运动感知计算模型的研究与应用

Motion Perception Model and Its Application

院(系)： 生命科学技术学院

专业： 生物医学工程

学号： 5060809026

姓名： 周博磊

指导教师： 张丽清

运动感知计算模型的研究与应用

摘要

视觉运动感知与分析是计算神经科学与计算机视觉交叉研究领域的基本问题之一，其主要任务是以神经生理学和认知科学的研究成果为基础，通过理论分析和计算机仿真，模拟人类视觉信息处理的计算机制，构造出新的视觉感知计算模型和人工视觉信息处理系统。视觉信息处理机制和计算原理的研究不仅对揭示大脑神经系统的理论原理，建立新型计算模型具有重要的意义，而且对推动信息技术的发展，如人工视觉系统、失明患者的视觉功能修复、机器认知、新型脑机接口系统等也具有积极的作用。另外，具有仿脑计算的运动检测算法在模式识别、身份验证、安全监控、智能交通系统等领域有着广泛的应用前景。

运动感知是视觉感知的重要组成部分，负责主动感知和获取环境信息，并将信息有效地传递给高级认知中枢，以便其作出迅速响应和反馈。经典视觉理论将大脑视觉通路分为what和where通路，what通路负责传输视觉刺激的内容，而where通路则传递和表征物体位置和运动信息。这种what和where通路的设定，使得视觉刺激得到了有效表征，其where通路对物体运动的表征与计算的机制，正是本文的研究重点。在人的日常生活中，人对运动物体的敏感程度远大于静止物体，运动感知无处不在。正因为运动感知对于人类的重要性，如何提取出运动感知的计算机理，并将其实现的算法应用到机器视觉系统中，是计算机视觉研究领域的一个关键问题，也是尚未解决的一个问题。

本文首先对人的运动感知行为进行心理物理学实验，对人在运动感知过程中的差异性和一致性进行数值量化分析，随后以受试者的标定数据作为为基准（ground-truth）建立了一个测试运动感知和检测算法的公共测试数据库；然后，结合心理物理学实验的结果，我们推导和建立了一个新的运动感知计算模型，并对计算模型的算法进行了深入的理论分析；最后，我们在先前建立的公共数据库上测试了该运动感知算法的性能和效果，并与当前流行的运动检测模型和算法进行了比较实验，实验结果证明了该运动感知模型的优越性能，及其所具有的广泛工程应用潜力。总的来说，本文的主要贡献和创新点体现在以下几个方面：

首先，我们采集和建立了一套新的运动感知测试数据库。这套数据库由20个不同自然场景中的视频片断组成，视频中的运动物体种类多样，共有11个受试者独立对其中的运动物体进行标定。随后，我们对11个受试者的标定数据进行了数值统计分析，计算出了人在自然场景中运动感知所存在的个体差异性及其一致性的大小。更进一步，我们以11个受试者的标定数据作为基准（benchmark)，结合改良后的评估准侧,建立了一套新的用于对运动检测机器算法进行测试和比较的公共平台。相对于现有的运动检测算法比较的数据库，我们的运动感知测试平台有如下优势和特点：第一，数据库中的视频片段均是自然场景，自然场景的种类多样，并且，多种运动物体的种类也多样化，包含开动的汽车、自行车、行人以及运动员等，运动检测算法的泛化能力（generalization）和鲁棒性（robustness）可以得到很好的测试；其二，由于数据库中的视频都是由移动中的摄像机拍摄，这去掉了”摄像机固定”的假设条件，使得

开发的新算法能够在更广泛的范围下正常工作；第三，我们的数据库中的视频由多人标定，我们对标定数据的稳定性进行了测试，并将标定数据的个体差异性和一致性纳入到设计测试准则之中，使得在这个数据库上测试出的结果能更加客观和准确。我们相信，这个公开的运动检测算法测试平台将对相关算法的研究和开发有推动性作用。

随后，结合建立运动感知数据库的实验过程中所得到的定性结果，我们通过两套不同的理论推导和分析方法，建立了新型的运动感知计算模型。这个运动感知计算模型的核心是利用傅里叶空间的相位差成分来提取和表征自然场景中物体的运动成分。这种表征可以有效地绕开运动物体本身的外形复杂度而专一性地提取物体的运动信息，与人类两条视觉通络中"what"和"where"之间的分离的机制有着概念上的联系。物体运动信息在我们的运动感知计算模型中能简明地表征出来，不仅使得模型本身的复杂度很低，也减少了模型中暗含的假设，使得模型更具鲁棒性，也避免了over-fitting的问题。最后，这个新型的运动感知模型的算法由9行MATLAB代码简明地实现出来。随后，我们通过大量实验对这个运动感知模型进行了测试和比较。在与同类运动检测算法的比较实验中，我们的运动感知模型算法取得了较大的优势。一系列的实验说明，该运动感知算法的良好的稳定性、鲁棒性，以及极快的计算速度（普通笔记本上达到70fps），使得它具有了广泛的工程应用前景。

在模型的应用中，我们将运动感知计算模型的算法移植到了ARM9嵌入式平台上进行了实现，算法在嵌入式Linux系统中运行良好，所需硬件环境极低，表现出了在工程系统中的应用潜力。随后，为了更进一步说明运动感知计算模型对大脑神经生理计算机制的契合，我们利用运动感知模型拟合了视觉注意力和眼动的生理学实验数据，在与同类的神经生理学计算模型的评测中，我们的运动感知计算模型的拟合结果最准确，这说明了运动感知模型不仅可以应用到工程问题之中，还能对神经生理学现象进行拟合和预测，这对揭示人类视觉系统计算机制有着积极的意义。

关键词： 运动感知，运动检测，运动表征，显著性图

Motion Perception Model and Its Application

ABSTRACT

Motion perception and analysis remains one of the fundamental issues in the interdisciplinary research of neural computation and computer vision. Its purpose is to build novel computational models for motion processing based on experimental results of neurophysiology and psychology, with solid theoretic analysis and computer stimulation. The study on visual signal processing and motion analysis not only helps to reveal the principles of brain networking and computing, but also contributes greatly to the development of information engineering and technology, such as artificial visual system, visual rehabilitation of the blind, and brain-computer interface system *etc*. Furthermore, brain-like motion detection algorithms also have wide application potential in pattern recognition, identity verification and intelligent transportation systems.

Generally speaking, motion perception and analysis is one of the most important modules in human vision system. It actively perceives and captures the information in the external environments, represents those information in an efficient way, and finally generalizes them to the higher visual cortexes which account for decision making. The classical vision science dichotomizes the human visual pathways into TWO parts, the ventral stream which is called as *what pathway* and the dorsal stream which is called as *where pathway*. The former processes and represents the appearance of the objects and the latter represents the spatial position and motion. This dichotomy of visual pathways help input channels efficiently represent all-round properties of attentative objects. In our research, we concentrate on exploring what the computational mechanism of *where pathway* is. Furthermore, our visual system is more sensitive to moving objects than still objects, motion perception is an important task in our daily lives. Because of the importance of motion analysis towards human, how to extract the computational mechanism of motion perception and then to implement it into the algorithms on artificial visual system become a meaningful challenge to be solved.

This thesis first introduces the psychophysical experiments on human motion perception in natural scenes, in which the quantitative numerical analysis is performed to test the consistency of the subjects' labeling data. After that, the labeling data from different subjects are used to form the ground truth of a new benchmark for performance evaluation of motion detection algorithms. Then, along with the result of former psychophysical experiments on human motion perception, a novel computational model for motion perception is constructed. Furthermore, this model presents a surprisingly simple algorithm to detect moving objects in dynamic scenes. After evaluating it on the proposed benchmark, our model shows priority over other state-of-art motion detection algorithms compared, the great efficiency of our computational model for motion perception also reveals great potentials in wide engineering applications. Main contributions of this thesis are listed as follows.

First of all, we collect and build a new database for the evaluation of motion detection algorithms, it would be available to public in future. This database contains 20 video clips collected in natural scenes with hand-hold camera, the moving objects in those clips are various, such as auto, bicycle, pedestrian and sport players. 11 subjects are instructed to annotate the moving objects in these clips, respectively. After that, numerical analysis is performed on those labeling data to test the database consistency, the result proves the fact of inter-subject difference on interpreting object movement. On the other hand, deriving from this fact, a generalized evaluation metric is normalized with these labeling data to form a compact benchmark for motion detection algorithm evaluation. Our motion perception database holds following properties compared with other similar test databases: 1) The total of videos are clipped in natural scenes, the class of scenes and category of objects are various, such as the auto and bicycle on the road, the pedestrian on the street and the sport players on the ground. This variation of targets could solidly measure algorithm's generalization ability and robustness; 2)since all of the videos are clipped by hand-hold camera, the assumption of static camera which is common among other databases is released, so that the newly developed algorithms should be encouraged to work under dynamic scenes. 3)The video clips in database is labeled by 11 subjects, the statistical consistency and inconsistency of subject on interpreting motion are taken into consideration when setting the evaluation methodology of database, which would make the evaluation results more accurate and objective. We believe, this database would make real contribution to the research of motion detection

Moreover, from two different views of theoretic analysis our novel computational model of motion perception is well justified. The fundament of our theory is that the Fourier phase discrepancy part represents solely the frame-to-frame movement of objects in spite of the spatial complexity of object appearance. Generally this fact could be understood in the context of the dichotomy of *what pathway* and *where pathway* in human visual system, in which the *where pathway* encodes the motion information efficiently. While motion could be represented efficiently by phase discrepancy part, the model complexity would be much reduced, which make our model avoid over-fitting. Finally, our motion perception model could be implemented as computer algorithm in 9 lines of MATLAB code. In the following experiment part, this algorithm is evaluated in our proposed database along with other state-of-art motion detection algorithms. Our model achieves the greatest performance. The efficiency, robustness and ultra speed of our algorithm make it hold a wide engineering application potential.

At last, our motion perception algorithm is implemented on ARM9 embedded system for engineering test. It is no doubt that the high-efficiency in computing and low-requirement in hardware resource make this algorithm demanding in engineering application. Furthermore, we apply this model to fit and predict the eye saccade data of human being. In comparison to other two computational models of visual perception, our model performs better to fit the behavior of human eye saccade. This indicates the our model's generalization ability in predicting neuropsychological phenomena. In the future, it is hoped to see more experiments are performed to explore the connection between our computational model and the neurological mechanism of *where pathway* in analyzing object motion.

Key words: motion perception, motion detection, motion representation, saliency map

目 录

第一章　绪论

视觉运动感知与分析是计算神经科学与计算机视觉交叉研究领域的基本问题之一，其主要任务是以神经生理学和认知科学的研究成果为基础，通过理论分析和计算机仿真，模拟人类视觉信息处理的计算机制，构造出新的视觉感知计算模型和人工视觉信息处理系统。视觉信息处理机制和计算原理的研究不仅对揭示大脑神经系统的理论原理，建立新型计算模型具有重要的意义，而且对推动信息技术的发展，如人工视觉系统、失明患者的视觉功能修复、机器认知、新型脑机接口系统等也具有积极的作用。另外，具有仿脑计算的运动检测算法在模式识别、身份验证、安全监控、智能交通系统等领域有着广泛的应用前景。

运动感知是视觉感知的重要组成部分，负责主动感知和获取环境信息，并将信息有效地传递给高级认知中枢，以便其作出迅速响应和反馈。经典视觉理论将大脑视觉通路分为what和where通路，what通路负责传输视觉刺激的身份，而where通路则传递和表征物体位置和运动信息。这种what和where通路的设定，使得视觉刺激得到了有效表征，其where通路对物体运动的表征与计算的机制，正是本文的研究重点。在人的日常生活中，运动感知无处不在，且人对运动物体的敏感程度远大于静止物体。正因为运动感知对于人类的重要性，如何提取出运动感知的计算机理，并将起实现的算法应用到机器视觉系统中，是计算机视觉研究领域的一个关键问题，也是尚未解决的一个问题。

本文首先从人的运动感知进行心理物理学实验出发，对人在运动感知过程中的差异性和一致性进行数值量化分析，随后以受试者的标定数据作为为基准（ground-truth）建立了一个可以测试运动感知和检测算法的公共数据库；然后，结合心理物理学实验的结果，推导和建立了一个新的运动感知计算模型，并对计算模型的算法进行了深入的理论分析；最后，我们在先前建立的公共数据库上测试了该运动感知算法的性能和效果，并与当前流行的运动检测模型和算法进行了比较实验，实验结果证明了该运动感知模型的优越性，及其所具有的广泛工程应用潜力。

本章绪论首先介绍视觉运动感知与计算模型研究的背景和意义，提其研究目标，并针对该问题综述当前国内外研究成果及研究进展，最后给出全文的组织结构安排。

1.1　研究背景及意义

1.1.1　视觉理论

在认知科学领域，“视知觉从哪里开始”是本源性问题。目前，国际上在视知觉计算研究领域中具有重要地位的理论是以Marr为代表的特征分析理论[1]。该理论认为视知觉是有局部性质到整体性质进行的，首先感知到的是局部几何特征。与之相反，以Gestalt理论为主的拓扑知觉理论[2]则认为视知觉过程是由大范围整体性质到局部性质进行的。首先感知到的是大范围的拓扑特征。本部分简单综述这两种理论的内容。

Marr的视觉计算理论要点是：1）视觉系统是一种复杂的信息加工系统，怎么对信息进行有效表征是核心问题。2）复杂的信息系统需要从三个不同层次上加以描述：计算理论层次，算法层次和硬件层次。3）视觉过程从视网膜上成像开始，经过三个阶段的加工，得到物体的三维结构。4）功能模块是结构上形成的专门快速处理某一视觉特性的模块，是进化和

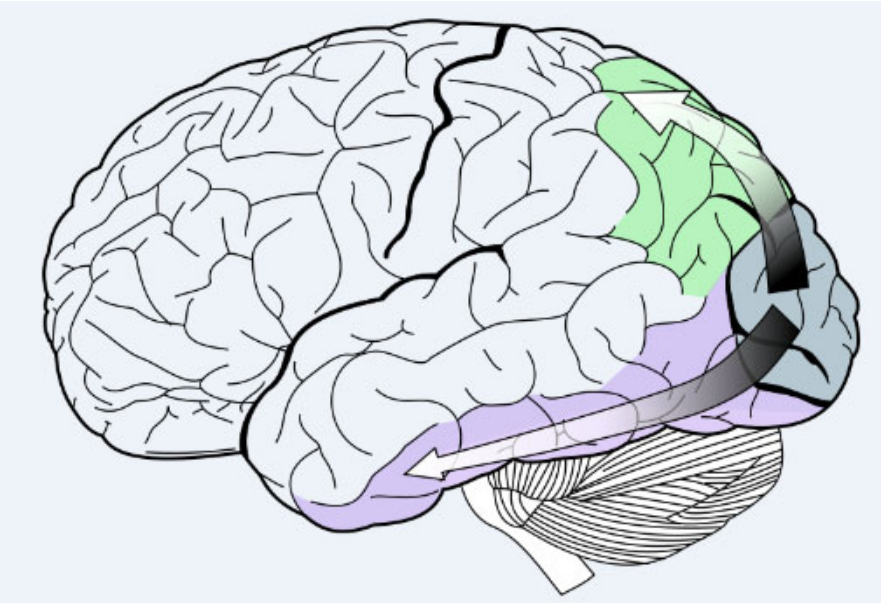


图 1.1 视觉系统where和what两条通路。绿色的通路代表dorsal通路，主要负责物体的位置和运动信息传递。紫色的通路代表ventral通路，主要负责物体的外表特征。

发育过程中形成的功能，是“软件硬化”在神经系统中的结构。早期视觉的功能模块（边框检测、光流检测、体视等）已获充分研究，并在理论上去的重大进展。

另一方面，拓扑知觉理论的观点是，自然界的许多复杂刺激是由相对独立的局部图像组成的整体图像。例如，一幅人脸图像中的各关键部分（如眼睛、鼻子和嘴等）都有独立的局部特征，只有当它们满足一定的拓扑关系时，才能作为一个整体唯一地标示一张人脸。这种重视知觉组织，淡化局部特征，强调对象之间的拓扑关系和整体意义的理论被称为拓扑知觉理论。作为拓扑知觉理论的主流学派，Gestalt学派提出了一些知觉组织的规律，包括：领接律、相似律、封闭律、连续律等。这些规律共同形成视觉计算整合的原则。

这两套里程碑式的视觉计算理论相辅相成，对从计算机制和原理上理解视觉系统的结构和生理学现象起到了巨大的推动作用。我们的这套运动感知计算模型就是基于Marr视觉计算理论的核心假设，即视觉系统如何对物体的运动进行有效表征才是核心研究问题。

1.1.2 视觉系统的**what**和**where**通路

在经典的视知觉的研究中，确定了大脑皮层中加工相同刺激的不同方面是由分离的神经通路，即what和where通路[3]，进行处理的。其中what通路从枕叶中的初级视皮层向颞叶下行（ventral stream），主要负责加工视觉刺激的颜色、形状和特性（即刺激的身份或刺激是什么）。where通路则从枕叶向顶叶上行（dorsal stream），主要负责物体位置和运动信息的加工。这样，为了识别外界环境中的物体和所发生的事件，特征信息至少要输送到两个不同的系统，如图1.1所示。

本文研究的主要内容即是对视觉系统的where通路的运动处理机制进行仿脑研究。我们认为，人的视觉系统在进化过程中将视觉系统对物体外貌特征和位置运动特征分离，是外界信息获取最大化（Informaiton Maximization）原则[4]的一个体现。在信息最大化的准则下，我们可以通过仿脑计算，尝试对where通路的运动处理机制或者人的运动感知机制进行计算建模，并且开发出高效的运动感知机器算法，应用于诸多工程问题之中去。

1.2 国内外研究进展综述

运动感知是视觉感知的重要组成部分，负责主动感知和获取环境信息，并将信息有效地传递给高级认知中枢，以便其作出迅速响应和反馈。在人的日常生活中，人对运动物体的敏感程度远大于静止物体，运动感知无处不在。正因为运动感知对于人类的重要性，如何提取出运动感知的计算机理，并将起实现的算法应用到机器视觉系统中，是计算机视觉研究领域的一个关键问题，也是尚未解决的一个问题。并且，运动感知计算模型与一些重要工程应用息息相关，如运动检测，运动物体捕捉与跟踪，运动物体分割，监控系统等等。

在计算机视觉研究领域中，由于运动检测的工程重要性，许多运动检测模型和算法已经被建立起来。总的来说运动检测的模型和算法可以大致分为四大类：背景建模法（background modeling），视角几何法（view geometry)，物体识别法（detection by recognition），显著性检测法（saliency-based detection）。

1.2.1 背景建模法（background modeling)

背景建模法的核心是对场景的背景成分建立统计模型，在一帧新的图像中任何与背景模型存在显著性差异的像素部分则被当成运动物体。例如, Stauffer和Grimson [5]利用高斯混合模型（mixture of gaussian)对图像像素进行建模。在这个方法中，当前帧的每一个像素将于每一个背景成分的高斯分布进行对比，看是否能找到一个匹配的背景分布成分。如果能找到，则这个像素被分类为背景成分，这个被匹配的背景成分的高斯分布的均值和方差也进行在线更新，如果达不到匹配的阈值，则这一像素点被分类为前景成分。随后的一些方法[6]对高斯分布的假设进行了扩展，利用核函数拟合背景模型得统计分布，提高了背景建模的稳定性。但是背景建模法都依赖于一个核心假设，即处理的视频必须是由固定静止的摄像机拍摄。在摄像机运动的情况下，伴随前景背景的任意视角和场景深度变化，背景建模法处理的效果非常差[7]。

1.2.2 视角几何法（view geometry)

视角几何法是基于对场景摄像机拍摄的几何参数进行估计，然后对摄像机移动进行补偿[8] [9]。在一定的3D几何限制下，摄像机移动的变换矩阵参数可以被估计出来，利用估计出来的变换矩阵，可以补偿摄像机移动导致的运动，从而从运动残差中分离出运动物体[10]。然而，视角几何法的弱点在于变换矩阵的参数估计计算量很大，在自然场景下，由于摄像机的移动无法预测，则每一帧都需要重新进行估计，并且参数估计的过程容易受到场景中的噪声干扰，故本方法很难在工程方面有很大的应用。

1.2.3 物体识别法（detection by recognition）

另外一类很流行的运动物体检测算法是基于物体识别或监督学习（supervised learning）的计算框架。通过对标定好的特定物体在各种视角的图片进行训练，分类器可以在视频中的每一帧直接识别特定的物体，从而对前景背景成分进行分离。例如，一些算法可以对特定类别物体，如人脸[11]或者行人[12]，进行识别。但是，这类算法通常需要离线的训练，且只能处理一小类特定类别的物体，并且，研究具有不变性、克服光照变化、遮挡的物体表征子本身就是计算机视觉领域的一个难题。

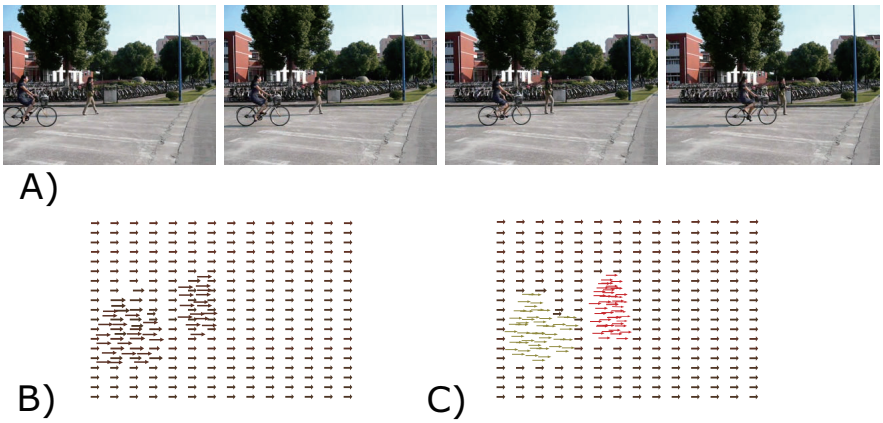


图 1.2 从光流分析的角度看待运动物体检测问题. **A**): 一个具有摄像机自动（ego-motion）和运动物体的视频序列。**B**): 相应的光流示意图。**C**): 通过分类场景中光流的来源种类，可以检测出运动物体的轮廓。

1.2.4 显著性检测法（**saliency-based detection**）

显著性检测法依赖假设，物体具有与背景成分存在显著性统计差异的特征。对于多数显著性检测模型，显著性的差异特征，如颜色、朝向[13]，稀疏基底响应[14]，或者时间特征[15]被用来生成物体显著性图（saliency map），进而判断出运动物体的位置。显著性检测法的优势是它不对被检测的物体的外形进行建模，因此算法可以不需要提前训练而对多类物体进行检测。然而，处于计算效率的考虑，现在多数的显著性检测算法并没有很好整合运动信息，以至于在视觉监控中的实际表现不尽如人意。

实际上，视觉系统**只需要**利用运动信息对运动物体进行检测，即使是在摄像机有自动（ego-motion）的情况下，如图1.2所示。在获得了所有光流信息后，物体检测的任务就变成了从所有光流信息中分离出前景运动导致的光流成分。但是，准确获得场景全部光流信息的计算量过于庞大，使得基于光流的方法很难用于实际工程应用中，

1.3 论文组织结构

本文从视觉计算的角度，先通过心理物理学实验研究人在自然场景下对运动感知的差异性和一致性，并结合标定数据建立了运动检测算法的评测数据库，随后，我们对运动感知计算模型进行理论分析和推导，并将模型实现为可供测试的计算机仿真算法。在此基础上，我们在运动感知数据库上测试了该算法，并与现有的运动检测同类算法进行了对比，实验结果证明，我们的算法在所有参与比较的运动检测算法中取得了最好的成绩，最后，我们的算法在嵌入式系统中的实际测试证明了算法所具有的良好工程应用前景。本论文的具体组织结构如下：

第二章重点介绍了运动感知数据库的标定与数值分析结果。这个数据库的优点和特点主要来源于两个方面：其一，在这个数据库中我们对人在自然场景中的运动感知情况进行定量统计分析。数据库中的视频片段均是自然场景，我们让11个受试者对数据库中的运动物体进行标定，然后统计分析这些受试者对运动感知的情况。其二，我们提供了一套新的运动感知和检测算法的测试平台。在分析完受试者的运动感知情况后，受试者的手工标定数据可以作为基准（ground-truth）对运动检测的机器算法进行测试。由于数据库中包含多种运动物体，

如车辆，行人等，运动检测算法的泛化能力(generalization)可以得到测试。另外，由于数据库中的视频都是由移动中的摄像机拍摄的，去掉了“摄像机固定”的假设条件，使得新开发的算法能具有更广泛的用途。以后这个数据库将在网上公开，为研究同行测试运动检测算法提供了一个新的平台，这无疑将推动运动感知模型和算法的研究。

第三章重点对运动感知计算模型进行了理论推导和分析。我们通过两种不同的理论分析方法对运动感知计算模型进行了推导和解释，最后得到了统一的结论，即：傅里叶空间的相位差成分在运动感知的过程中具有重要理论作用，它有效地表征出了物体的运动。这也正是我们的运动感知计算模型的核心。随后，运动感知模型的算法由9行MATLAB代码简明地实现出来。

第四章主要是实验部分，介绍我们如何在数据库上测试运动感知算法以及分析比较同类运动检测算法。定性实验结果和定量实验结果均证明了我们这个运动感知模型的优越性和其具有的广泛工程应用前景。

在第五章中，我们将运动感知计算模型的算法在ARM9嵌入式平台上进行了实现，算法在嵌入式Linux系统中运行良好，发挥出色，表现出了在工程系统中的应用潜力。然后，我们利用运动感知模型拟合了视觉注意力和眼动的生理学数据，在同类算法中评测结果最好，说明了运动感知模型不仅具有广泛的工程学应用，还能对神经生理学现象进行预测，对揭示人类视觉系统计算机制有着一定意义。

最后一章总结全文的工作及创新点，分析本论文工作的特点与不足，并对将来的工作进行展望。

第二章　运动感知数据库的标定与分析

2.1　引言

运动感知不仅是心理物理学，也是计算机视觉研究领域关注的一个重要问题。心理物理学对运动感知的研究方法主要是用简单的运动刺激，如点、线，或者运动幻觉[3]，记录受试者的眼动，受试者的按键反应或者受试者的EEG脑电信号变化[16]。这种方法论具有两点不足：

1，实验的结果通常只是定性的，而缺少定量化的分析。

2，给予的运动刺激过于简单，而忽略了其他因素对人的运动感知的影响，得出的结论在自然场景中不具有很大普适性。比如，人在自然场景中的运动感知受到诸如场景的上下文[17]，以及运动物体的外表的显著性的影响以及场景中其他物体对视觉注意力的干扰等。

另一方面，计算机视觉领域也在对运动感知的计算机理进行建模研究。因为运动感知和检测的过程对一些关键的计算机视觉工程应用，如运动结构重构（structure from motion)，物体跟踪和识别，视频剪辑等，起着重要的作用。近几年新的运动感知的计算模型和算法不断出现，如何对这些计算模型和算法进行评估和比较，分析出缺陷和不足以便开发出新的计算模型和算法，就成了一个重要的问题。近几年建立了一些评估运动检测和跟踪算法的公共数据库，如PETS [18] 和CAVIAR [19]。有了这些测试平台，研究人员就可以对自己的算法进行测试，以便开发出新的更高效的算法。虽然这些数据库对运动感知和运动检测做出了巨大贡献，但仍存在以下一些不足：

1，数据库中包含的运动物体较单一，如只包含行人或者车辆。这种设定限制了对运动检测算法的普适性方面的测试。

2，数据库中的视频都是通过固定的摄像机拍摄。这样使得多数运动检测算法都含有“摄像机固定”的假设，限制了运动检测算法更广泛的应用。比如车载的视觉系统受到车辆自动，颠簸等影响，要求运动检测的算法在摄像机移动状态下也能正常工作。

基于以上几点，我们建立了这个运动感知数据库。这个数据库的特点有如下两点：

第一，数据库中的视频中的运动物体由11个受试者单独标定，我们对这些标定数据的稳定性的进行了测试，并将这种运动感知的个体存在的差异性和一致性整合到了数据库的评估准则之中，使得数据库测试平台能够对运动检测算法进行更客观和准确的比对。

第二，数据库中包含多种运动物体，如车辆，行人等，可以有效地测试运动检测算法的泛化能力(generalization)和鲁棒性（robustness)。

第三，由于数据库中的视频都是由移动中的摄像机拍摄的，去掉了“摄像机固定”的假设条件，使得新开发的算法能在更具挑战性的情况下正常工作。以后这个数据库将在网上公开，为研究同行测试运动检测算法提供了一个新的平台，这无疑将推动关于运动检测模型和算法的研究和开发。

2.2　数据库的采集与标定

运动感知数据库中采集的视频主要包含室内和室外两大类场景(见图2.1)。所有视频片段均是通过一个手持家用摄像机拍摄，拍摄过程中摄像机有一定自动（ego-motion)，摄像机的

参数是：索尼DCR-SR47E，采样率20Hz。这些视频片段中包含了多种类别的运动物体，如行人，汽车和自行车，打篮球的运动员等。相对于以前只包含单一物体的运动检测数据库，这增加了运动感知数据库的多样性和对算法的检测难度。

在数据库的标定过程中，在保持相同采样率的情况下，视频片段中物体的运动情况基本相似，因此就不必每一帧都人工标定。在MATLAB环境下我们编写了一个具有GUI的人工标定程序，受试者以0.5s的标定间隔对所有数据库中的视频片段进行标定，对运动物体标定的标准是：要求受试者用方框框出视频中自己觉得在运动的物体。共有11位受试者参与标定了所有的视频。在随后的讨论中，我们把人工标定数据库的人称为受试者。数据库的统计性质如表2.1所示。

特性	数量
视频	20
帧数	2557
受试者	11
关键帧	297
标定框	4785

表 2.1 运动感知数据库概要.

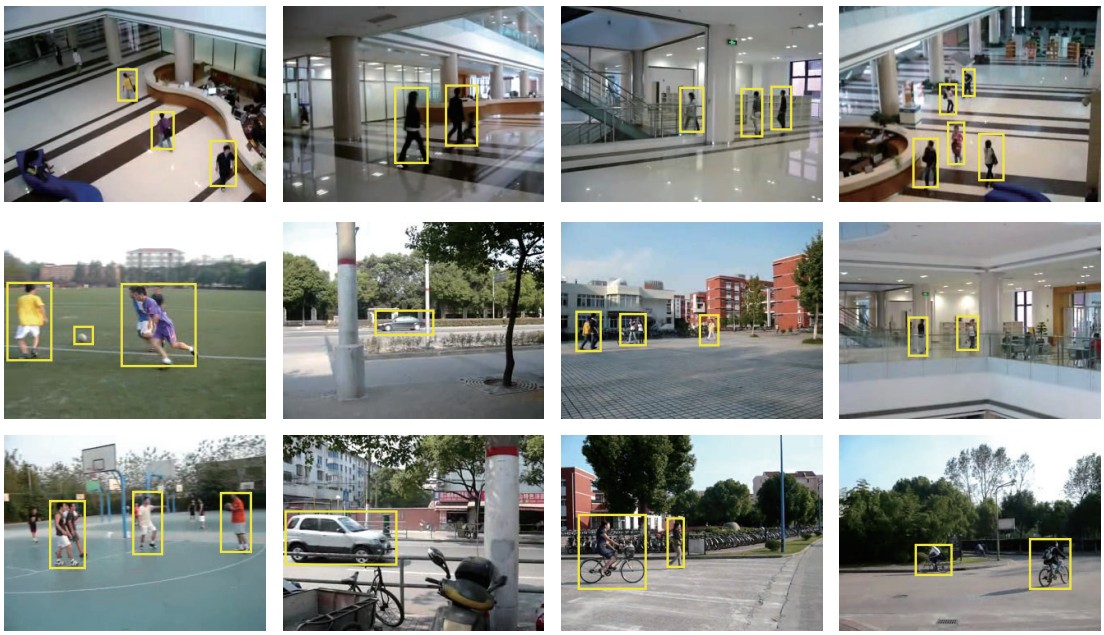


图 2.1 数据库中的部分视频样本。由于摄像机具有自动，场景和其中的典型物体均在运动。

2.3 运动感知的定量数值分析

人对外界客观环境的感知普遍存在差异。尤其在自然场景下，人的视觉系统对外界的感知（perception)和解析(interpretation)会受到诸多影响，如场景本身的复杂度，注意力集中程度

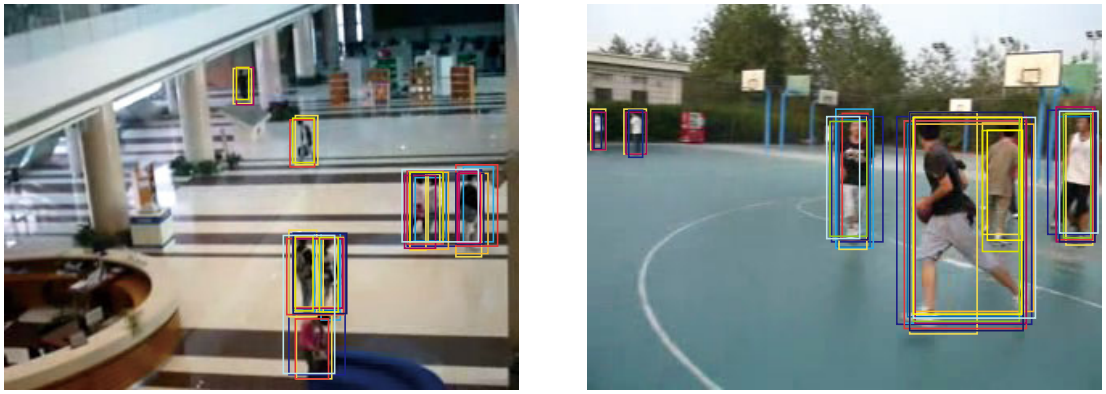


图 2.2 运动感知的个体差异。这两张图重叠了11个受试者对同个视频关键帧的标定。相同颜色的标定框指代同一受试者。我们可以明显看出不同受试者对自然场景中的运动物体有着不同的判断和标注，这种个体差异不能够被忽略。

等，以及一些主观对运动判断标准的差异。但是，在这些个体差异性之后，我们仍然可以观察到一些共有的视觉感知计算的统计特性，以分析视觉计算的通有机制。

在标定好的运动感知数据库中，我们观察到了明显的运动感知的个体差异，如图2.2所示。为了揭示人在自然场景中的运动感知的个体上的差异性和统计上的一致性，我们对视觉感知数据库中的标定数据做了如下测试：

在评估准则下，轮流以i^{th} 受试者的标定结果作为基准(ground-truth)，测试j^{th} 受试者的标定结果。我们有11个受试者的标定，每个受试者的标定可以与其他10个受试者的标定进行测试，那么我们可以得到110组测试结果。具体测试过程和110组测试结果的统计分析在下面两小节详述。

2.3.1 评估准则

运动感知数据库的测试评估准则与公认的PETS [20]中的测试评估准则一致。以R_{GT} 指代作为基准（ground truth）的受试者标注的方框区域。R_D指代作为测试组的受试者标注的方框区域。一次测试被当做True Positive，如果满足：

$$\frac{Area(R_{GT} \cap R_D)}{Area(R_{GT} \cup R_D)} \geq Th, \tag{2.1}$$

这里$Th = 0.5$ 。否则计为False Positive。这样，测试组受试者标注的所有方框被计为True Positive或者False Positive。

当i^{th} 受试者标定作为测试组, j^{th} 受试者的标定作为基准（ground truth）, 对于n^{th} 视频，我们用$GT_n^{i,j}, TP_n^{i,j}, FP_n^{i,j}$ 分别指代ground truth的计数, true positive的计数, 和false positive的计数。总的Detection Rate(DR) 和False Alarm Rate (FAR) 这样计算:

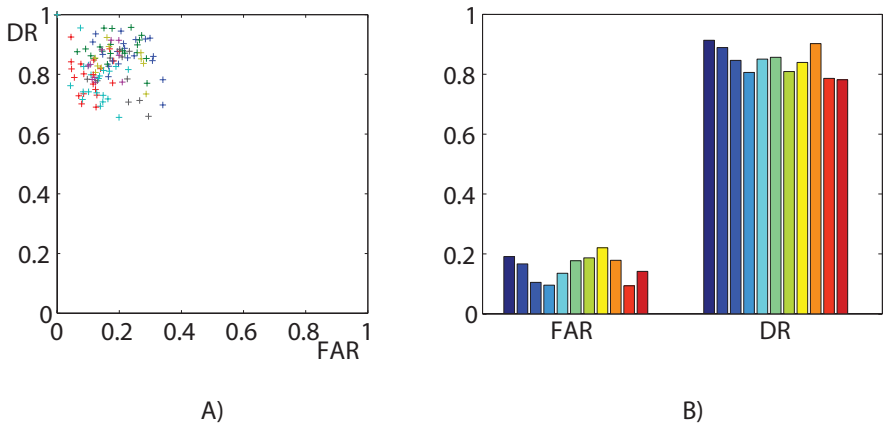


图 2.3 **A**): 表示所有11 位受试者测试结果的110对数据点投射到FAR-DR空间。相同颜色的+指代某个受试者的测试结果。对于不同的受试者，运动感知的测试结果DR 在0.65 至1 波动, FAR 在0 至0.4 波动。人的运动感知能力在这个数据库上的平均表现为：$FAR = 0.15 \pm 0.08$, $DR = 0.84 \pm 0.08$。**B**): 不同颜色的竖条代表不同受试者的FAR 和DR。

$$
\begin{aligned}
DR^n_{i,j} &= \frac{TP^n_{i,j}}{GT^n_{i,j}} \\
FAR^n_{i,j} &= \frac{FP^n_{i,j}}{TP^n_{i,j} + FP^n_{i,j}} \\
DR_{i,j} &= \frac{\sum_{n=1}^{N} DR^n_{i,j}}{N} \\
FAR_{i,j} &= \frac{\sum_{n=1}^{N} FAR^n_{i,j}}{N}
\end{aligned}
$$

DR, FAR即作为测试的分数。DR表明正确率，FAR代表误判率，DR越高，FAR越小则测试的结果越好。对于一帧关键帧里有多个方框的标注，寻找等式2.1中与作为基准的方框的匹配非常难。一个近似解决办法是，将测试组中的方框与基准中的方框一一进行匹配，然后再计算。实际实验中这种解决办法对最后结果的影响非常小。

2.3.2 结果

在上节的评估准则下，我们可以得到110组$(FAR_{i,j}, DR_{i,j})$，将这些点投射到FAR-DR空间，如图2.3所示。我们可以看出，对于不同的受试者，运动感知的测试结果DR 在0.65 至1 波动, FAR 在0 至0.4 波动。人的运动感知能力在这个数据库上的平均表现为：$FAR = 0.15 \pm 0.08$, $DR = 0.84 \pm 0.08$。这即表明了人的运动感知能力在自然场景中的差异。我们通过该数值分析将运动感知的差异定量化计算了出来，可以看出，个体差异在实验中很显著，那么随后在采样该视觉感知数据库测试运动检测的机器算法时需要将受试者在标定过程中的个体差异性考虑进去。在后面的实验部分中，我们修正了评估准则，

另外，注意到在等式2.1中, 实验阈值参数$Th = 0.5$ 的选择是我们人工设定的。这个参数的高低决定了实验的难度界限。为了估计Th 对实验结果的影响, FAR和DR 可以计算成以Th (见等式2.1)的函数。计算结果如表2.2所示。

Th	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
Human Detection Rate	0.92	0.91	0.91	0.90	0.88	0.84	0.77	0.62	0.37	0.15	0.00
Human False Alarm Rate	0.07	0.07	0.07	0.08	0.11	0.15	0.22	0.37	0.62	0.85	1.00

表 2.2 人的运动感知的实验结果(DR,FAR)随实验阈值Th 的变化。

2.4 本章小结

本章论述了建立运动感知数据库的整个过程，以及对数据库人工标定数据的数值分析结果。通过定量的统计分析，我们证明了人在自然场景中运动感知存在显著性差异的结论，并且计算出了人在自然场景下运动感知的差异性大小，这为后面建立和推导运动感知计算模型打下了基础。

从人的运动感知在本数据库上的表现结果来看，即使是人，在该数据库上对运动物体进行检测也不能做到100%完美的表现，那么当运动检测机器算法在这个数据库上进行测试时，则需要把这种数据库本身的稳定性考虑进去。实际上，对于机器算法测试的数据库，并没有绝对客观的基准，作为测试基准的手工标定的数据本身，就存在显著的差异性(如上面的人的运动感知测试实验的所得结论)。然而，公共数据库手工标定数据的稳定性问题并没有受到广泛的重视，List *et.al.* [21] 分析了CAVIAR [19]数据库中手工标定数据的统计性质, 指出不同人的手工标定数据存在较大差异性，这种差异性已经显著性得影响到最后机器算法测试的结果。在我们的数据库中，我们也观察到了同一个视频可以被不同受试者做出不同的感知，如图2.2.A所示，一些受试者把几人并排走的情况用一个方框标定出来，而一些受试者则倾向于把每个行人都用一个方框标定出来。

在随后的计算模型及算法的比较实验中，我们将以这个运动感知数据库为平台，测试我们建立的这个运动感知算法，并与同类运动检测算法进行对比。

第三章　运动感知模型及其算法

3.1　引言

在这一章中，我们建立了一个新的运动感知计算模型，并对模型进行了理论推导和分析，最后给出了相应的计算机算法实现代码。

3.1.1　相关工作

2001年，Vernon [22] 提出了一个利用傅里叶空间变换来分解场景中物体运动的理论。在他的理论中，运动物体的分离和物体的速度都可以通过求解一个线性系统方程而获得。根据傅里叶变换的时移性质，一个物体在空间中的运动带来的仅仅是物体在傅里叶表征下的位移差变化。对于具有m个物体的场景，准确计算物体的分离和速度需要求解一个具有$2m$个未知数的线性方程。这个方法有如下不足：1物体数量m必须事先知道；2，建立这个线性方程需要观察到$2m$帧图像；3，在这2m帧时间序列图像中物体的速度必须保持一致。这些限制使得Vernon的方法很难在实际工程中得到应用。

3.1.2　运动感知模型概述

我们的观点与Vernon的类似: 在像素空间里分布的复杂信息可以在傅里叶空间里简明有效地表征出来。与其为一个限定性问题找到准确解，不如在尽可能少的假设下寻找这个问题的近似解。

为了在动态场景中分离出运动物体，我们的模型跟随了预测编码（predictive coding）的思想。首先，我们通过预测背景成分的运动重构出下一帧图像。然后将我们的预测结果与下一帧的真实观察值进行对比，这样就使得代表前景运动成分的像素因为预测的错误而凸现出来。根据严格的理论分析，我们将新推导的运动模型的算法用9行MATLAB代码实现出来。

3.2　理论推导

我们把$f(\mathbf{x},t)$ 当成在t时刻的图像采样[1], 这里$\mathbf{x}=[x_1,x_2]^\top$ 是表征像素位置的两维向量空间. 一幅图像的像素表征可以写成如下形式:

$$f(\mathbf{x},t)=\sum_{\mathbf{x}_i\in\mathcal{I}} f(\mathbf{x},t)\cdot\delta_{\mathbf{x}_i}(\mathbf{x}), \tag{3.1}$$

这里$\mathcal{I}$ 指代像素的集合. 对于任意一幅自然场景图像，可以二分为$\mathcal{I}=\{\mathcal{F}_t,\mathcal{B}_t\}$. 每个像素属于前景$\mathcal{F}_t$ 或者背景$\mathcal{B}_t$. 指示函数$\delta_{\mathbf{x}_i}(\mathbf{x})$ 定义为:

$$\delta_{\mathbf{x}_i}(\mathbf{x})=\begin{cases}1 & \text{如果}\mathbf{x}=\mathbf{x}_i,\\ 0 & \text{其他.}\end{cases}$$

对于具有普通采样率的家用摄像头,我们可以假设：摄像头移动造成的帧到帧的运动可以近似为背景的一致性的平移. 如果我们知道这种平移的位移大小$\mathbf{v}=[v_1,v_2]^\top$, 根据像素强

[1]为了简明现在是以灰度图像的形式进行理论推导，RGB图像的处理在随后的算法实现部分有介绍

度一致性（*intensity constancy*）的假设条件[23],即像素在图像空间的平移不改变像素值的大小:

$$f(\mathbf{x},t)=f(\mathbf{x}+\mathbf{v},t+1). \tag{3.2}$$

我们就可以预测背景在下一帧图像上的位置。这个假设条件的成立需要像素$\mathbf{x}$ 在t时刻和$\mathbf{x}+\mathbf{v}$ 在$t+1$ 属于背景. 在这里我们用$\check{\mathcal{B}}_t$ 和$\hat{\mathcal{B}}_{t+1}$ 指代所有满足公式.3.2的像素集合.

只要我们计算出了摄像头自动(ego-motion)所带来的偏移量(即背景的偏移量), 这样我们就可以通过将当前帧的所有像素$\mathbf{x}$ 移动到$\mathbf{x}+\mathbf{v}$，以此重构出下一帧. 但是这种重构在当前帧的某些处于$\mathcal{I}-\check{\mathcal{B}}_t$的像素集合, 即自由运动的前景成分，会出现较大的差错. 因此, 出线重建差错的位置即是前景运动物体的空间位置。换句话来说，这种重构的差错图$s(\mathbf{x},t)$ 可以被当成运动感知的显著性图*(saliency map)* [13],严格的定义如下:

$$s(\mathbf{x},t)=\Big[f(\mathbf{x}+\mathbf{v},t+1)-f(\mathbf{x},t)\Big]^2. \tag{3.3}$$

3.2.1 傅里叶相位差与视角移动

为了生成显著性图, 我们需要知道背景位移的向量$\mathbf{v}$. 在傅里叶表征空间, 在等式.3.2中的空间位移可以简明有效地表征为傅里叶谱的相位差.

以$F_{\mathbf{x},t}(\boldsymbol{\omega})=\mathscr{F}\big(f(\mathbf{x},t)\cdot\delta_{\mathbf{x}_i}(\mathbf{x})\big)$ 指代为一个像素的2-D离散傅里叶变换, 这里$\boldsymbol{\omega}=[\omega_1,\omega_2]^\top$. 整幅图像的傅里叶变换$F_t(\boldsymbol{\omega})$ 可以由此获得:

$$F_t(\boldsymbol{\omega})=\sum_{\mathbf{x}_i\in\mathcal{I}}F_{\mathbf{x}_i,t}(\boldsymbol{\omega})$$

由傅里叶变换的时移性质[24], 空间域的平移只是带来傅里叶相位的改变，而保持傅里叶频谱不变:

$$F_{\mathbf{x}+\mathbf{v},t+1}(\boldsymbol{\omega})=F_{\mathbf{x},t}(\boldsymbol{\omega})e^{-i\cdot\Phi(\mathbf{v})}, \tag{3.4}$$

这里$\Phi(\mathbf{v})=\boldsymbol{\omega}^\top\mathbf{v}=\omega_1v_1+\omega_2v_2$, 在随后的讨论中，我们将它定义为相位差（*phase discrepancy*）。

因为整个背景具有一致的平移$\mathbf{v}$, 等式3.4 中的$\check{\mathcal{B}}_t$具有完备的形式:

$$\sum_{\mathbf{x}_i\in\hat{\mathcal{B}}_{t+1}}F_{\mathbf{x}_i,t+1}(\boldsymbol{\omega})=\sum_{\mathbf{x}_i\in\check{\mathcal{B}}_t}F_{\mathbf{x}_i,t}(\boldsymbol{\omega})e^{-i\cdot\Phi(\mathbf{v})}. \tag{3.5}$$

这样我们可以对此做如下分解:

$$\begin{aligned}
&F_{t+1}(\boldsymbol{\omega})=\sum_{\mathbf{x}_i\in\mathcal{I}}F_{\mathbf{x}_i,t+1}(\boldsymbol{\omega})\\
=&\sum_{\mathbf{x}_i\in\check{\mathcal{B}}_t}F_{\mathbf{x}_i,t}(\boldsymbol{\omega})e^{-i\cdot\Phi(\mathbf{v})}+\sum_{\mathbf{x}_i\in\mathcal{I}-\hat{\mathcal{B}}_{t+1}}F_{\mathbf{x}_i,t+1}(\boldsymbol{\omega})\\
=&F_t(\boldsymbol{\omega})e^{-i\cdot\Phi(\mathbf{v})}-\sum_{\mathbf{x}_i\in\mathcal{I}-\check{\mathcal{B}}_t}F_{\mathbf{x}_i,t}(\boldsymbol{\omega})e^{-i\cdot\Phi(\mathbf{v})}\\
&+\sum_{\mathbf{x}_i\in\mathcal{I}-\hat{\mathcal{B}}_{t+1}}F_{\mathbf{x}_i,t+1}(\boldsymbol{\omega}).
\end{aligned} \tag{3.6}$$

虽然这里看来如果不知道背景和前景的分离情况，是不可能计算$\Phi(\mathbf{v})$, 在随后的章节中我们将证明在一定的限定内，这里的相位差可以很好的近似计算.

3.2.2 傅里叶相位差的近似计算

由于无法量化和预测在$\mathcal{I}-\check{\mathcal{B}}_t$集合里的像素的形式和大小, 我们假设这些像素的位置和灰度值具有同一统计分布（uniform distribution). 在傅里叶空间，这个假设指示了$F_{\mathbf{x}_i,t}(\boldsymbol{\omega})$ 是满足同一分布的:

$$|F_{\mathbf{x}_i,t}(\boldsymbol{\omega})| \sim U(0,1)$$
$$\angle F_{\mathbf{x}_i,t}(\boldsymbol{\omega}) \sim U(0,2\pi).$$

以$\{z_i\}, i=1,2,\ldots,n$ 指代一系列独立的随机量, 这里$|z_i| \sim U(0,1), \angle z_i \sim U(0,2\pi)$. $\{z_i\}$ 具有期望μ 和方差σ^2. 以$Z_n=\sum_{i=1}^{n} z_i$ 指代这一系列随机变量的和。根据中心极限定则[25]，当n足够大时, Z_n 的渐进分布是高斯分布，即Z_n: $Z_n \sim N(n\mu, n\sigma^2)$. 由于随机变量具有相反的相位差，则它们互相抵消, 使得$\mu=0$. 由此可知, $|Z_n|$ 符合一个具有2个自由度的χ 分布:

$$p(|Z_n|=x) = \sqrt{n}\sigma x e^{-x^2/2}. \tag{3.7}$$

因此, 频谱的期望是由集合中像素的数量决定的，即:

$$\frac{E(|F_t(\boldsymbol{\omega})|)}{E(|\sum_{\mathbf{x}_i\in\check{\mathcal{B}}_t} F_{\mathbf{x}_i,t}(\boldsymbol{\omega})|)} = \frac{\sqrt{\#(\mathcal{I})}}{\sqrt{\#(\check{\mathcal{B}}_t)}}. \tag{3.8}$$

前景和背景集合中的像素数量可以由前面的运动感知数据库的标定结果估计。平均来说，我们前景运动物体的标定框大约占据了一帧图像的5% . 这个结果，结合等式3.6 和3.8可知，一幅图像在傅里叶空间的频谱主要由背景成分决定，不会受到前景成分的较大影响。

因此，我们可以把等式.3.6中的相位差近似为:

$$\tilde{\Phi}(\mathbf{v}) = \angle F_{t+1}(\boldsymbol{\omega}) - \angle F_t(\boldsymbol{\omega}). \tag{3.9}$$

其中，估计的错误由前景成分的像素和背景成分中被遮挡的像素组成。这里在频率$\boldsymbol{\omega}$ 累计的错误可以当成在等式.3.6中为$F_t(\boldsymbol{\omega})e^{-i\cdot\Phi(\mathbf{v})}$ 增加的噪声η , 即:

$$\eta = -\sum_{\mathbf{x}_i\in\mathcal{I}-\check{\mathcal{B}}_t} F_{\mathbf{x}_i,t}(\boldsymbol{\omega})e^{-i\cdot\Phi(\mathbf{v})} + \sum_{\mathbf{x}_i\in\mathcal{I}-\check{\mathcal{B}}_{t+1}} F_{\mathbf{x}_i,t+1}(\boldsymbol{\omega}).$$

在等式.3.8中, 我们把$F_t(\boldsymbol{\omega})e^{-i\cdot\Phi(\mathbf{v})}$ 设定为1，以其确定η 的分布:

$$\begin{aligned} E(|\eta|) &= \frac{\sqrt{2\#(\mathcal{I}-\check{\mathcal{B}}_t)}}{\sqrt{\#(\check{\mathcal{B}}_t)}} \approx \sqrt{0.1} \\ \angle\eta &\sim U(0,2\pi). \end{aligned}$$

因此，这个误差的上界$\tilde{\Phi}(\mathbf{v})$ 计算为:

$$\max\left[\Phi(\mathbf{v})-\tilde{\Phi}(\mathbf{v})\right] = \max\left\{\tan^{-1}\left[E(|\eta|)\right]\right\} \approx 0.31. \tag{3.10}$$

只要等式3.9 的假设成立, 我们可以由$F_t(\boldsymbol{\omega})$ 和$\tilde{\Phi}$重构出$\tilde{F}_{t+1}(\boldsymbol{\omega})$:

$$\begin{aligned} \tilde{F}_{t+1}(\boldsymbol{\omega}) &= F_t(\boldsymbol{\omega})e^{-i\cdot\tilde{\Phi}(\mathbf{v})} \\ &= |F_t(\boldsymbol{\omega})|\cdot e^{-i[\angle F_t(\boldsymbol{\omega})+\tilde{\Phi}(\mathbf{v})]} \\ &= |F_t(\boldsymbol{\omega})|\cdot e^{-i[\angle F_{t+1}(\boldsymbol{\omega})]} \end{aligned}$$

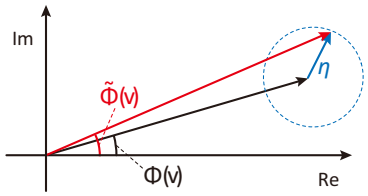


图 3.1 计算相位角度误差的示意图. 当$E(|\eta|)=\sqrt{0.1}$, 相位角度误差的上界为0.31 (17.6°), 相位角度误差的中值为0.21 (12.3°)

最后，运动感知的显著性图有如下简明的形式:

$$\begin{aligned} s(\mathbf{x},t) &= \left\{\mathscr{F}^{-1}\left[F_{t+1}(\boldsymbol{\omega})\right]-\mathscr{F}^{-1}\left[\tilde{F}_{t+1}(\boldsymbol{\omega})\right]\right\}^2 \\ &= \left\{\mathscr{F}^{-1}\left[\left(|F_{t+1}(\boldsymbol{\omega})|-|F_t(\boldsymbol{\omega})|\right)\cdot e^{-i\angle F_{t+1}(\boldsymbol{\omega})}\right]\right\}^2 \end{aligned} \tag{3.11}$$

3.2.3 消除边缘效应

2-D 离散傅里叶变换隐含“信号具有周期性”的假设。这个性质使得等式3.2中在边缘的像素在下一帧找不到对应值. 这样，使得运动感知的显著性图通常在边缘有较大的匹配错误，以致影响最后运动检测的结果(见图3.2C所示)，我们称之为边缘效应.

假设我们有前后两帧图像. 我们以$\mathcal{C}_1$ 和$\mathcal{C}_2$ 来指代导致边缘效应的像素. 这样:

$$\begin{aligned} \mathcal{C}_1 &= \{\mathbf{x}_i \mid \mathbf{x}_i \in \mathcal{B}_1, \mathbf{x}_i+\mathbf{v} \notin \mathcal{I}\} \\ \mathcal{C}_2 &= \{\mathbf{x}_i \mid \mathbf{x}_i \in \mathcal{B}_2, \mathbf{x}_i-\mathbf{v} \notin \mathcal{I}\} \end{aligned}$$

如果我们以第1 帧图像来估计第2 帧图像(如等式.3.11 所示), 我们在$\mathcal{C}_1$ 中有较大的错误。然而, 用等式.3.11来估计$\mathcal{C}_2$ 中的像素则没有错误. 同理可得, 如果我们调换时间序列$\longrightarrow$ 以第2 帧图像来估计第1 帧图像, 那么仅仅只有$\mathcal{C}_2$ 中的像素产生边缘效应.

写成严格的形式, 我们把以观测到第1 帧来估计第2 帧图像而产生的正时间序列的运动感知显著性图定义为$\overrightarrow{s}(\mathbf{x},t)$, 以及以观测到第2 帧来估计第1 帧图像而产生的反时间序列的运动感知显著性图记为$\overleftarrow{s}(\mathbf{x},t+1)$. 那么:

$$\begin{aligned} \overrightarrow{s}(\mathbf{x}_i,t) &> \varepsilon, && \text{当}\mathbf{x}_i \in \mathcal{C}_1 \\ \overleftarrow{s}(\mathbf{x}_i,t+1) &\le \varepsilon, && \text{当}\mathbf{x}_i \in \mathcal{C}_1 \\ \overrightarrow{s}(\mathbf{x}_i,t) &\le \varepsilon, && \text{当}\mathbf{x}_i \in \mathcal{C}_2 \\ \overleftarrow{s}(\mathbf{x}_i,t+1) &> \varepsilon, && \text{当}\mathbf{x}_i \in \mathcal{C}_2, \end{aligned}$$

这里ε 由等式.3.10约束.

我们最后以组合正反时间序列的运动感知显著性图来消除边缘效应，写成统一的形式即为:

$$s(\mathbf{x},t)=\sqrt{\overrightarrow{s}(\mathbf{x},t)\cdot\overleftarrow{s}(\mathbf{x},t+1)} \tag{3.12}$$

对于$\forall\mathbf{x}_i \in \mathcal{C}_1\bigcup\mathcal{C}_2$, 容易看出$s(\mathbf{x}_i,t)\to 0$ 当$\overrightarrow{s}(\mathbf{x}_i,t)\to 0$, 或者$\overleftarrow{s}(\mathbf{x},t+1)\to 0$. 通过等式.3.12产生的混合运动感知显著性图如图.3.2-D所示.

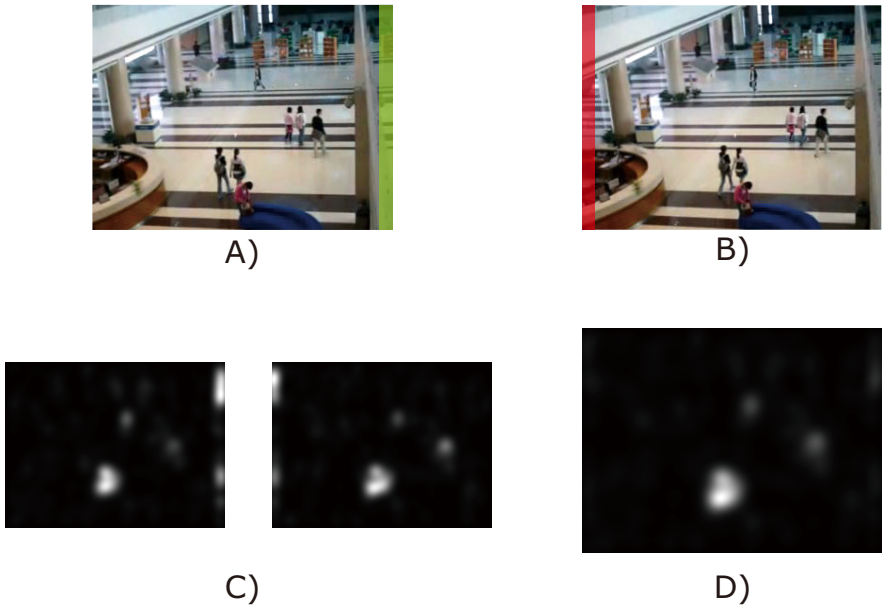


图 3.2 边缘效应的示意图. **A) & B)**: 两张相邻帧. 在各自帧中绿色和红色的区域分别指代$\mathcal{C}_1$和$\mathcal{C}_2$。**C)**: 单独根据正反时间序列产生的运动感知显著性图。可以看出边缘效应产生的误差明显影响了最后结果，不能忽略。**D)**: 混合运动感知显著图.

3.3 另一种理论推导

经过分析，我们的运动感知计算模型还可以从傅里叶空间前景背景成分分离的角度进行理论推导，过程如下：

假设$f_t(x,y)$ 是在t^{th}时刻的一帧图像的像素空间表征，它由1个背景成分和m(未知) 个前景成分组成,即：

$$f_t(x,y) = f_t^0(x,y) + \sum_{k=1}^{m} f_t^k(x,y) \tag{3.13}$$

$f_t^0(x,y)$ 定义为背景成分, $f_t^k(x,y)$ 是前景成分。

在时间间隔δt后,

$$f_{t+\delta t}(x,y) = f_{t+\delta t}^0(x,y) + \sum_{k=1}^{m} f_{t+\delta t}^k(x,y) \tag{3.14}$$

这里我们可以假设在δt时间间隔内像素是平移的。这个假设即是常见的用在光流计算中的像素强度一致性假设（intensity constancy）[23]:

$$f_{t+\delta t}^i(x,y) = f_t^i(x - v_x^i \delta t, y - v_y^i \delta t) \tag{3.15}$$

这里$i = 0, 1, ..., m$, (v_x^i, v_y^i) 是i^{th} 图像成分在空间域的速度。这里需要强调的是(v_x^0, v_y^0) 背景成分在空间域的速度。

每个图像成分的傅里叶变换表征为$F_t^i(w_x, w_y)$:

$$F_t^i(w_x, w_y) = \mathscr{F}(f_t^i(x,y)) \tag{3.16}$$

反傅里叶变换为:

$$f_t^i(x,y) = \mathscr{F}^{-1}(f_t^i(x,y)) \tag{3.17}$$

根据傅里叶空间的时移不变性[24]，

$$\mathscr{F}(f_t^i(x-v_x^i\delta t, y-v_y^i\delta t) = F_t^i(x,y)e^{-i(w_x v_x^i \delta t + w_y v_y^i \delta t)} \tag{3.18}$$

因此我们定义$\Delta\Phi_{\delta t}^i = e^{-i(w_x v_x^i \delta t + w_y v_y^i \delta t)}$, 所得：

$$F_{t+\delta t}^i(x,y) = F_t^i(x,y)\Delta\Phi_{\delta t}^i \tag{3.19}$$

对等式3.13 和等式3.14进行傅里叶变换，我们得到

$$F_t(w_x,w_y) = F_t^0(w_x,w_y) + \sum_{k=1}^{m} F_t^k(w_x,w_y) \tag{3.20}$$

$$F_{t+\delta t}(w_x,w_y) = F_t^0(w_x,w_y)\Delta\Phi_{\delta t}^0 + \sum_{k=1}^{m} F_t^k(w_x,w_y)\Delta\Phi_{\delta t}^k \tag{3.21}$$

当等式3.21两边同除$\Delta\Phi_{\delta t}^0$, 然后减去等式3.20 我们得到

$$\sum_{k=1}^{m} F_t^k(w_x,w_y)(\frac{\Delta\Phi_{\delta t}^k}{\Delta\Phi_{\delta t}^0} - 1) = F_{t+\delta t}(w_x,w_y)\frac{1}{\Delta\Phi_{\delta t}^0} - F_t(w_x,w_y) \tag{3.22}$$

这里我们定义

$$PDI_{\delta t}(w_x,w_y) = \sum_{k=1}^{m} F_t^k(w_x,w_y)(\frac{\Delta\Phi_{\delta t}^k}{\Delta\Phi_{\delta t}^0} - 1) \tag{3.23}$$

$F_t(w_x,w_y)$ 在频率域的相位差增量PDI（Phase Discrepancy Increment)。由等式3.23 我们可以看到相位差增量衡量PDI的是前景成分$\Delta\Phi_{\delta t}^k$和背景成分$\Delta\Phi_{\delta t}^0$ 在频率域的相角差。值得一提的是当$\Delta\Phi_{\delta t}^k = \Delta\Phi_{\delta t}^0$, 则$k^{th}$ 前景成分对PDI不进行作用。

因此，PDI的反傅里叶变换时，

$$ipd_{\delta t}(x,y) = \mathscr{F}^{-1}(\sum_{k=1}^{m} F_t^k(w_x,w_y)(\frac{\Delta\Phi_{\delta t}^k}{\Delta\Phi_{\delta t}^0} - 1)) \tag{3.24}$$

它表征了前景成分的空间特性，如运动位置，相对运动的强度等等。

进一步，容易证明的是当$|F_t^0(w_x,w_y)| \gg |F_t^k(w_x,w_y)|_{k=1,...,m}$，则$\Delta\Phi_{\delta t}^0 \approx \Delta\Phi_{\delta t}$ ，这里$|F_t^i(w_x,w_y)|$ 是$F_t^i(w_x,w_y)$ $i=0,1,...,m$ 的傅里叶频谱，$\Delta\Phi_{\delta t}$ 是$F_t(w_x,w_y)$ 在δt 时刻的整幅图像的傅里叶相位增量。这个假设可以解释为前景成分的傅里叶谱能量远远小于背景成分的谱能量，并且，前景成分更具不确定性及比背景成分还有更多引起视觉注意力的信息。在这个假设之下，

$$\begin{aligned} & F_{t+\delta t}(w_x,w_y)\frac{1}{\Delta\Phi_{\delta t}^0} - F_t(w_x,w_y) \\ = \;& F_{t+\delta t}(w_x,w_y)\frac{1}{\Delta\Phi_{\delta t}} - F_t(w_x,w_y) \\ = \;& |F_{t+\delta t}(w_x,w_y)|\Phi_{t+\delta t}\frac{1}{\Delta\Phi_{\delta t}} - F_t(w_x,w_y) \\ = \;& |F_{t+\delta t}(w_x,w_y)|\Phi_t - F_t(w_x,w_y) \end{aligned}$$

这里$\Phi_t = e^{i\phi(w_x,w_y)}$是$F_t(w_x,w_y)$的傅里叶相位角. 因此再由等式3.22可得，相位差增量PDI（Phase Discrepancy Increment）可以简明地由此计算：

$$PDI_{\delta t}(w_x,w_y) = |F_{t+\delta t}(w_x,w_y)|\Phi_t - F_t(w_x,w_y) \tag{3.25}$$

$F_t(w_x, w_y)$ 在t时刻观察到的一帧图像$f_t(x, y)$的傅里叶空间表征, $F_{t+\delta t}$ 是在随后δt时间间隔内观察到的一帧图像$f_{t+\delta t}(x, y)$的傅里叶空间表征。$|F_{t+\delta t}(w_x, w_y)|\Phi_t$ 意味着以$F_t(w_x, w_y)$的相位代替$F_{t+\delta t}(w_x, w_y)$ 的相位。

这样，我们从傅里叶空间成分分析的角度也推导出了运动感知计算模型。这说明了我们这个模型的普适性和理论分析的多样性。

3.4 算法实现

在MATLAB中, 基于相位差的运动感知模型可以以如下简明的算法实现:

```
FFT1=fft2(Frame1);
FFT2=fft2(Frame2);
Amp1=abs(FFT1);
Amp2=abs(FFT2);
Phase1=angle(FFT1);
Phase2=angle(FFT2);
mMap1=abs(ifft2((Amp2-Amp1).*exp(i*Phase1)));
mMap2=abs(ifft2((Amp2-Amp1).*exp(i*Phase2)));
mMap=mat2gray(mMap1.*mMap2);
```

`Frame1` 和`Frame2` 是相邻的两帧. 在我们的实验中, 视频的每一帧为120×160 的灰度图像. 在普通的2.2GHz Core 2 Duo 笔记本计算机上, 这个算法的处理速度高达75 fps。

对于处理彩色图像，一种自然的算法推广是把RGB每一个颜色通道单独处理，最后把3张生成的运动感知显著图线性叠加。然而，这样会使得处理时间增加3倍，且相对于灰度图的处理效果，结果只有微弱提高，这在随后的实验部分有详细说明。由于我们的这个算法是强调算法的高效速度，所以在随后的实验中我们都是以灰度图像为主进行算法测试。

另外，我们注意到在自然场景中，像素强度一致性（*intensity constancy*）的假设[23]受到诸多因素的影响，如背景成分的随机运动（如树叶的随风飘动）、采样伪迹、CCD噪声等。一个去掉这些噪声的方法是线性叠加多帧的运动感知显著性图。但是如果过度叠加，这种方法会对运动物体的定位产生较大误差。在实验中，我们叠加了从连续5帧图像产生的运动感知显著性图。在采样率20Hz，5帧耗费0.25秒，这个方法有效地减小了噪声，而没有引起较大的运动物体位置偏移误差(结果见图3.3).

3.5 本章小结

在这一章中，我们通过两种不同的理论分析方法对运动感知计算模型进行了推导和分析，最后得到了统一的结论，即：傅里叶空间的相位差成分在运动感知的过程中具有重要理论作用，它有效地表征出了物体单独的运动成分。这也正是我们的运动感知计算模型的核心。随后，运动感知模型的算法由9行MATLAB代码简明地实现出来。在随后的实验部分，我们将在自然场景中测试该算法的效果和效率，以证明我们的运动感知模型的优越性能。

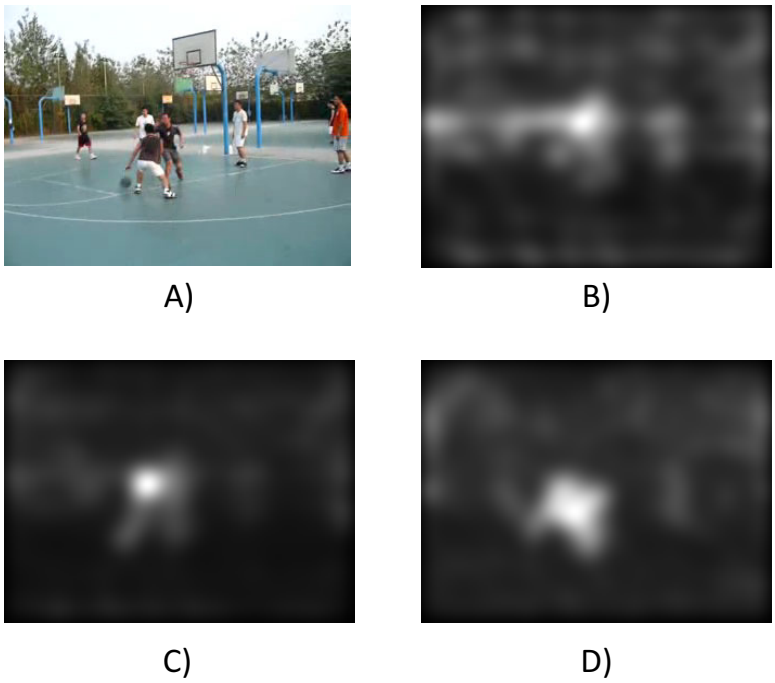


图 3.3 叠加多帧运动感知显著性图对结果的影响. **A**): 单帧的运动感知显著性图. 背景噪声明显可见. **B**): 叠加5帧运动感知显著性图的结果. 背景噪声被有效抑制，检测到的运动物体的位置依然准确。**C**:) 叠加20帧运动感知显著性图的结果. 因为叠加的时间窗口为1s, 运动物体在这1s时间区域内会有较大的移动，造成运动物体的位置定位不准确。

第四章　实验分析

4.1　引言

在上一章中，我们推导出了新的运动感知模型及其算法，并对计算模型进行了理论分析。这一章中，我们将在建立好的运动感知数据库平台之上测试该算法的性能，并与其他同类运动检测的算法进行比较。

4.2　评估准则

这里机器算法的评估准则依据第二章中建立的测试准则，即：以R_{GT} 指代作为基准（ground truth）的人工标注的方框区域。R_D指代算法检测出的的运动区域。一次测试被当做True Positive，如果：

$$\frac{Area(R_{GT} \cap R_D)}{Area(R_{GT} \cup R_D)} \geq Th, \tag{4.1}$$

这里$Th = 0.5$ 。否则计为False Positive。这样，算法检测出的所有方框被计为True Positive或者False Positive。

对于n^{th} 视频, 以i^{th} 个受试者的标注作为ground truth, 我们用GT_n^i, TP_n^i, FP_n^i 分别指代ground truth的计数, true positive的计数, 和false positive的计数。算法的测试结果Detection Rate(DR)和False Alarm Rate (FAR) 这样计算:

$$\begin{aligned} DR_n &= \frac{\sum_i TP_n^i}{\sum_i GT_n^i} \\ FAR_n &= \frac{\sum_i FP_n^i}{\sum_i TP_n^i + FP_n^i}. \end{aligned}$$

DR,FAR即作为测试的分数。DR表明正确率，FAR代表误判率，DR越高，FAR越小则测试的结果越好。对于一帧关键帧里有多个方框的标注，寻找等式4.1中与作为基准的方框的匹配非常难。一个解决办法是，将测试组中的方框与基准中的方框一一进行匹配，然后再计算。实际实验中这种解决办法对最后结果的影响非常小。

4.3　实验结果

为了从运动感知算法产生的显著性图（saliency map)上分割出不同的运动检测区域，一个分割算法需要知道图像尺度，区域数量等参数。为了使得我们的这个运动感知算法具有普适性，我们利用Non-Maximal Suppression [26]方法来从显著性图上分割运动检测区域。NMS算法有3个参数$[\theta_1, \theta_2, \theta_3]$。显著图分割的过程如下：

首先，该算法在整幅显著性图上找出半径为θ_1的局部极大点。每一个值大于θ_2 的局部极大点被选成一个检测方框的种子点。随后，显著性图根据阈值θ_3的大小被二值化。最后包裹检测方框种子点的白色区域按照其边缘的大小划分成一个方框。

我们可以假设，NMS算法参数调整的过程是独立于运动感知数据库中的20个视频文件的。因此，我们利用交叉效验（cross-validation）的方法来避免过度拟合（over-fitting）。在每次交叉效验循环中，我们把取出19个视屏文件作为训练组调整NMS参数，使得极大化:

$$\sum_{m\in\{training\}} DR_m(1-FAR_m),$$

然后利用剩下的1个视频作为算法的测试组。这样，算法的最后结果DR 和FAR 是在20次交叉效验中测试组视频的平均。图4.1中展示了一些我们的运动感知算法检测运动物体的定性结果。算法定量的结果在表4.1中。

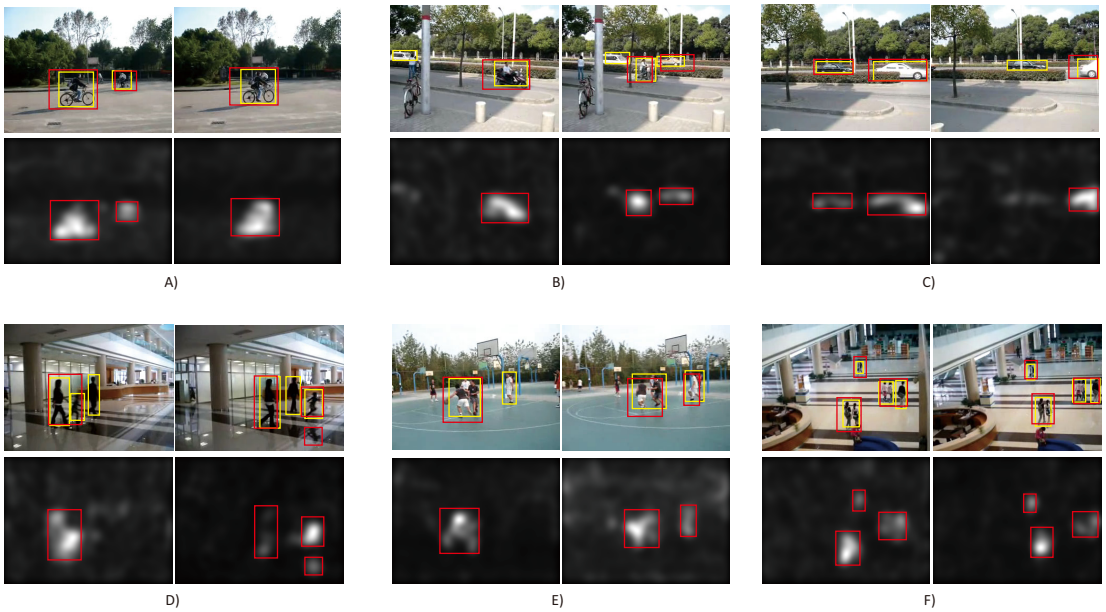


图 4.1 运动感知显著性图以及检测出的运动物体/区域。在每帧图像中，红色的方框是我们计算模型检测出的的运动区域，黄色的方框是数据库中人工标定的运动区域。

4.4 运动检测方法比较

	Detection Rate	False Alarm Rate
Human average	0.84 ± 0.08	0.15 ± 0.08
我们的模型（Phase Discrepancy）	0.46 ± 0.14	0.58 ± 0.24
动态注意力（Dynamic Visual Attention）[14]	0.32±0.22	0.86±0.10
贝叶斯惊讶（Bayesian Surprise）[15]	0.12±0.09	0.92±0.04
图像显著性（Saliency）[13]	0.09±0.08	0.98±0.01
混合高斯（Mixture of Gaussian）[5]	0.00±0.00	1.00±0.00

表 4.1 不同模型的测试结果。我们的模型取得了最佳分数。

为了更好的评估我们的运动感知计算模型，我们在相同的测试平台和评估准则下用4种具有代表性的运动检测方法进行比较: 混合高斯模型（the Mixture of Gaussian model）[5], 动态

注意力模型（the Dynamic Visual Attention model）[14], 贝叶斯惊讶模型（the Bayesian Surprise model）[15]和图像显著性模型（the Saliency model）[13]. 所有比较模型的MATLAB/C++算法实现均来自各作者的网页。各个算法生成的运动物体检测图的定性结果如图4.2所示. 在数据库上定量测试时, 分割运动物体检测图的算法依旧是NMS, 其参数调整的方法以交叉效验分别进行，各个算法的定量结果如表4.1所示. 我们的运动感知模型在所有模型中取得了最好的分数。

值得一提的是，这里比较的模型中，并非所有的模型算法最初的设计目的都是针对在动态场景中对运动物体进行检测。实际上，一个算法的表现是由它事先对数据本身的假设决定的。在我们的运动物体感知数据库上,“物体”的定义是由它与背景成分的相对运动的差异所决定，没有假设“物体”拥有具体的特性或者背景是单调一致的。所以，这也是某些算法在我们的运动感知数据库的测试中表现不好的原因之一。

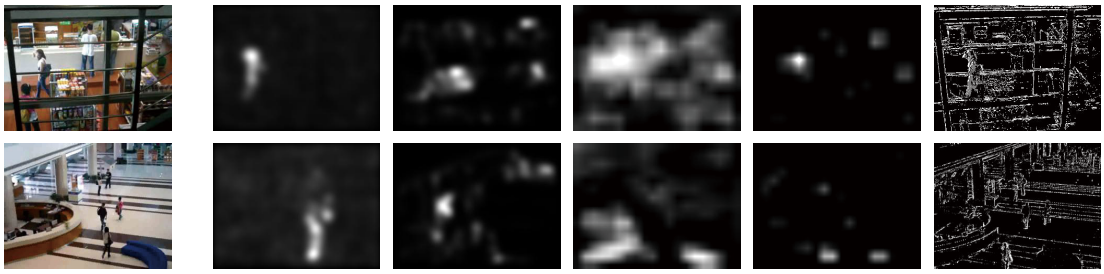


图 4.2 由不同算法产生的运动检测区域图。**A**): 原始帧. **B**): 我们的模型（phase discrepancy). **C**): 动态注意力（Dynamic Visual Attention）[14]. **D**): 贝叶斯惊讶（Bayesian Surprise）[15]. **E**): 图像显著性（Saliency）[13]. **F**): 混合高斯（Mixture of Gaussian）[5].

4.5 结果分析

4.5.1 算法与不同受试者的标定数据测试结果

为了测试数据库中不同标定者的个体标定差异对机器算法测试结果的影响，我们依次以每个受试者的标定数据作为基准测试了我们的算法，其结果如图4.3所示。算法的平均结果是$FAR = 0.59 \pm 0.12, DR = 0.47 \pm 0.19$。

可以看出，单独以某个受试者的标定数据测试机器算法时存在较大误差，故我们在用这个运动感知数据库测试机器算法时，是整合了所有受试者的标定数据（如等式4.2所示），测试所得的结果更准确，更具实际对比意义。另一方面，这也证明了我们的观点，人在运动感知过程中存在显著性差异，在实验过程中必须考虑。

4.5.2 实验阈值Th与算法的表现程度的关系

另外，注意到在等式2.1中, 实验阈值参数$Th = 0.5$ 的选择是我们人工设定的。这个参数的高低决定了实验的难度界限。为了估计Th 对实验结果的影响, FAR和DR 计算成以Th（见等式4.1)的函数，结果如表4.2所示。其结果的变化程度也表示了模型的稳定性。

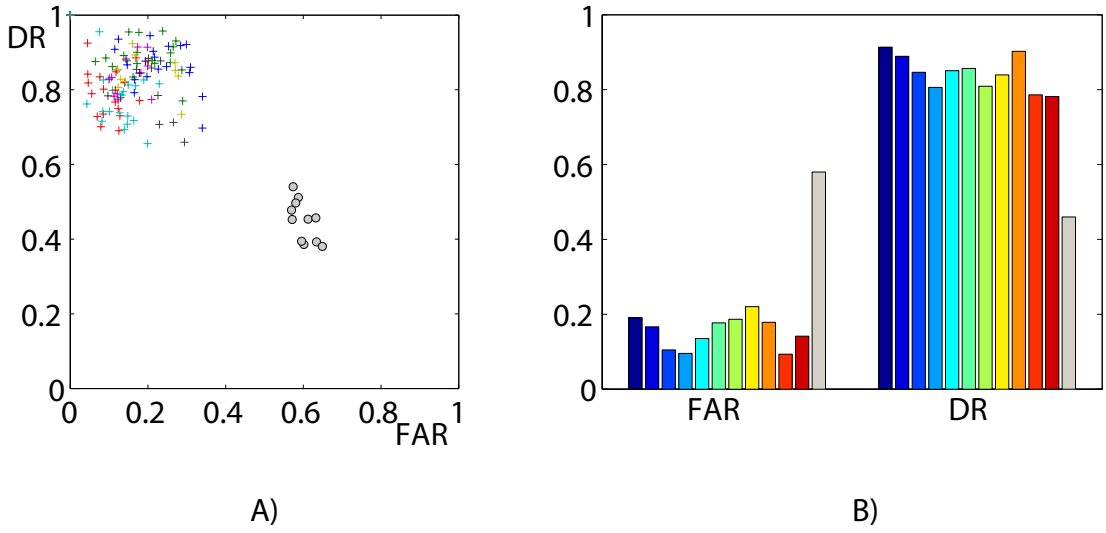


图 4.3 **A)**: 11 个受试者的测试结果和我们算法的测试结果投射到FAR-DR空间。每个相同颜色的+ 指代某个受试者的测试结果。○指代对于不同受试者的标定数据作为基准测试的结果。**B)**: 不同颜色的竖条代表不同受试者的平均FAR 和DR。灰色竖条代表算法的平均测试结果FAR和DR。

Th	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
human Detection Rate	0.92	0.91	0.91	0.90	0.88	0.84	0.77	0.62	0.37	0.15	0.00
Human False Alarm Rate	0.07	0.07	0.07	0.08	0.11	0.15	0.22	0.37	0.62	0.85	1.00
Model Detection Rate	0.83	0.82	0.80	0.75	0.63	0.46	0.20	0.07	0.02	0.00	0.00
Model False Alarm Rate	0.18	0.20	0.24	0.31	0.43	0.58	0.80	0.93	0.98	1.00	1.00

表 4.2 人的运动感知的实验结果(DR,FAR)和模型的运动感知的实验结果(DR,FAR)随实验阈值Th 的变化。

4.5.3 运动物体大小对算法的影响

在模型理论推导的等式3.10中, 模型的上界是关于物体大小的函数。为了提供一个对于较大运动物体本算法也能正常工作，我们在运动感知数据库中选出了2个具有较大运动物体的视频，来测试我们的算法。前景物体的平均区域大小大约是图像大小的10% (对比于原实验占据5%)。新的实验结果列于表4.3中.

	原实验	较大物体实验
Detection Rate	0.46 ± 0.14	0.41 ± 0.14
False Alarm Rate	0.58 ± 0.24	0.65 ± 0.08

表 4.3 较大运动物体检测比较实验。可以看出，算法的实验结果只是下降了很小一点。

4.5.4 误差来源

实验中一个很大的挑战是对邻近的或者部分被遮挡的多个物体运动检测区域的分割（如图4.1.F所示). 如果要解决这个难题，则需要采用物体跟踪算法或者更强大的对显著性图的区域分割算法。

在某些情况下，我们也需要整合来自物体识别的自上而下(top-down)的信息。由于我们算法产生的显著性图是基于像素自身特征的，它更倾向于单独检测出物体运动的部分（如挥动的手），而非整个物体。一个有趣的例子在图4.1.D中: 我们的算法检测出了地板上的一个运动物体的倒影。然而在数据库标定的过程中，没有受试者把运动物体的倒影标记成运动中的物体。

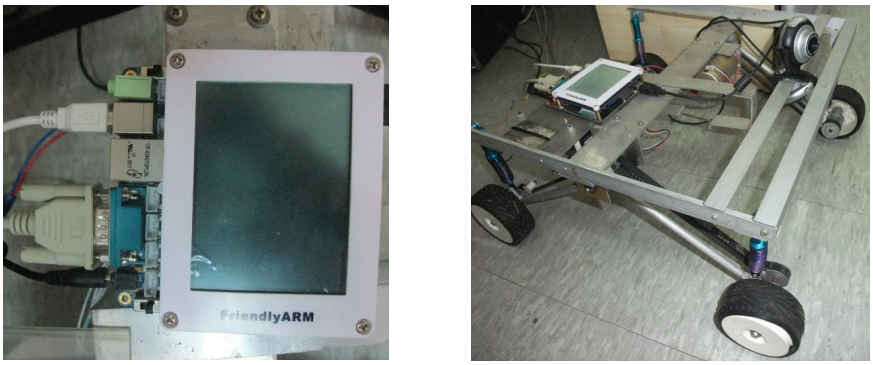
4.6 本章小结

在这章中，我们对推导出的运动感知计算模型的算法进行了一系列的实验测试，测试的数据库平台来自第二章手工标定的运动感知数据库。同时，我们在这个数据库上对比测试了4个具有代表性的运动检测模型和算法，定量的数据结果证明了我们这个运动感知模型的良好性能。

第五章　模型的应用

5.1　硬件实现和实时测试

5.1.1　硬件平台



A)　　B)

图 5.1 **A)**:基于ARM9嵌入式系统平台MINI2440，我们在这个平台上测试运动感知算法的计算能力。**B)**:将装载有运动感知算法的机器视觉系统与智能小车的控制系统整合，这样小车可以根据机器视觉系统提供的线索进行自动控制。

为了实际测试运动感知算法的工程应用特性，我们在嵌入式平台MINI2440上实现了该算法。

如图5.1所示，MINI2440是一块ARM9开发板，其硬件资源特性如下：

CPU 处理器

- Samsung S3C2440A，主频400MHz，最高533Mhz

SDRAM 内存

- 在板64M SDRAM
- 32bit 数据总线
- SDRAM 时钟频率高达100MHz

FLASH 存储

在板256M/1GB Nand Flash, 掉电非易失(用户可定制64M/128M/256M/512M/1G)

在板2M Nor Flash，掉电非易失，已经安装BIOS

LCD 显示

- 板上集成4 线电阻式触摸屏接口，可以直接连接四线电阻触摸屏
- 支持黑白、4 级灰度、16 级灰度、256 色、4096 色STN 液晶屏，尺寸从3.5 寸到12.1 寸，屏幕分辨率可以达到1024x768 象素；
- 支持黑白、4 级灰度、16 级灰度、256 色、64K 色、真彩色TFT 液晶屏，尺寸从3.5 寸到12.1 寸，屏幕分辨率可以达到1024x768 象素；
- 标准配置为统宝3.5”真彩LCD，分别率240x320，带触摸屏；接口和资源

- 1 个100M 以太网RJ-45 接口(采用DM9000 网络芯片)
- 3 个串行口
- 1 个USB Host
- 1 个USB Slave B 型接口
- 1 个SD 卡存储接口
- 1 路立体声音频输出接口，一路麦克风接口；
- 1 个2.0mm 间距10 针JTAG 接口
- 4 USER Leds
- 6 USER buttons(带引出座)
- 1 个PWM 控制蜂鸣器
- 1 个可调电阻，用于AD 模数转换测试
- 1 个I2C 总线AT24C08 芯片，用于I2C 总线测试
- 1 个2.0 mm 间距20pin 摄像头接口
- 板载实时时钟电池
- 电源接口(5V)，带电源开关和指示灯

系统时钟源

- 12M 无源晶振

扩展接口

1 个34 pin 2.0mmGPIO 接口

1 个40 pin 2.0mm 系统总线接口

规格尺寸

- 100 x 100(mm)

操作系统支持

- Linux2.6.32.2 + Qtopia-2.2.0
- WindowsCE.NET 6.0(R3)

我们用ARM ADS(ARM Developer Suite1.2)来编译运动感知算法的C语言实现，并通过H-JTAG进行仿真调试，加载到MINI2440进行测试。另外，我们通过Fedora系统与嵌入式系统内的Linux进行同步化，这样编写的C程序可以直接下载到MINI2440的Linux环境中执行，比前一种方法更为简单。故我们的算法实现都是在嵌入式Linux环境下进行。

5.1.2 嵌入式环境测试结果

我们在嵌入式Linux系统内加载了一段足球场踢球的视频（见图5.2），这段视频里既有多个运动物体，又具有自动（ego-motion），对运动检测的算法测试难度非常高。用已经加载于嵌入式Linux内核中的C语言实现的运动感知算法对其进行处理，部分连续处理的结果如图5.2所示。可以看出在嵌入式环境下我们的运动感知算法依然表现出色，对场景中的运动物体进行了有效检测。这表明了我们开发的运动检测算法有广泛的工程应用前景。

5.1.3 与机器视觉小车的整合

运动感知算法可以作为机器视觉系统的前级模块，整合到机器人或者智能小车系统之中。如图5.1所示，我们正在尝试把运动感知算法连同嵌入式系统MINI2440与机器小车的动力系统进行整合，这样，在人工智能控制系统的调控下，根据运动感知算法对外界环境的感知，智能小车可以自动进行导航和制定行进策略。NASA火星车的智能系统就是结合了计算

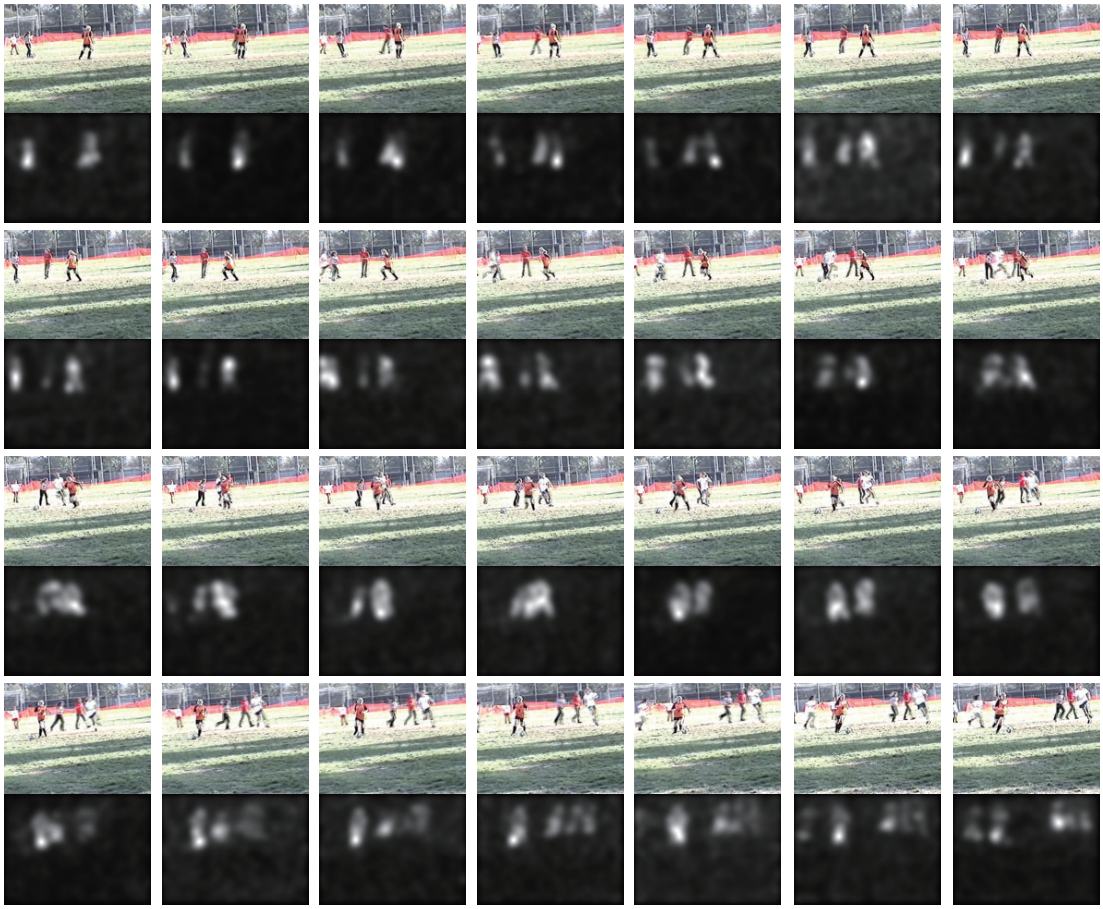


图 5.2 在ARM9嵌入式Linux环境下对运动感知算法进行测试结果。上一行是加载的足球场踢球视频样本帧，视频片段中具有多个运动物体，且摄像机本身具有较大的自动（ego-motion）。下一行是运动感知算法检测到的运动显著性图。可以看出，我们的运动感知算法在嵌入式环境下依然工作良好，表现出了广泛的工程应用前景。

机视觉算法对外界环境的检测结果进行综合决策，从而控制火星车的正常行进，如图5.3所示。

5.2 运动感知计算模型对眼动的预测

2006年，Itti [15]等研究人员用他们的Bayesian Surprise模型来预测和拟合视觉注意力和眼动的数据，以此来测试计算模型对视觉生理现象的预测情况。在神经计算研究领域，一个可以更好预测神经生理学现象的计算模型才是更有生理学依据的模型。因此，我们也利用眼动的数据来测试运动感知模型的预测能力。

实验的具体设定如[15]中所述，首先计算在人眼动注视位置对应的运动感知显著性图上的强度分布q_s，然后在显著性图上随机采样形成强度分布q_r，然后采用Kullback－Leibler divergence 度量[27]来计算两个分布的距离：

$$D_{\mathrm{KL}}(q_r \| q_s) = \sum_i q_r(i) \log \frac{q_r(i)}{q_s(i)} \tag{5.1}$$

图 5.3 火星车外形图（图片来源于NASA）。其感知系统包括雷达、视觉、激光等，其视觉导航系统所起的作用非常大。我们的运动感知算法有望应用到这些智能车系统之中去。

KL（Kullback－Leibler divergence）的值越大，说明计算模型对眼动位置的区分度越大，计算模型对眼动现象的预测越准确。其结果如图5.4所示，我们对比了另外两个视觉计算模型：图像显著性模型（Saliency Model）[13]和动态注意力（Dynamic Visual Attention Model）[14]，我们的运动感知模型取得了最好的预测结果。

5.3 本章小结

在这一章中，我们将运动感知计算模型的算法在ARM9嵌入式平台上进行了实现，算法在嵌入式Linux系统中运行良好，发挥出色，表现出了在工程系统中应用潜力。在以后的工作中，我们将把运动感知计算模型加载到智能小车之中去，从系统层面上测试计算模型的性能。

另一方面，我们利用运动感知模型来拟合了视觉注意力和眼动的生理学数据，在同类计算模型中拟合结果最好，说明了运动感知模型不仅具有广泛的工程学应用，还能对神经生理学现象进行预测，对揭示人类视觉系统计算机制有着积极的意义。在以后的研究中，我们希望看到有神经生理学的相关实验来验证我们的运动感知计算理论是否可以用来揭示人类视觉系统的运动检测机制。

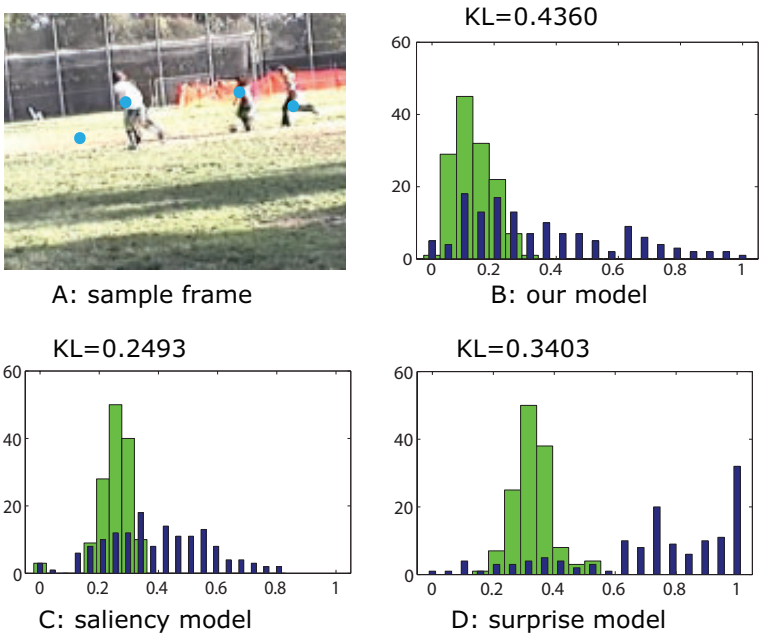


图 5.4 眼动的数据和视频刺激来自于[15]。视频刺激中包含运动的物体和自动。在这个视频刺激中，共有137次眼睛的扫视（saccade）被记录，如A中的蓝色点。我们把我们的模型与另外2个计算模型进行了对比实验：在获得模型生成的显著性图后，在眼动位置的显著性强度分布（蓝色的窄条带）和在随机位置的显著性强度分布（绿色的宽条带）可以被算出。这两个分布的KL距离用来衡量每个模型的分数。可以看出，我们的运动感知模型取得了最好成绩。)

第六章　结论

本文首先对人的运动感知行为进行心理物理学实验，对人在运动感知过程中的差异性和一致性进行数值量化分析，随后以受试者的标定数据作为为准则（ground-truth）建立了一个测试运动感知和检测算法的公共测试数据库；然后，结合心理物理学实验的结果，我们推导和建立了一个新的运动感知计算模型，并对计算模型的算法进行了深入的理论分析；最后，我们在先前建立的公共数据库上测试了该运动感知算法的性能和效果，并与当前流行的运动检测模型和算法进行了比较实验，实验结果证明了该运动感知模型的优越性能，及其所具有的广泛工程应用潜力。总的来说，本文的主要贡献和创新点体现在以下几个方面：下面我们将本论文的主要工作和贡献总结如下：

在第二章中，我们采集和建立了一套新的运动感知测试数据库。这套数据库由20个不同自然场景中的视频片断组成，视频中的运动物体种类多样，共有11个受试者独立对其中的运动物体进行标定。随后，我们对11个受试者的标定数据进行了数值统计分析，计算出了人在自然场景中运动感知所存在的个体差异性及其一致性的大小。更进一步，我们以11个受试者的标定数据作为基准（benchmark)，结合改良后的评估准侧,建立了一套新的用于对运动检测机器算法进行测试和比较的公共平台。相对于现有的运动检测算法比较的数据库，我们的运动感知测试平台有如下优势和特点：第一，数据库中的视频片段均是自然场景，自然场景的种类多样，并且，多种运动物体的种类也多样化，包含开动的汽车、自行车、行人以及运动员等，运动检测算法的泛化能力（generalization）和鲁棒性（robustness）可以得到很好的测试；其二，由于数据库中的视频都是由移动中的摄像机拍摄，这去掉了”摄像机固定”的假设条件，使得开发的新算法能够在更广泛的范围下正常工作；第三，我们的数据库中的视频由多人标定，我们对标定数据的稳定性进行了测试，并将标定数据的个体差异性和一致性纳入到设计测试准则之中，使得在这个数据库上测试出的结果能更加客观和准确。我们相信，这个公开的运动检测算法测试平台将对相关算法的研究和开发有推动性作用。

在第三章中，结合建立运动感知数据库的实验过程中所得到的定性结果，我们通过两套不同的理论推导和分析方法，建立了新型的运动感知计算模型。这个运动感知计算模型的核心是利用傅里叶空间的相位差成分来提取和表征自然场景中物体的运动成分。这种表征可以有效地绕开运动物体本身的外形复杂度而专一性地提取物体的运动信息，与人类两条视觉通络中”what”和”where”之间的分离的机制有着概念上的联系。物体运动信息在我们的运动感知计算模型中能简明地表征出来，不仅使得模型本身的复杂度很低，也减少了模型中暗含的假设，使得模型更具鲁棒性，也避免了over-fitting的问题。最后，这个新型的运动感知模型的算法由9行MATLAB代码简明地实现出来。

在第四章中，我们通过大量实验对这个运动感知模型进行了测试和比较。在与同类运动检测算法的比较实验中，我们的运动感知模型算法取得了较大的优势。一系列的实验说明，该运动感知算法的良好的稳定性、鲁棒性，以及极快的计算速度（普通笔记本上达到70fps），使得它具有了广泛的工程应用前景。

在第五章中，我们将运动感知计算模型的算法移植到了ARM9嵌入式平台上进行了实现，算法在嵌入式Linux系统中运行良好，所需硬件环境极低，表现出了在工程系统中的应用潜力。随后，为了更进一步说明运动感知计算模型对大脑神经生理计算机制的契合，我们利用运动感知模型拟合了视觉注意力和眼动的生理学实验数据，在与同类的神经生理学计算模

型的评测中，我们的运动感知计算模型的拟合结果最准确，这说明了运动感知模型不仅可以应用到工程问题之中，还能对神经生理学现象进行拟合和预测，这对揭示人类视觉系统计算机制有着积极的意义。

在以后的工作中，一方面我们计划对更有效的运动感知计算模型进行研究，提出新的运动表征子，并应用到如运动检测、物体识别和跟踪等计算机视觉应用和问题之中去。另一方面，我们将从系统层次的角度建立视频信号数据挖掘的框架，使得数据挖掘技术和模式识别技术和算法能够有效地整合到视频分析之中去，为视觉监控系统、智能交通系统的建立提供技术框架和解决方案。

参考文献

[1] MARR D. Vision: A Computational Investigation into the Human Representation and Processing of Visual Information[J]. 1982.

[2] PERLS F, STEVENS J. Gestalt therapy verbatim[M].[S.l.]: Real People Press Lafayette, CA, 1969.

[3] PALMER S. Vision science: Photons to phenomenology[M].[S.l.]: MIT press Cambridge, MA., 1999.

[4] SIMONCELLI E, OLSHAUSEN B. Natural image statistics and neural representation[J]. Annual review of neuroscience, 2001, 24(1):1193–1216.

[5] STAUFFER C, GRIMSON W. Adaptive background mixture models for real-time tracking[C]:Proc. IEEE Conf. on Computer Vision and Pattern Recognition. .[S.l.]: [s.n.] , 1999，2:246–252.

[6] MITTAL A, PARAGIOS N. Motion-based background subtraction using adaptive kernel density estimation[C]:Proc. IEEE Conf. on Computer Vision and Pattern Recognition. .[S.l.]: [s.n.] , 2004，2.

[7] CHEUNG S, KAMATH C. Robust techniques for background subtraction in urban traffic video[J]. Video Communications and Image Processing, SPIE Electronic Imaging, 2004, 5308:881–892.

[8] TIAN T, TOMASI C, HEEGER D. Comparison of approaches to egomotion computation[C]:Proc. IEEE Conf. on Computer Vision and Pattern Recognition. .[S.l.]: [s.n.] , 1996:315–320.

[9] HAN M, KANADE T. Reconstruction of a scene with multiple linearly moving objects[J]. International Journal of Computer Vision, 2004, 59(3):285–300.

[10] IRANI M, ANANDAN P. A unified approach to moving object detection in 2D and 3D scenes[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(6):577–589.

[11] VIOLA P, JONES M. Robust real-time face detection[J]. International Journal of Computer Vision, 2004, 57(2):137–154.

[12] DOLLÁR P, WOJEK C, SCHIELE B, et al. Pedestrian detection: A benchmark[C]:Proc. IEEE Conf. on Computer Vision and Pattern Recognition. .[S.l.]: [s.n.] , 2009:304–311.

[13] ITTI L, KOCH C, NIEBUR E, et al. A model of saliency-based visual attention for rapid scene analysis[J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 1998, 20(11):1254–1259.

[14] HOU X, ZHANG L. Dynamic Visual Attention: Searching for coding length increments[J]. Advances in Neural Information Processing Systems, 2008, 21:681–688.

[15] ITTI L, BALDI P. Bayesian Surprise Attracts Human Attention[J]. 2006:547–554.

[16] COCHIN S, BARTHELEMY C, LEJEUNE B, et al. Perception of motion and qEEG activity in human adults[J]. Electroencephalography and Clinical Neurophysiology, 1998, 107(4):287–295.

[17] TORRALBA A, MURPHY K, FREEMAN W, et al. Context-based vision system for place and object recognition[C]:Proceedings of the Ninth IEEE International Conference on Computer Vision. .[S.l.]: [s.n.] , 2003:273.

[18] http://ftp.pets.rdg.ac.uk

[19] http://homepages.inf.ed.ac.uk/rbf/CAVIAR/

[20] BASHIR F, PORIKLI F. Performance Evaluation of Object Detection and Tracking Systems[C]//IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS2006). .[S.l.]: [s.n.] , 2006.

[21] LIST T, BINS J, VAZQUEZ J, et al. Performance Evaluating the Evaluator[C]:Proc. IEEE Joint Workshop on Visual Surveillance and Performance Analysis of Video Surveillance and Tracking. .[S.l.]: [s.n.] , 2005.

[22] VERNON D. Fourier vision: segmentation and velocity measurement using the Fourier transform[M].[S.l.]: Kluwer Academic Publishers, 2001.

[23] BLACK M, ANANDAN P. A framework for the robust estimation of optical flow[C]:Proc. IEEE Conf. on International Conference of Computer Vision. .[S.l.]: [s.n.] , 1993:231–236.

[24] MALLAT S. A wavelet tour of signal processing[M].[S.l.]: Academic Press, 1999: 37–38.

[25] GRIMMETT G, STIRZAKER D. Probability and random processes[M].[S.l.]: Oxford University Press, USA, 2001: 194–195.

[26] SONKA M, HLAVAC V, BOYLE R. Image Processing, Analysis, and Machine Vision[M].[S.l.]: Cengage-Engineering, 2007. .

[27] THOMAS J, COVER T. Elements of information theory[M].[S.l.]: Wiley-Interscience, 2006.

附录一　个人简历及在学期间发表论文

A.1　个人简历

1987年4月10日出生于四川省内江市，2000年9月至2003年7月初中入读四川省内江市第六中学，2003年9月至2006年7月高中入读四川省成都市第七中学，2006年9月因获得生物竞赛赛区一等奖保送上海交通大学生命科学技术学院，入读生物医学工程系。2010年8月将入读香港中文大学信息工程系，攻读研究型硕士学位。研究兴趣包括：计算机视觉，神经生物学，计算神经学，感知网络，机器学习等。

个人电子邮箱：zhoubolei@gmail.com

A.2　发表论文

B.Zhou and L.Zhang. A hierarchical model for visual perception.
The 2rd International Conference on Cognitive Neurodynamics (ICCN), 2009, EI index

B.Zhou and L.Zhang. Scene Gist: a holistic generative model of natural image.
The 9th Asian Conference on Computer Vision (ACCV), 2009, EI index

B.Zhou and L.Zhang. A phase discrepancy analysis of object motion.
The 10th Asian Conference on Computer Vision (ACCV), 2010 (submitted)

谢辞

在本文完成之际，首先我要感谢我的导师张丽清教授，是他将我引入计算机视觉和机器学习的研究领域。本文的工作以及其他未包括在本文中的研究工作都是在张老师的精心指导下完成。张老师严谨求实的治学态度、儒雅和勤勉的学者风范都是我终身学习的榜样。在此谨向张老师表示衷心的感谢和深深的敬意。

在我开始研究性学习阶段，我曾在童善宝老师的神经工程实验室参与过脑电采集课题研究工作，童教授给予了我极大的关心和帮助，手把手指导我怎么从课程学习转变到研究型学习、如何调研相关文献以及如何提出新的研究想法，可以说，是他引领我进入科学研究的大门。借此机会对童善宝老师表示衷心感谢。

实验室良好的学习和讨论氛围。是在所有师兄的共同努力下创造的，我从中受益匪浅。感谢他们在学期间积极参与研究课题讨论，无私分享他们的研究成果、算法代码及工具，使我少走弯路，能够有效开展研究工作。特别感谢祝文俊、吴强以及李俊华学长等给予我的帮助和指导。

另外，在计算机视觉研究过程中，侯晓迪和曹阳同学一直在学术上给予他们的支持，在与他们的讨论过程中，我开阔了自己的研究思路和视野，并逐渐形成自己的研究风格和方向。感谢他们，希望他们早日完成学业，拥有美好的前程。

最后，感谢所有关心和帮助过我的老师、同学及朋友!