d_i, i=1, 2, \dots, N$ 表示 N 个双目视差相位视差。

此部分对应图 1 右侧的双目视差部分,通过计算视频各帧左右图像的视差图,确定各帧对应的时刻运动物体在双目不同二维平面中的差异,量化观看视频的整个过程中双目观看物体运动在视网膜上产生偏移的变化情况。

2.3 时空整合模块

在观看 3D 视频的过程中,时间感知模块获取双眼分别捕获的立体视频在视频时长内物体运动的二维矢量场;空间感知模块获取每帧左右眼图像的视差,模拟人双目匹配图像的过程。运动物体在真实三维世界 $S(x, y, z)$ 同投射在视网膜 $I(u, v)$ 上的影像的对应关系是一个三维坐标转化为二维坐标的过程。我们借助坐标变换将视觉观看中获取的多方信息结合起来,总结出立体运动位置 L 的计算公式,如式(3)所示:

$$L(x, y, z, t, \varphi) = C(u, v, d^\varphi(u, v, \varphi), v(u, v, t)) \quad (3)$$

其中, $C(\cdot)$ 表示由坐标系 I 转换为 S 的函数。

此部分对应图 1 的中间部分,在三维空间计算运动物体的位置。已知首尾帧之间左右眼分别观看运动物体的光流变化,以及首尾帧时刻双目观看物体的视差,通过射影几何变化定律,我们可以在三维空间坐标系上确定物体的坐标、位置和运动方向。

为了判断模型与人类视觉系统感知情况的拟合程度,本研究设计了主观实验,为后续的对工作打下基础。对立体深度运动的感知,物体的运动速度、运动方向、运动前后相对位置等参数都是重要的影响因素。同时在大脑的反应机制中也有对这些参数敏感的细胞响应。当目标深度位置发生偏移时,人对刺激目标的敏感度下降,例如远离屏幕的双眼视差等因素可能会影响感觉的不确定性;而当目标水平位置发生偏移时,加大的双眼视差会影响视网膜图像匹配的不确定性。故实验将物体的运动限制在水平面内,通过改变目标速度和参考物体的旋转半径,改变目标同参考物体的相对位置关系,探究不同运动方向的感知特征,以寻找评判模型设计效果的主观对比标准。

3 主观感知实验

3.1 准备工作

3.1.1 实验人员

选择用《立体视觉检查图》和《标准对数视力表》对被试的立体视力和双目视力进行检查。共有 31 名被试参加了本次实验,除 1 名被试立体视异常,其余被试均立体视正常且双目矫正正常,被试中 14 人为男性。参加实验的被试平均年龄为 23.6 岁,且均未参与过类似的实验。

3.1.2 实验设备与环境

实验使用了极米 Z8X 3D 投影仪作为显示设备,被试佩戴配套 3D 眼镜。实验使用了 MATLAB 和心理物理学工具箱 (Psychtoolbox) 编写的自定义代码控制,在 Lenovo 拯救者 Y7000 上运行代码,图形处理器型号为 GeForce GTX 1050 Ti。利用 HDMI 线将电脑与投影仪相连接,投射电脑屏幕于实验环境中的白墙上(投影仪距离投影墙面 1.5 m,投影仪投射比为 1.2:1)。投影仪提供的屏幕分辨率为 1920×1080 像素(每只眼睛 960×1080 像素),刷新率为 60.027Hz。实验场景如图 2 所示。

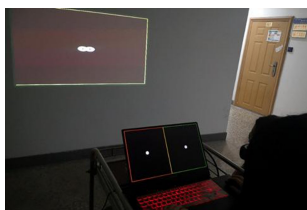


图 2 实验场景

Fig. 2 Experimental scene

3.2 立体视觉刺激

为了记录被试对立体视频中运动的感知情况,设计可量化保存被试反应的立体视觉刺激视频。在所有实验中,被试距离投影墙面 2m。由于人眼观看立体视频存在辐辏冲突,当物体出屏运动时,视觉不舒适感增强,因此将视觉刺激的呈现范围设定在屏幕内部,呈现范围为 $(2.60\text{m} \times 0.70\text{m} \times 2.60\text{m})$ 。受文献[20]的启发,考虑了单目(大小、隐现)和双目线索(视差),设计刺激在透视投影的情况下呈现,刺激视频俯视示意图如图 3 所示。

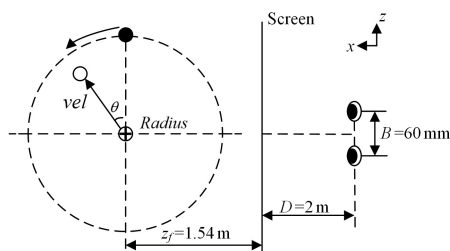


图 3 立体运动视觉刺激示意图

Fig. 3 Schematic diagram of stereoscopic motion visual stimulation

刺激开始时,将距离屏幕 1.54 m 处显示的十字(图 3 中黑色十字)作为初始标识,要求被试观看标识,明确目标初始位置后手动触发目标球体(图 3 空心黑边圆,直径 4.55cm)开始运动。目标初始位置与十字标识位置重叠,出现在视觉场景的中心,选择随机方向 $\theta(0^\circ \leq \theta \leq 360^\circ)$ 出射,匀速移动 6s 后消失。目标球体在 y (垂直)方向上不变,运动轨迹位于 x 方向(水平方向)和 z 方向(深度方向)所在的平面,以十字标识为中心,跨越 360° 随机选择出射方向匀速运动,方向示意如图 3 所示。

水平视差可以分解为深度运动和横向运动两部分,其计算式如式(4)和式(5)所示:

$$d_z = \frac{z_f + z_s}{z_f + z_s + D} \cdot \frac{B}{2} \quad (4)$$

$$d_x = \frac{D + x_s}{z_f + z_s + D} \quad (5)$$

其中, B, D, z_f 分别表示双目间距、眼睛视觉中心与屏幕间距离、十字标识初始深度, x_s, z_s, d_z, d_x 表示目标球体与屏幕中心在 x 和 z 方向上的相对距离以及深度运动、横向运动引起的视差。根据深度运动视差和水平运动视差可以计算得到双眼的水平视差。

3.3 主观实验任务

本实验旨在探究被试立体深度运动感知的典型特征。考虑到 Fulvio 等发现头部抖动有利于人对立体运动的观测,因此未对被试头部进行固定^[20]。被试佩戴 3D 眼镜,在与投影屏幕保持固定距离情况下,观看 3D 投影仪播放的立体视觉刺激视频。在观看过程中,通过被试移动外围固定轨迹的参考球体对目标球体进行拦截,以量化被试对目标球体方向的判断。在目标球体消失后,出现一个用于拦截的红色参考球体(图 3 中黑色实心球体,该球体直径为 5.85 cm),以下简称红球。红球的运动轨迹不可见,但是轨迹固定(图 3 虚线),它以十字标识为中心,在水平面内做圆周运动。被试利用键盘的左右箭头移动红球,如果目标球体沿着既定方向继续前进,就会到达红球拦截的位置。被试可以按照自己对于目标球体的消失位置的感知,对红球进行调整以拦截目标。该过程可以缓慢进行,以确保被试对反应位置判断的准确性。在红球调整期间被试不需要注视,直到对红球的位置设置满意时,按下空格键确定红球拦截位置。接着启动响应反馈——目标重新出现在其最后可见的位置,沿着原定轨迹继续前进。如果目标击中了红球,响应反馈会在屏幕上给出“successful!”的提示;否则,会给出“Miss the sphere!”的提示。至此,一次测试完成,进入下一次测试,被试继续注视十字标识,按任意按键启动下一次实验。

为了探究立体刺激中目标球体的速度以及参考红球旋转半径对被试感知立体深度运动的影响,设置了两组实验。在实验 1 中,我们测试了目标球体的不同运动速度(记为 vel)对感知立体深度运动的影响。为确保目标球体不会运动超出边界,并考虑到屏幕本身的大小,设置了 3 组不同的目标球体运动速度,运动相同时间后目标消失,由被试感知其消失的位置,操纵红球对目标进行拦截。在实验 2 中,我们研究了拦截目标的参考红球旋转半径(记为 $Radius$)对感知立体深度运动的影响。考虑到屏幕的大小,我们均匀选择了 3 组不同的红球旋转半径,保持目标运动速度不变,由被试感知红球与目标的相对位置,预估红球的拦截位置。总共设置 5 组实验,一组为对照组实验 ($vel=0.099\text{m/s}, Radius=1.30\text{m}$, 记为实验 CG); 两组以目标球体运动速度为变量的实验组实验 ($vel=0.132\text{m/s}, Radius=1.30\text{m}$, 记为实验 MidV), ($vel=0.165\text{m/s}, Radius=1.30\text{m}$, 记为实验 TopV); 两组以红球运动半径为变量的实验组实验, ($vel=0.099\text{m/s}, Radius=1.04\text{m}$, 记为实验 MidR), ($vel=0.099\text{m/s}, Radius=0.78\text{m}$, 记为实验 SmaR)。

被试在实验之前进行了 5~10 次练习实验以熟悉这项任务。熟悉任务后,所有被试以随机顺序分别完成 5 项实验,其中每项实验包含 15 次测试。被试在实验中一共完成 75 次

测试,为了防止长期佩戴 3D 眼镜带来的不适感,中间有短暂休息。

4 主观实验结果与分析

4.1 实验条件探究

在实际测试过程中,我们对实验条件的选择进行如下探究:经分析,本次主观实验任务中有两大起决定性作用的实验条件,分别是判定拦截成功阈值和参考红球调整步长。判定拦截成功阈值定义为:在被试做出判断后,目标球体按原定方向运动至红球附近时,在红球周围多大范围内判定为拦截成功(即感知正确)。参考红球调整步长定义为:被试每按动一次键盘左键(或右键)控制红球移动的范围。二者均以度为单位。

共设置两组实验条件:第一组测试了 7 人,其中判决拦截成功阈值设置为 2.29° ,该角度对于拦截精度要求较高;红球调整步长设置为较粗略的 0.05° ,该组实验记为 FE。考虑实际中人类肉眼可以分辨的最小角度,以及红球调整步长与人感知长度的差异,第二组条件选择判决拦截成功阈值为 8° ,红球的调整步长为 0.01° ,该组实验记为 SE,由剩余被试进行测试。

分别计算两组实验的平均拦截成功次数及标准差,结果如图 4 所示。由图 4 可知:SE 实验平均拦截成功的次数远多于 FE,明显改进了感知正确性。由此可以明确,人对于运动物体的感知是相对粗略的,当选择过于精确的判决成功阈值时,感知效果反而不佳;反之,选择接近人肉眼分辨率的阈值,可极大地改进感知效果。此外,参考红球的移动步长更精细,也可以起到提升感知效果的作用。另外,两组实验的标准差达到 1~2 次左右,误差较大,这可能是实验数据较少、被试的学习能力不同以及对实验的熟悉程度参差不齐等因素导致的。

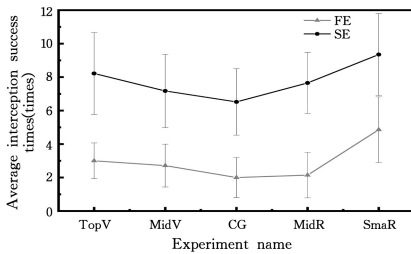


图 4 不同实验条件感知效果对比

Fig. 4 Comparison of perception effects under different experimental conditions

由于 SE 实验表现出良好的感知效果,因此后续分析采用实验 SE 中的实验条件以及相应数据(剔除了立体视异常数据)。

4.2 详细实验分析

4.2.1 目标与参考相对距离的影响

由图 4 可知,纵观这 5 次实验,实验 SmaR 和实验 TopV 被试正确感知判断的次数最多,实验 CG 的正确感知次数最少。根据实验 1 可知,目标球体的运动速度越快,实验的平均拦截成功次数越多;根据实验 2 可知,红球旋转半径越小,实验的平均拦截成功次数越多。

立体视觉舒适度随着运动速度的加快而降低,误差随着运动速度的加快而变大^[21]。立体舒适度与被试的正确感知息息相关,而实验 1 中体现的却是运动速度越快,感知正确而成功拦截的平均次数越多。仔细分析可知,目标球体运动速度改变,运动时间恒定,因此在此期间目标球体的运动距离不同,运动速度越快,运动距离越远。而红球的旋转半径固定时,当运动距离越大时,二者的相对距离越近,做出正确判断的可能性越大。在实验 2 中,红球的旋转半径变小,成功拦截的平均次数变多,也印证了这一点,即红球与目标球体消失位置的相对距离越小时成功拦截的可能性越大。

我们探究了目标消失位置与红球相对距离与被试感知结果之间的关系,结果如图 5 所示。其中横坐标表示各组实验,左侧纵坐标表示目标消失位置与参考物体的平均相对距离,用不同颜色的柱状图来区分成功和失败的情况,右侧的纵坐标则表示各组实验中感知成功的次数,用折线图表示。单独观察每组实验,发现拦截成功(感知结果正确)的相对距离要小于拦截失败的相对距离;对比各组实验,发现相对距离越大,拦截成功的次数越少。综上可以判断:目标与参考之间的相对位置关系极大地影响着被试对运动物体的感知正确性,相对位置距离越小,感知正确率越高。

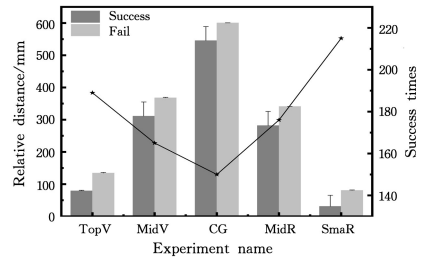


图 5 目标与参考物体相对距离的影响

Fig. 5 Effect of relative distance between target and reference

4.2.2 被试感知情况分析

我们采用散点图来分析感知方向和实际方向的关系,如图 6 所示。其中每个数据点代表某位被试的一次实验,数据点的横纵坐标分别表示该次实验中运动物体的实际运动方向和被试感知到物体的运动方向,用虚线大致区分为向左(left)、向右(right)、远离(away)、靠近(near)4 个方向,图中一共描绘了 2325 个数据点。

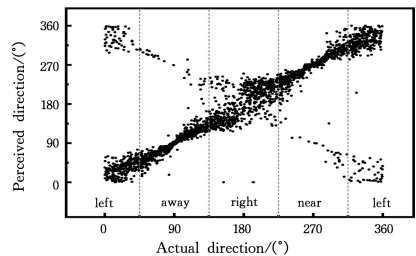


图 6 感知方向同实际方向的关系

Fig. 6 Relationship between perceived direction and actual direction

仔细观察散点图可以发现,大多数数据点沿着正对角线分布,表示目标运动的横向(x)和深度(z)分量的准确报告。由图 6 还可以明显看出,在深度运动分量更大的 90° 和 270° 附近,被试感知方向更加贴近实际方向,而在横向运动分量更大

的 0° 和 180° 附近,被试感知方向的偏差更大。除此之外,还有少部分数据点分布在负对角线上,表示目标运动深度(z)的不准确报告,此类数据点占总数的8.13%。Fulvio等的实验中对被试表现出的这种混淆深度运动方向的现象进行了详细的分析,并得出结论:在3D运动中都存在这种混淆情况,在深度方向上尤其明显,且这种感知错误是由于受到了感官不确定性的限制,这种感官不确定性来自各种感官噪声源^[22]。

针对两种行为测量方法总结了被试在任务中的表现。第一个行为测量方法是计算目标拦截的百分比,即捕获在整个实验中估计和报告目标的3D运动方向的总体准确性。如果目标击中了参考红球周围 8° 范围内的任何位置,则被试的此次感知判断被确定为一次成功拦截,从而产生2.2%的机会拦截率。我们做了如下分析:首先计算每个被试在不同实验条件下成功拦截目标的百分比,成功拦截目标意味着对目标的深度运动和横向运动的准确估计;其次分析了不同运动方向下目标拦截的百分比情况。

为了细化数据,以 15° 为单位将 360° 分为24份,计算各个方向下的目标拦截成功率,实验结果如图7所示, a 与 b 表示实验1和实验2中目标拦截成功百分比的情况。根据两组图像的趋势可以看出,当目标直接接近或离开被试时(即运动方向为 90° 和 270° 附近时),拦截成功的百分比特别高。

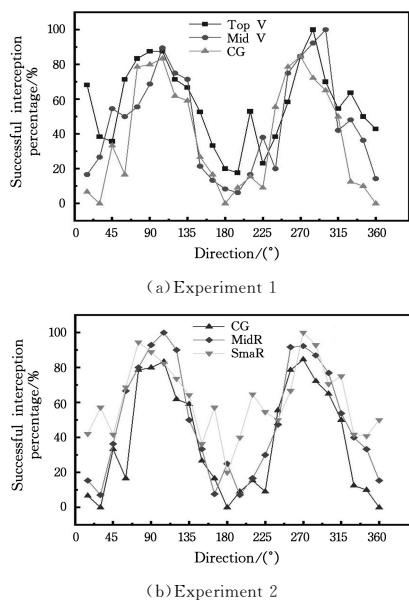


图7 两组实验成功拦截百分比

Fig. 7 Percentage of successful interception in two groups of experiments

以 90° 为单位比较横向和深度方向的拦截成功百分率,结果如表1所列。这些结果表明,与目标横向移动时相比,当目标直接接近或离开被试时更有可能被成功拦截。因此被试在不同目标方向上的感知能力是不对称的,深度运动相比横向运动引起的感知更为明显,其正确拦截率比横向运动的正确拦截率高约42.67%~47.01%。考虑到大脑皮层中神经元对运动的方向选择性,这种明显的感知能力差异可能是由于深度运动方向更贴合大脑皮层感知的最优方向,刺激了细胞响应的缘故。结合这一发现,我们将人对运动感知方向不对称的特性作为模型拟合人的感知程度的主观标准之一。

表1 各组实验在前后左右4个方向的拦截成功率

Table 1 Interception success rate of each group of experiments in four directions

Exp	0° / %	90° / %	180° / %	270° / %	Var
TopV	51.08	77.50	35.29	68.49	0.0351
MidV	38.30	69.00	19.05	68.43	0.0559
CG	10.67	64.81	13.89	68.13	0.0653
MidR	24.36	82.56	19.32	74.19	0.1081
SmaR	12.90	74.19	45.45	74.42	0.0855
Ave	27.46	73.61	26.60	70.13	—
Var	0.0294	0.0049	0.0176	0.0016	—

为了量化被试感知在不同方向的偏差,第二个行为测量方法选择计算均方根误差。分别在 $0^\circ, 90^\circ, 180^\circ, 270^\circ$ 前后 45° 计算运动的均方根误差,计算结果如图8所示。横向运动体现出较小波动($var=0.0127, \theta=0^\circ$; $var=0.0009, \theta=180^\circ$),而深度方向运动体现出了较大的波动($var=0.1062, \theta=90^\circ$; $var=0.1023, \theta=360^\circ$)。经过详细分析可以推断出,目标在深度方向运动时,感知准确而成功拦截的可能性更大,但一旦感知错误,引起的偏差往往更大,均方根误差超出约0.1583~0.3665。纵观两组实验,实验MidR和实验SmaR在各个方向的整体均方根误差变化不大,实验TopV的整体均方根误差较小,实验MidV和实验CG在各个方向的整体均方根误差变化剧烈。

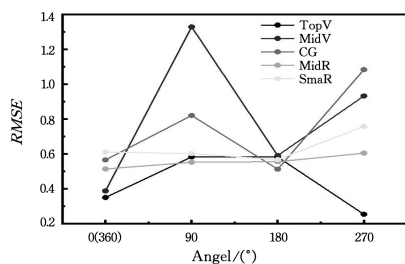


图8 各个方向均方根误差

Fig. 8 Root mean square error in each direction

结束语 目前的研究探索了被试运动感知的显著特征,提供了对3D环境中运动感知过程的见解。特别地,我们的研究结果揭示了被试在3D视频观看的过程中对运动物体不同运动方位具有感知差异,这种感知差异与大脑皮层的方向选择特性是相互契合的。这一发现为设计3D运动感知模型提供了新思路:结合人眼视觉特性和大脑皮层中特定细胞的方向选择性等特征是有必要的,这能够更加贴近人眼真实感知情况,因此我们提出了模型的初步框架,并将运动感知的不对称性作为判断模型效果的主观标准之一。未来,我们将继续深入扩展主观实验的研究角度,提出更全面的描述人立体感知能力的指标。

参考文献

- [1] WHEATSTONE C. Contributions to the physiology of vision [J]. Royal Society, 1838, 28: 371-394.
- [2] HMUEL A, GRINVALD A. Functional organization for direction of motion and its relationship to orientation maps in cat area 18 [J]. Journal of Neuroscience, 1996, 16(21): 6945.
- [3] POGGIO G F, MOTTER B C, SQUATRITO S, et al. Responses of neurons in visual cortex (V1 and V2) of the alert macaque to

- dynamic random-dot stereograms[J]. *Vision Research*, 1985, 25(3):397-406.
- [4] TANAKA K, SAITO H. Analysis of motion of the visual field by direction, expansion/contraction, and rotation cells clustered in the dorsal part of the medial superior temporal area of the macaque monkey[J]. *Journal of Neurophysiology*, 1989, 62(3):626-641.
- [5] XU H, WALLISCH P, BRADLEY D C. Spiral motion selective neurons in area MSTd contribute to judgments of heading[J]. *Journal of Neurophysiology*, 2014, 111(11):2332-2342.
- [6] HUBEL D H, WIESEL T N. Receptive fields of single neurones in the cat's striate cortex[J]. *PhysiolLund*, 1959, 148:547-591.
- [7] HYSON M T, JULESZ B, FENDER D H. Eye movements and neural remapping during fusion of misaligned random-dot stereograms[J]. *Journal of the Optical Society of America*, 1983, 73(12):1665-1673.
- [8] DING J, LEVI D M. A unified model for binocular fusion and depth perception[J]. *Vision Research*, 2021, 180(2):11-36.
- [9] XIE M, NIEHORSTER D C, LAPPE M, et al. Roles of visual and non-visual information in the perception of scene-relative object motion during walking [J]. *Journal of Vision*, 2020, 20(10):15.
- [10] JULESZ B. Binocular depth perception of computer-generated patterns[J]. *Bell System Tech*, 1960, 39:1125-1162.
- [11] CORMACK LK, CZUBA TB, KNÖLL J, et al. Binocular Mechanisms of 3D Motion Processing[J]. *Annual Review of Vision Science*, 2017, 3:297-318.
- [12] KIM T, RYU K, SONG K, et al. Loop-Net: Joint Unsupervised Disparity and Optical Flow Estimation of Stereo Videos With Spatiotemporal Loop Consistency[J]. *IEEE Robotics and Automation Letters*, 2022, 5(4):5597-5604.
- [13] MALTZ M V, STUBBS K M, QUINLAN D J, et al. Familiar size affects the perceived size and distance of real objects even with binocular vision[J]. *Journal of Vision*, 2021, 21(10):21.
- [14] THOMPSON L, JI M, ROKERS B, et al. Contributions of binocular and monocular cues to motion-in-depth perception [J]. *Journal of Vision*, 2019, 19(3):2.
- [15] WHRITNER J A, CZUBA T B, CORMACK L K, et al. Spatio-temporal integration of isolated binocular three-dimensional motion cues[J]. *Journal of Vision*, 2021, 21(10):2.
- [16] PETER S, ANDREW G. Combining cues to judge distance and direction in an immersive virtual reality environment[J]. *Journal of Vision*, 2021, 21(4):10.
- [17] PENG Y, LEE H, SHU T, et al. Exploring biological motion perception in two-stream convolutional neural networks[J]. *Vision Research*, 2021, 178:28-40.
- [18] CHESSA M, SABATINI S P, SOLARI F. A systematic analysis of a V1-MT neural model for motion estimation[J]. *Neurocomputing*, 2016, 173:1811-1823.
- [19] CHESSA M, SOLARI F. A computational model for the neural representation and estimation of the binocular vector disparity from convergent stereo image pairs[J]. *International Journal of Neural Systems*, 2019, 29(5):1-20.
- [20] FULVIO J M, ROSEN M L, ROKERS B. Sensory uncertainty leads to systematic misperception of the direction of motion in depth[J]. *Attention Perception & Psychophysics*, 2015, 77(5):1685-1696.
- [21] WAN W F. Image Quality evaluation for human visual perception characteristics[D]. Xi'an: Xidian University, 2020.
- [22] FULVIO J M, MIAO H, ROKERS B. Head jitter enhances three-dimensional motion perception [J]. *Journal of Vision*, 2021, 21(3):12.



LU Ping, born in 1999, postgraduate. Her main research interests include 3D stereo video motion perception and so on.



ZHANG Di, born in 1988, Ph.D. Her main research interests include 3D stereo vision and virtual reality technology.

(责任编辑:何杨)