

本科生学士学位论文

视觉和惯性定位结合的目标追踪算法研究

Object Tracking Based On Combination of Vision and Inertial Localization

姓 名： 杜大有

学 号： 1300012870

院 系： 信息科学技术学院

专 业： 计算机科学与技术

指导老师： 梁云 特聘研究员

二〇一七年六月

## 摘 要

运动目标检测及跟踪是机器视觉领域备受关注的前沿课题之一。一方面，可以使用基于计算机视觉的方法，在摄像机采集的图像中搜索特定目标，提取目标区域，标定目标位置。另一方面，也可以使用惯性定位的方法，结合移动设备上的惯性测量单元(IMU)，从而精确地定位目标的室内位置。

然而，基于计算机视觉的定位方法通常会由于短暂无法检测到目标（如目标被遮挡，或超出摄像机范围等）而失去目标的匹配，尤其是在多目标检测时；另一方面，单纯地使用移动设备上的传感器（加速度计、地磁计、陀螺仪等）进行惯性定位，会造成由两重积分带来的累计误差。因此，本文设计并实现了一种基于卡尔曼滤波器的位移融合算法，旨在消除惯性定位的累计误差的同时，增强视觉定位的稳定性。实验表明，融合后的轨迹能与目标的实际轨迹较好地吻合；在摄像头因为某种原因暂时无法检测到目标时，惯性定位能够持续反馈出较为准确的位置，从而重新匹配目标。

本文的主要工作如下：

* 分析了当前基于计算机视觉和惯性定位的目标跟踪方法的局限性。
* 设计并实现了一种视觉和惯性定位结合的传感器融合系统。该系统通过获取目标携带的具有惯性测量单元的装置（如手机等）的传感器信息和深度摄像头检测的视觉信息，首先使用扩展卡尔曼滤波器融合校正惯性装置姿态，然后对目标进行实时的惯性定位，并利用视觉信息进行融合校正，最后将校正后的位置信息反馈给视觉信息提供者。
* 通过以人作为追踪目标，手机作为惯性数据提供者，带有深度摄像头的机器人作为视觉信息提供者和输出接收者，在多种模拟的实际情况下进行跟踪实验，验证了本文的传感器融合系统的有效性。

关键词：传感器融合，运动目标追踪，卡尔曼滤波，惯性测量单元

## Abstract

This thesis presents a moving object tracking system with a Kalman Filter-based algorithm. A software tool is developed to track a moving object in a dynamic region, possibly among other similar objects (e.g. tracking a specific person in a supermarket). A Kalman Filter-based algorithm is used to predict and update the position of the object, utilizing the data received from both the IMU, which takes along by the object, and from a depth-camera, which also gives the position information about the object through computer vision techniques. The experiment shows that the tool can track the object with high accuracy, meanwhile solves the mismatch problem (wrong match and match lost) that commonly happened in computer vision-based multiple-object tracking systems.

The contributions of this thesis are as follows.

* This thesis analyze the limitations of the current object tracking methods based on computer vision and inertial positioning.
* Design and implement a visual and inertial positioning combined sensor fusion system. The system first uses an Extended Kalman Filter (EKF) to compute the attitude and predict the position of the inertia device (such as a cell phone, etc.) by acquiring the sensor information. Then the visual information is used to update and correct the position, through a Kalman Filter process. Finally, the corrected position information is fed back to the visual information provider.
* The tool is tested through the experiments that given a human as the tracking target, the mobile phone as an inertial data provider, a robot carried with a depth-camera as a visual information provider and output receiver. A variety of practical conditions are considered in the experiments to verify the effectiveness of the system.

Key Words: Sensor Fusion, object tracking, Kalman Filter, Inertial Measurement Unit(IMU)

## 目 录

[摘 要 2](#_Toc484631200)

[Abstract 3](#_Toc484631201)

[目 录 4](#_Toc484631202)

[第一章 引言 1](#_Toc484631203)

[1.1 运动目标检测及跟踪的应用现状 1](#_Toc484631204)

[1.2 现有目标追踪算法的分析与局限 1](#_Toc484631205)

[1.2.1 基于计算机视觉的目标检测与追踪 1](#_Toc484631206)

[1.2.2 基于IMU的惯性定位 2](#_Toc484631207)

[1.3 本文的主要工作与结构 3](#_Toc484631208)

[1.3.1 本文的主要工作 3](#_Toc484631209)

[1.3.2 本文的结构 3](#_Toc484631210)

[第二章 基于9轴IMU的手机姿态测算 5](#_Toc484631211)

[2.1 总述 5](#_Toc484631212)

[2.2 手机传感器坐标定义 7](#_Toc484631213)

[2.3 9轴传感器姿态融合算法 8](#_Toc484631214)

[2.4 姿态初始化算法 10](#_Toc484631215)

[2.5 融合加速度计和地磁计的姿态测算原理 12](#_Toc484631216)

[2.5.1 建立初始姿态的公式推导 12](#_Toc484631217)

[2.5.2 使用初始旋转矩阵初始化卡尔曼滤波器的公式推导 13](#_Toc484631218)

[2.5.3 使用加速度计或者地磁计更新后验姿态公式推导 14](#_Toc484631219)

[第三章 惯性定位与视觉定位结合的位移测算 18](#_Toc484631220)

[3.1 视觉定位系统简述 18](#_Toc484631221)

[3.2 惯性定位与视觉定位融合算法 19](#_Toc484631222)

[3.3 惯性定位与视觉定位融合滤波原理 21](#_Toc484631223)

[第四章 实验与效果分析 23](#_Toc484631224)

[4.1 姿态测算实验 23](#_Toc484631225)

[4.1.1 实验方案 23](#_Toc484631226)

[4.1.2 实验结果 24](#_Toc484631227)

[4.1.3 实验结果讨论 26](#_Toc484631228)

[4.2 位移融合滤波实验 28](#_Toc484631229)

[4.2.1 实验方案 28](#_Toc484631230)

[4.2.2 实验结果与实验结果讨论 29](#_Toc484631231)

[4.3 视觉定位漏检实验 33](#_Toc484631232)

[4.3.1 实验方案 33](#_Toc484631233)

[4.3.2 实验结果与实验结果讨论 33](#_Toc484631234)

[第五章 工作总结与展望 35](#_Toc484631235)

[5.1 本文工作总结 35](#_Toc484631236)

[5.2 未来工作展望 35](#_Toc484631237)

[参考文献 37](#_Toc484631238)

[致 谢 38](#_Toc484631239)

## 第一章 引言

1.1 运动目标检测及跟踪的应用现状

运动目标检测是机器视觉领域备受关注的前沿课题之一，被广泛应用于交通、安防、机器人等领域。其中，

1.2 现有目标追踪算法的分析与局限

1.2.1 基于计算机视觉的目标检测与追踪

目标检测（object detection）是指从图像中提取出目标位置与大小的过程，其中一个重要分支为人物检测，后者被广泛应用于交通、安防、机器人领域中。常见的目标检测方法有特征提取与分类、前景提取和基于深度学习的目标检测方法等。

基于目标特征提取与分类的目标检测方法相对传统而成熟。早期阶段有基于Haar特征的人脸检测[\*\*\*][\*\*\*]算法，后续经过[\*\*\*]的改进增加了基于Haar-like特征的弱分类器和基于Adaboost方法训练的级联分类器增强了目标检测算法的速度。Dalal和Triggs提出了梯度直方图（Histogram of Gradient, HOG）作为图像检测特征并结合支持向量机（Supporting Vector Maching, SVM）进行分类极大提高了算法精度[\*\*\*]。在此基础之上，Felzenszwalb等人提出了可拆分部件模型（Deformable Part Model, DPM）[\*\*\*][\*\*\*][\*\*\*]并将其应用于人物目标检测，其对人物特征的检测准确率优于HOG方法，但是同时计算量较大，无法实现实时检测。后续有Gadeski等人[\*\*\*][\*\*\*]使用GPU对此方法进行并行加速，试图解决实时性的问题，同时Dollár[]等人提出了积分特征通道（Integral Channel Features）等概念进一步提升目标检测的精度。

基于深度学习的目标检测方法为近年兴起的技术，具有精确度高、计算量大等特点。当前广泛使用的完成物体检测任务的一种方式是使用卷积神经网络(CNN)[\*\*\*]。卷积神经网络最早出现在上世纪80年代，最初应用于数字手写识别，取得了一定的成功。然而，由于受硬件的约束，卷积神经网络的高强度计算消耗使得它很难应用到实际尺寸的目标识别任务上。Wisel和Hubel在猫视觉系统研究工作的基础上提出了简单、复杂细胞理论，设计出来一种人工神经网络[\*\*\*]，之后发展成为卷积神经网络。卷积神经网络中卷积层的滤波器是各个位置共享的，因而可以大大降低参数的规模，这对防止模型过于复杂是非常有益的，另一方面，卷积操作保持了图像的空间信息，因而特别适合于对图像进行表达。基于卷积神经网络，R-CNN[\*\*\*],SPPNET[\*\*\*],Fast-RCNN[\*\*\*]，Faster-RCNN[\*\*\*],YOLO[\*\*\*]等一系列方法被开发出来，运用到了物体检测任务中，达到了相当好的效果，这些方法在ImageNet[\*\*\*],PASCAL VOC[\*\*]等比赛中取得了相当好的效率和精确度。

人物（目标）跟踪是指获取指定人物（目标）运动轨迹的过程，可以分为基于特定对象的跟踪（Detection Free Tracking, DFT）和基于检测的跟踪（Detection Based Tracking, DBT）[\*\*\*]。前者只跟踪给定的对象（在开始时指定），后者同时跟踪图像中的所有对象。人物（目标）跟踪面临多方面的挑战，比如目标被物体遮挡，区分图像中多个目标等。需要保证目标离开视野（被遮挡或走出拍摄范围）时仍能继续追踪目标并最终使之回到视野（即解决漏检问题），和多目标检测时，目标恢复检测后与原目标相匹配关联的问题（即解决轨迹切换问题）。

在本文中使用的基于视觉的目标追踪算法为，首先使用基于DPM[\*\*\*]的人物检测算法提取特征，构建特征金字塔，星型级联分类、检测窗口融合等步骤检测图像中所有的人物，然后使用多特征近邻匹配人物匹配跟踪方法建立对象与轨迹的关联[\*\*\*]，最后筛选出需要跟踪的特定目标[\*\*\*]，所面临的主要问题是失去目标视野的漏检问题和恢复检测后的轨迹匹配问题[\*\*\*]。

1.2.2 基于IMU的惯性定位

基于惯性测量单元（Inertial Measurement Unit，IMU）的惯性定位为相对传统的室内定位方法。其主要特点有采样周期短、数据处理快、受环境影响小等优点。然而，由于其需要计算时间的累次积分得到最终定位位置，会随时间的延长产生不可忽略的累计误差[\*\*\*]。为了解决此问题，一种通用的方法是采用估计步长、计算步数，采用步长乘以步数的方法估计位移[\*\*\*][\*\*\*][\*\*\*]，另一种方法是对惯性定位进行校正。如Jimenez[]和Issac[]等人提出了零速率更新（Zero-Velocity Update, ZUPT）并结合扩展卡尔曼滤波器（Extended Kalman Filter, EKF）校正惯性定位的偏误等。但是，单纯使用IMU进行定位的方法无法完全消除其偏误，因此虽然上述方法在短距离内能取得较好的结果，但是在长距离跟踪定位时，会无可避免的出现累积的定位偏误问题。

因此，本文采用视觉定位和惯性定位结合的方式，一方面利用视觉定位长期稳定的特点消除惯性定位的累计误差，另一方面利用惯性定位短距离准确、受环境干扰低、采样稳定的特点解决视觉定位中产生的漏检和轨迹切换问题。

1.3 本文的主要工作与结构

1.3.1 本文的主要工作

本文的主要工作包括：

* 分析了当前基于计算机视觉和惯性定位的目标跟踪方法的局限性。
* 设计并实现了一种视觉和惯性定位结合的传感器融合系统。该系统通过获取目标携带的具有惯性测量单元的装置（如手机等）的传感器信息和深度摄像头检测的视觉信息，首先使用扩展卡尔曼滤波器融合校正惯性装置姿态，然后对目标进行实时的惯性定位，并利用视觉信息进行融合校正，最后将校正后的位置信息反馈给视觉信息提供者。
* 通过以人作为追踪目标，手机作为惯性数据提供者，带有深度摄像头的机器人作为视觉信息提供者和输出接收者，在多种情况下进行跟踪实验，验证了本文的传感器融合系统的有效性，能解决惯性定位的累计误差和视觉定位的漏检、轨迹切换等问题。

1.3.2 本文的结构

本文的结构如下：

第一章呈现了目标（人物）检测与追踪算法的应用现状，对基于视觉定位的目标检测追踪算法和基于惯性定位的目标检测追踪算法进行了分析，并指出了其存在的缺陷。

第二章分析了缓存的理论上限与当前性能间的差距及造成的原因，验证了实现浏览器的缓存优化是有意义而且可行的，并对可行的缓存优化手段进行了分析，提出了优化方案。

第三章基于面向对象的方法对本文的系统进行了分析与设计。

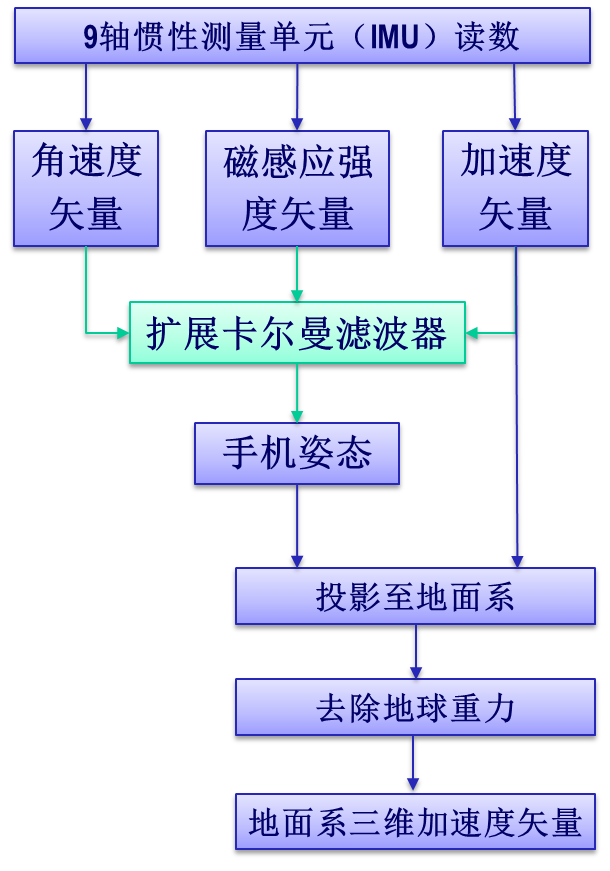
第四章对本系统进行了实验设计，并描述了自动化测试系统的设计与实现，展现了实验的结果，最后分析了本系统带来的额外开销。

第五章对全文的工作进行了回顾和总结，并提出了未来工作的三个方向。

## 第二章 基于9轴IMU的手机姿态测算

本章对基于9轴惯性测量单元（Inertial Measurement Unit, IMU）的各个部分进行了详细说明和理论推导，同时给出了相应的算法流程以在整体上说明各个部分的工作方式。

2.1 总述



图\*\*\* 手机姿态测算部分总览

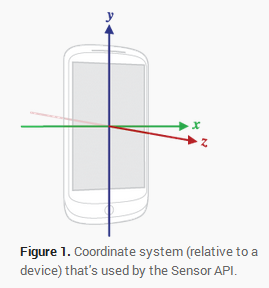
使用惯性定位计算目标物体位置，需要获得在地面系下的目标物体的加速度，然后经过两次积分得到相对于起始点的位移。这就需要获取手机的姿态，从而将手机传感器测量得到的加速度投影到地面系中（即计算得到手机三维坐标轴的旋转矩阵，然后使用姿态旋转矩阵左乘手机系三维加速度矢量，就可以得到地面系三维加速度）。

图\*\*\* 总体上展示了手机姿态测算部分的算法。从手机惯性测量单元（IMU）获取的数据共有九个维度，分为三组，分别为三维角速度矢量，三维磁感应强度矢量，三维加速度矢量（方向定义见\*\*\*）。其中，使用角速度积分结合初始姿态可以得到一组后续姿态；同时，由于地球重力场和地磁场的存在，使用加速度和磁感应强度数据也可以独立得到一组姿态。角速度计会由于陀螺仪的漂移产生累计误差，加速度计会由于人和手机相对于地球的运动产生一定误差，地磁计会由于外源磁场产生一定误差。为了使得测算的姿态尽量准确，本文中将不同方式的数据使用扩展卡尔曼滤波器进行融合，校正得到一组相对更为准确的姿态。得到手机姿态后，使用姿态旋转矩阵左乘手机系三维加速度矢量，然后减去地球重力的影响，就可以得到地面系三维加速度（纯运动加速度）。

总体而言，本文参照[\*\*\*]中的理论推导，使用四元数作为姿态的状态表示，以避免使用旋转矩阵表示造成的冗余或欧拉角表示产生的奇异点；使用扩展卡尔曼滤波器将角速度传感器数据和外源观测数据融合，消除角速度计的漂移。与[\*\*\*]中不同的是，本文中用以融合的数据为加速度计和地磁传感器数据，而不是恒星计的数据。比较而言，由于基于加速度计和地磁传感器的姿态测算原理利用静止状态下的地球重力场和地磁场，因此在手机运动时产生的误差更大，得到较为精确姿态的难度更高。

下文中将略去共性较大的部分，即姿态的四元数表示、四元数的相关运算、陀螺仪的误差模型、卡尔曼滤波器的状态方程和预测方程等，着重介绍与其不同的部分。具体而言，将在\*\*\* 介绍手机惯性测量单元得到的数据坐标定义，在\*\*\*介绍姿态初始化算法，在\*\*\* 介绍角速度计、加速度计、地磁计的姿态融合算法，在\*\*\* 对融合加速度计、地磁计的部分进行公式推导（即初始姿态求解过程、扩展卡尔曼滤波器的初始化过程，扩展卡尔曼滤波器的更新过程）。扩展卡尔曼滤波器的预测方程在\*\*\* 中有详细介绍，这里不再赘述。

2.2 手机传感器坐标定义



图\*\*\*手机传感器坐标轴图示

本文中一共读取了三种手机传感器数据：平动加速度计，角速度计和地磁传感器。这三类传感器数据均有三个维度，其对应定义如下：

平动加速度计：如图\*\*\*所示，其x坐标平行于手机短边，面向手机朝向右侧为正；其y坐标平行于手机长边，面向手机朝向上方为正；其z坐标穿过手机屏幕向外为正。

地磁传感器：和平动加速度计坐标定义相同。如图\*\*\*所示，其x坐标平行于手机短边，面向手机朝向右侧为正；其y坐标平行于手机长边，面向手机朝向上方为正；其z坐标穿过手机屏幕向外为正。

角速度计：绕平动加速度计x轴旋转为第一坐标（Pitch），顺时针旋转为正；绕平动加速度计y轴旋转为第二坐标（Roll），顺时针旋转为正；绕平动加速度计z轴旋转为第三坐标（Azimuth），顺时针旋转为正。

2.3 9轴传感器姿态融合算法

否

姿态初始化

读入传感器数据

使用角速度计数据更新先验姿态

计算线性加速度模长和磁感应强度模长

加速度模长小于阈值？

地磁计模长小于阈值？

使用加速计增量更新姿态

使用地磁计增量更新姿态

否

是

是

否

一次姿态计算循环

图\*\*\* 姿态融合算法流程图

如图\*\*\*所示，姿态融合算法将在开始时进行初始化，并在之后的时间内不断进行姿态计算循环。

姿态初始化：

在初始化期间采集一定时间的数据（约2s左右），并认为此时间内手机静止，主要目的是使用加速度计和地磁计计算初始姿态，并且消除角速度计的初始漂移。这一部分将在\*\*\*进行详细介绍。

姿态计算循环：

首先，在读取传感器数据后，对角速度进行积分得到手机先验姿态，并计算其相应的协方差矩阵，以同其余传感器进行扩展卡尔曼滤波消除陀螺仪漂移。

然后，计算加速度模长和磁感应强度模长。其理由是加速度和地磁计姿态解算的原理为认为手机近似静止，利用地球的重力场和地磁场来构建三维笛卡尔坐标系从而解算姿态（详见\*\*\*），而如果手机不在匀速运动（即带有加速度）或者外部有磁场干扰的条件下，姿态计算误差就会变大（见实验\*\*\*）。因此，本文中为加速度模长和磁感应强度模长分别设定了一个阈值。当真实读数小于对应的阈值时，认为其相对于地球重力场或磁力场的影响很小，可以使用相应读数来更新姿态；否则，本次循环中不使用该组数据来更新姿态。

每完成一个姿态计算循环后，等待下一组传感器信号输入，进行下一次循环。该扩展卡尔曼滤波器的计算量较小，其计算时间远小于传感器信号采样周期（约1~2ms），完全可以达到实时性的要求。

2.4 姿态初始化算法

读入传感器数据

分别累计角速度向量、加速度向量、地磁计向量

达到设定的初始化时长？

记录初始linux时间戳

对角速度计累计值取平均，存储作为陀螺仪偏移

对加速度累计值和地磁计累计值取平均

根据加速度向量和地磁计向量平均值计算初始姿态

是

否

初始数据采集阶段

初始姿态计算阶段

图\*\*\* 姿态初始化算法流程图

如图\*\*\*所示，姿态初始化算法分为初始数据采集阶段和初始姿态计算阶段。

在初始数据采集阶段，程序分别累计设定时间内（2秒）的角速度向量，加速度向量和地磁计向量。以试验中约700HZ计算，可以累计约1400组数据，并记录起始linux时间戳和结束linux时间戳。在此期间，假设手机完全静止，其姿态不发生变化。

在初始姿态计算阶段，首先，对角速度计累计值取平均，存储作为陀螺仪偏移。在之后使用任何陀螺仪读数前，均减去此偏移。然后对加速度累计值和地磁计累计值取平均，并根据加速度向量和地磁计向量平均值计算初始姿态（公式推导见\*\*\*），并初始化相应的协方差矩阵。在之后的姿态预测和更新过程中，此姿态（转换为四元数表示）作为初始值进入扩展卡尔曼滤波算法，使用角速度数据进行预测，使用加速度计和地磁计读数进行更新。

2.5 融合加速度计和地磁计的姿态测算原理

加速度计读数和地磁计读数有两部分作用：建立初始姿态和在卡尔曼滤波器中更新后验姿态。下面将分别对建立初始姿态过程、初始化卡尔曼滤波器过程和使用加速度计和地磁计读数更新后验姿态过程进行公式推导

2.5.1 建立初始姿态的公式推导

输入：初始数据采集阶段的加速计读数均值，地磁计读数均值，它们均为手机自然坐标系中的三维列向量，方向和正负见\*\*\*中定义。

输出：初始旋转矩阵

1. 对和进行归一化，即
2. 计算旋转矩阵
3. 验证旋转矩阵

令地球重力场单位向量为（地面系下）

令地磁场单位向量为（地面系下）

由旋转矩阵定义，将地球重力场和地磁场向量转换至手机自然坐标系有

即理论上等于手机测量值

2.5.2 使用初始旋转矩阵初始化卡尔曼滤波器的公式推导

输入：初始旋转矩阵，陀螺仪（角速度计）采样周期

参数：陀螺仪方差（variance of the gyro's output per Hz），陀螺仪偏误方差

输出：扩展卡尔曼滤波器初始状态量（姿态四元数*q*）和噪声协方差矩阵

1. 将旋转矩阵表示转换为四元数表示
2. 计算矩阵Q

1. 计算噪声协方差矩阵

令

则噪声协方差矩阵为

2.5.3 使用加速度计或者地磁计更新后验姿态公式推导

输入：某一时间点加速计读数，地磁计读数，先验姿态（四元数表示）和协方差矩阵*P*（*P*为2x2的3x3矩阵）；

参数：加速计测量偏误方差，地磁计测量偏误方差；

输出：更新后的后验姿态和协方差矩阵；

* 融合加速度计读数：

1. 验证加速度计模长不超出限定范围（保证融合准确性）
2. 将三维加速计读数归一化
3. 计算当前（先验）姿态中加速度向量
4. 计算感应矩阵
5. 计算卡尔曼增益
6. 更新姿态和协方差矩阵
7. 四元数归一化

* 融合地磁计计读数：

1. 检验磁感应强度测量矢量模长不超出限定范围（保证融合准确性）
2. 将三维磁感应强度读数归一化
3. 计算当前（先验）姿态中磁感应强度向量
4. 计算感应矩阵
5. 计算卡尔曼增益
6. 更新姿态和协方差矩阵
7. 四元数归一化

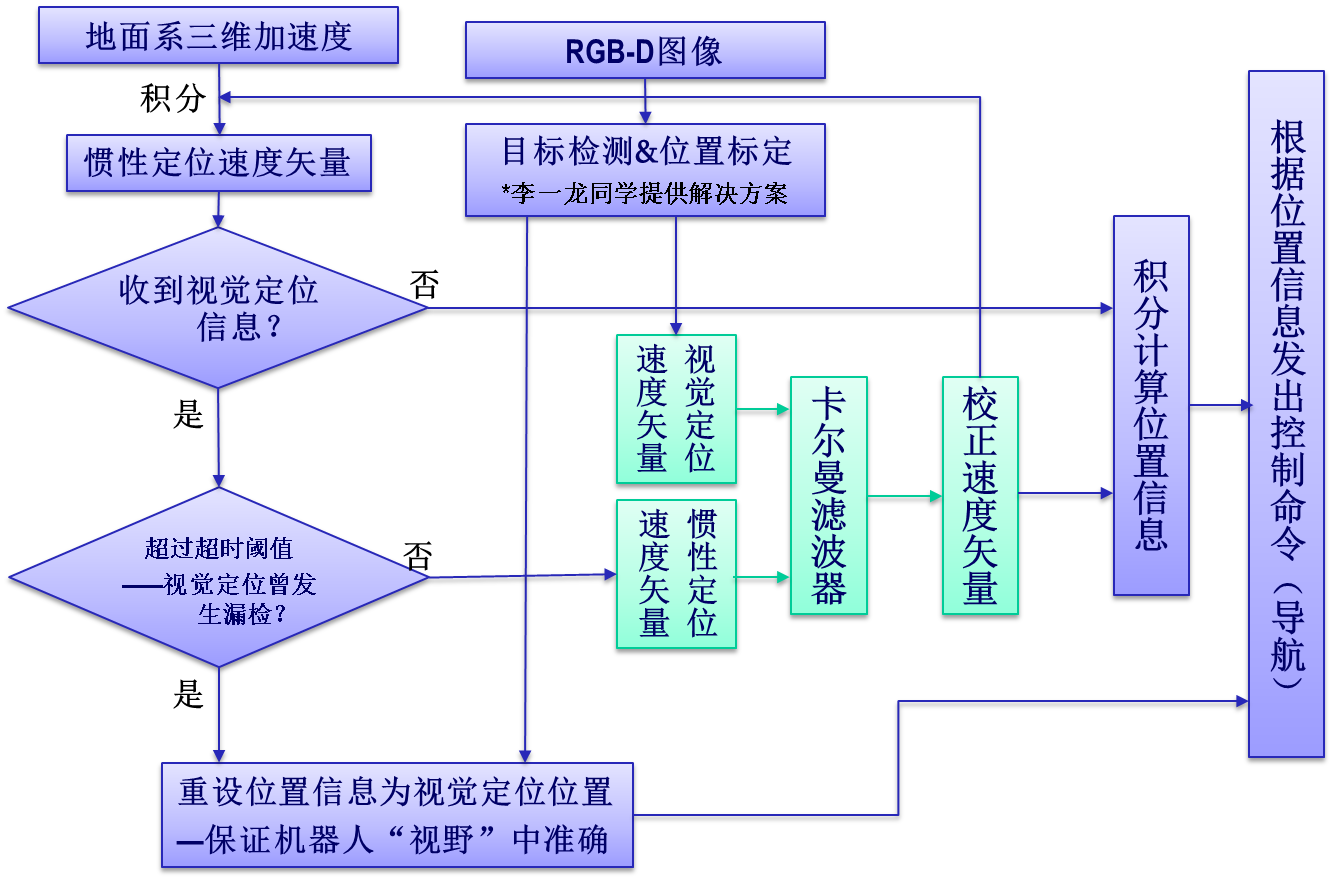
## 第三章 惯性定位与视觉定位结合的位移测算

本章首先简要介绍了视觉定位系统的原理和过程，然后详细介绍了惯性定位与视觉定位融合的算法流程以在整体上说明系统的运行过程，最后对其中应用的卡尔曼滤波器融合原理给出了理论推导。

3.1 视觉定位系统简述

本文采用了[\*\*\*]中提供的视觉定位系统

3.2 惯性定位与视觉定位融合算法



图\*\*\* 惯性定位与视觉定位融合部分总览

本节结合图\*\*\* 详细描述了惯性定位与视觉定位融合算法。

1. 首先，通过惯性测量单元得到的三维加速度投影到地面系，得到地面系三维加速度，然后进行积分，得到惯性定位速度矢量。
2. 一方面由于IMU采样周期比视觉定位采样周期短得多，另一方面也可能由于视觉定位漏测，在完成一个周期的姿态测算和速度计算后，不一定会接收到视觉定位数据。这时，如果没有接收到视觉定位数据，则直接由惯性定位速度矢量积分得到更新后的位置信息。因此，即使视觉定位发生漏检，此系统也依然可以实时反馈可供参考的位置信息，以此进行机器人导航和视觉定位重新匹配。
3. 如果接收到了视觉定位信息，则判断是否超过设定的超时阈值。超时阈值用来判定视觉定位是否曾经发生过漏检。举例而言，对于每秒30帧的拍摄视频，视觉定位信号周期应为30ms左右；此时，若设超时阈值为100ms，则两次接收到视觉定位信息超过超时阈值意味着曾经必定发生了漏检而不是数据传输问题（通常而言，发生漏检的造成的间隔一般在0.5s~2s之间）；另一方面，即使只有一到两帧漏检，其视觉定位数据仍然可用，因此本文中设定的超时阈值约为检测周期的三倍左右。
4. 如果两次同步时间超过超时阈值，认为曾经发生漏检，然后经过纯惯性导航追踪目标，视觉定位重新匹配到了目标。这时需要重新校正设定位置信息为视觉定位的位置坐标（因为此时一般视觉定位和纯惯性定位会产生一定差异），这是为了保证目标在摄像机中的位置相对准确。
5. 如果没有超过超时阈值，这时可以正常使用视觉信号进行融合校正速度信息和位置信息，以消除惯性定位中的累计误差。如图所示，在获取了三维惯性定位速度矢量和三维视觉定位速度矢量（若在平面上运动，则垂直于地面的坐标轴设置为零），使用一个卡尔曼滤波器融合后得到一个校正的速度矢量，然后进行积分得到位置信息。同时反馈校正速度矢量到下一轮的速度积分计算中，以修正惯性定位的累计误差。
6. 最后，根据反馈的位置信息，采用导航算法控制机器人向目标方向运动。

3.3 惯性定位与视觉定位融合滤波原理

本小节对使用卡尔曼滤波器将惯性定位和视觉定位进行融合的过程进行理论推导

1. 对于惯性定位过程建模，有

这里*s*为地面系中的三维位移矢量，v为地面系中的三维速度矢量，*t*为时间，为两次采样时间间隔，为IMU采样的三维加速度投影到地面系后减去地球重力影响得到的地面系三维加速度

1. 转换为矩阵形式，有
2. 使用x表示状态向量，A表示状态转移矩阵，B表示控制输入矩阵，w表示惯性传感器噪声（根据卡尔曼滤波器要求，假定为高斯噪声），有
3. 转换为离散形式，有
4. 建立视觉定位测量模型，其中为测量结果，为观测矩阵（这里根据摄像机相对于原始地面系转过的角度给出坐标系旋转矩阵），x为真实状态量（即三维位移和三维加速度），v为噪声（根据卡尔曼滤波器要求，假定为高斯噪声）。中的速度矢量由位移矢量除以时间得出。
5. 设定噪声参数
6. 卡尔曼滤波器预测阶段，计算观测前的先验值
7. 卡尔曼滤波器预测阶段，计算最小均方误差矩阵
8. 卡尔曼滤波器更新阶段，计算卡尔曼增益
9. 卡尔曼滤波器更新阶段，计算融合修正后的状态量
10. 卡尔曼滤波器更新阶段，计算融合后更新的最小均方误差矩阵
11. 反复执行步骤（7）至（11），每次更新状态向量和对应的最小均方误差矩阵

## 第四章 实验与效果分析

在本章中，为了验证融合系统的实际运行效果，并检验解决视觉定位中的漏检、失配问题和惯性定位中的累计误差问题：首先，本文使用手机在各种运动情况下的实测数据验证姿态还原效果；然后，与机器人一起进行跟踪实验，结合手机数据和视觉数据验证轨迹还原效果，同时检验消除累计误差；最后，在进行跟踪实验时模拟漏检的情形，以验证在一定时间内漏检仍能较好地恢复匹配和还原轨迹。

4.1 姿态测算实验

4.1.1 实验方案

在本节的实验中，采取以下方案：首先将手机平放于桌面上一定时间，然后绕某一固定轴做转动（垂直于手机屏幕长边、垂直于手机屏幕短边、垂直于手机屏幕向内），分别对应于\*\*\*所述的xyz轴，期间尽量保证其余两轴相对稳定。在进行姿态测算后，分别输出对应于绕xyz轴转动的旋转角Pitch、Roll、Azimuth。在初始时，手机摆放方位为：z轴指向地心，x轴指向北，y轴指向西。

下文共包含四组实验：前三组分别为绕xyz轴转动，输出在使用不同融合方式下对应的Pitch、Roll、Azimuth旋转角关于时间的曲线。最后一组将手机放于汽车内进行实验，（由于汽车外部为铁壳）以模拟外磁场的干扰情况。

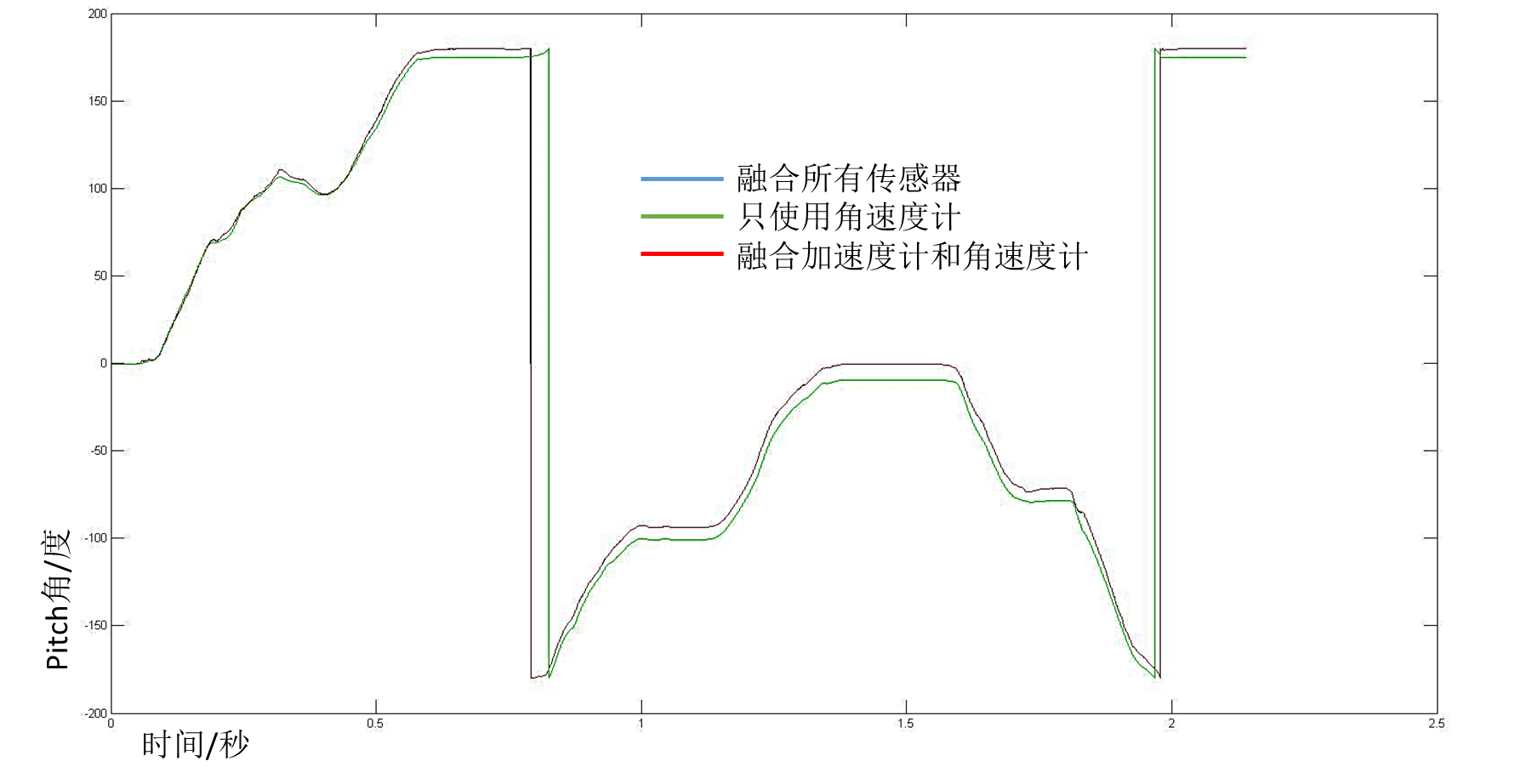
文中使用的测试用手机为HUAWEI P9 [\*\*\*]

这四组实验旨在说明：

1. 在无较强干扰时，在不同的融合方式下和不同的运动方式下，姿态还原均较为准确；
2. 在运动过程中，对应方向上有加速度会对融合结果有微小影响；
3. 在运动过程中，外在强磁场干扰会对融合结果产生较大影响，而这可以通过设定融合读数阈值范围去除。

4.1.2 实验结果

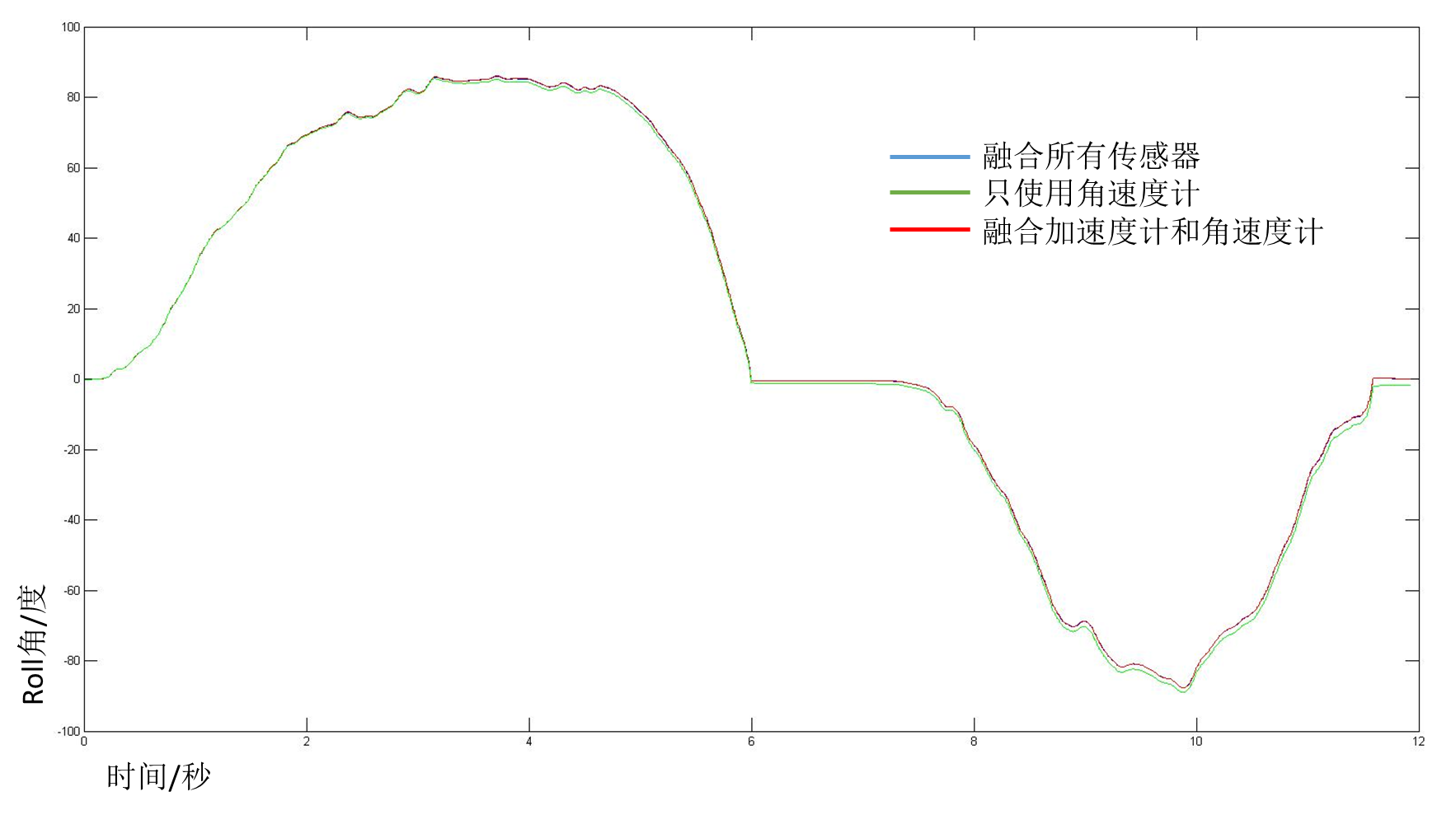
1. 沿x轴（轴沿南北方向，旋转改变Pitch角）旋转



图\*\*\* 手机沿x轴做旋转运动姿态还原效果

图\*\*\* 展示了手机沿x轴（轴沿南北方向）运动时，姿态还原出的Pitch角随时间变化曲线。其中，融合所有传感器（角速度计、加速度计、地磁传感器）的曲线与融合加速度计和角速度计的曲线重合（与实际相比相对准确），而只使用角速度计则有微小偏移。

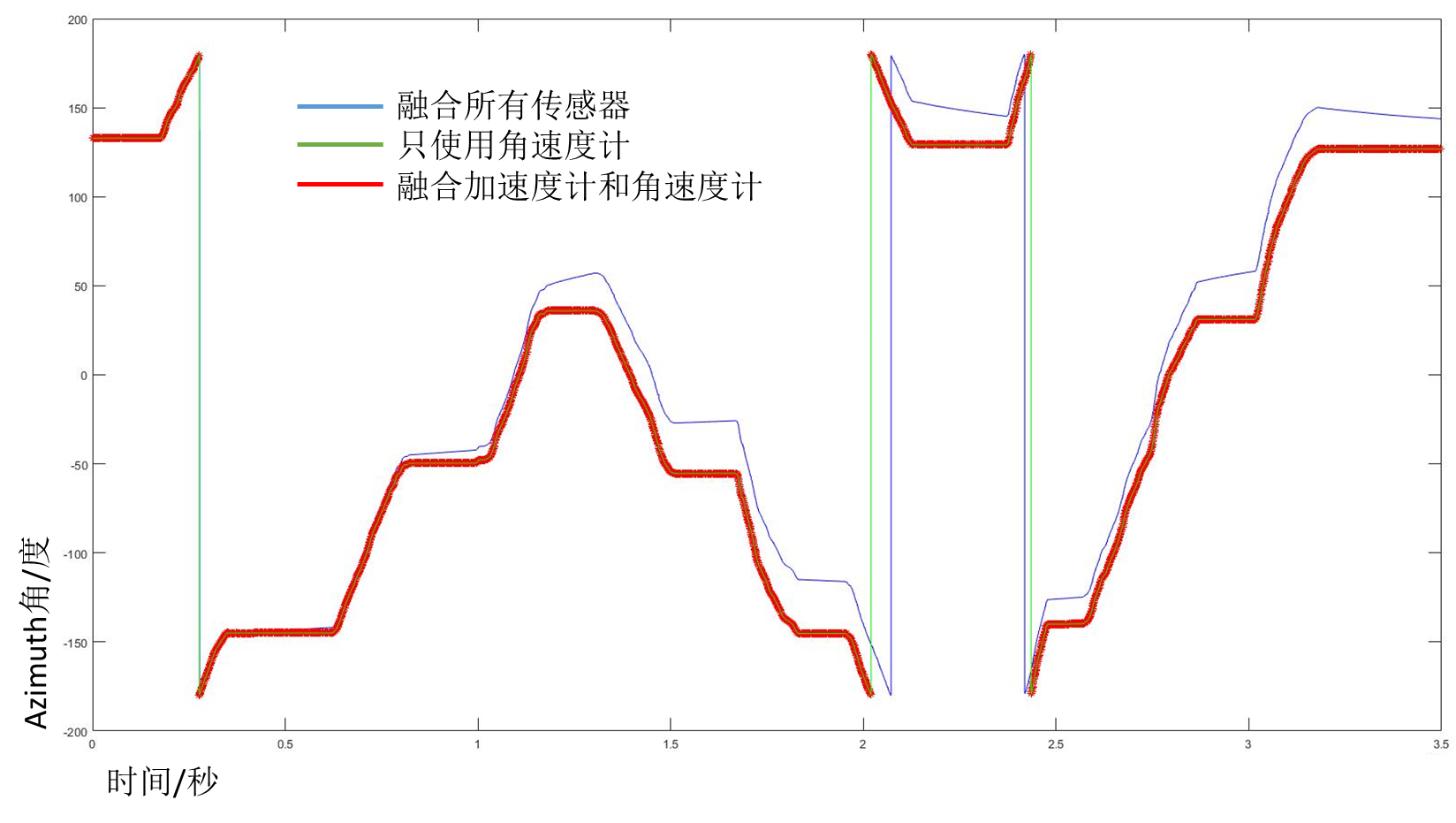
1. 沿y轴（轴沿东西方向，旋转改变Roll角）旋转



图\*\*\*手机沿y轴做旋转运动姿态还原效果

图\*\*\* 展示了手机沿y轴（轴沿东西方向）运动时，姿态还原出的Roll角随时间变化曲线。其中，融合所有传感器（角速度计、加速度计、地磁传感器）的曲线、融合加速度计和角速度计的曲线、只使用角速度计的曲线均大部分重合，仅有微小偏差（与实际相比相对准确）。

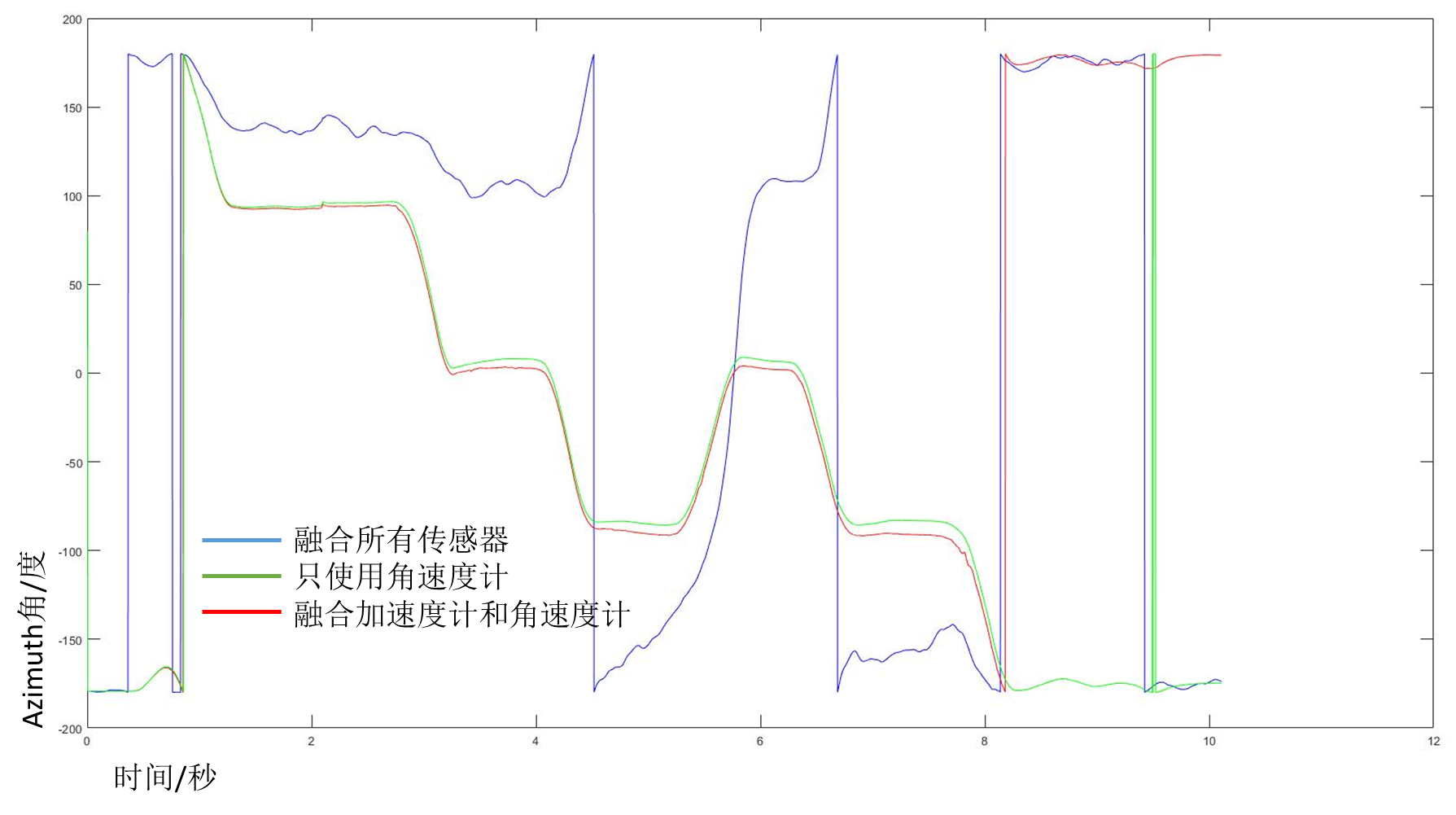
1. 沿z轴（轴指向地心，旋转改变Azimuth角）旋转



图\*\*\* 手机沿z轴做旋转运动姿态还原效果

图\*\*\* 展示了手机沿z轴（轴指向地心）运动时，姿态还原出的Roll角随时间变化曲线。其中，融合所有传感器（角速度计、加速度计、地磁传感器）的曲线有微小偏差，融合加速度计和角速度计的曲线、只使用角速度计的曲线均大部分重合（与实际相比相对准确）。

1. 强磁场干扰下沿z轴（轴指向地心，旋转改变Azimuth角）旋转



图\*\*\* 强磁场干扰下手机沿z轴做旋转运动姿态还原效果

图\*\*\* 展示了在强磁场干扰下（手机在汽车内）手机沿z轴（轴指向地心）运动时，姿态还原出的Roll角随时间变化曲线。由于汽车外壳的作用，地磁计无法正确地读取地球磁感应强度数据。其中，融合所有传感器（角速度计、加速度计、地磁传感器）的曲线有较大偏差，融合加速度计和角速度计的曲线、只使用角速度计的曲线均大部分重合（与实际相比相对准确）。

4.1.3 实验结果讨论

对于各组实验中，部分曲线产生误差的解释：

* 前两组实验中，融合所有传感器（角速度计、加速度计、地磁传感器）的曲线与融合加速度计和角速度计的曲线重合（与实际相比相对准确），而只使用角速度计则有微小偏移。此偏移是由于手机在旋转过程中地球重力在各个轴上的分布有较大分布变化，而没有使用加速度计进行重新校正，因此产生偏移。
* 第三组实验中，融合所有传感器（角速度计、加速度计、地磁传感器）的曲线有微小偏差，融合加速度计和角速度计的曲线、只使用角速度计的曲线均大部分重合（与实际相比相对准确）。此偏移是由于手机在旋转过程中，地磁场在各个轴上的分布有较大变化，而地磁计本身在运动时准确性较差，因此产生偏移。
* 第四组实验中，融合所有传感器（角速度计、加速度计、地磁传感器）的曲线有较大偏差，融合加速度计和角速度计的曲线、只使用角速度计的曲线均大部分重合（与实际相比相对准确）。此偏移是由于手机在旋转过程中，地磁场由于汽车外部铁壳的存在被屏蔽，因此实际上等价于存在强外加磁场干扰。与此同时，地磁计本身在运动时准确性较差，因此融合地磁计数据会极大地干扰姿态融合的准确性。

偏差的解决方法如\*\*\*节所述，在姿态融合阶段，会对加速度计数据和地磁计数据的模长进行判断，即在平动加速度较大时（干扰加速度计数据）不使用加速度计数据进行融合，在外磁场干扰较大时（干扰地磁计数据）不使用地磁计数据进行融合。在排除不准确数据后，融合的姿态就可以还原到各组中相对准确的数据曲线上。

综上所述，姿态测算部分实验表明：

1. 在无较强干扰时，在不同的融合方式下和不同的运动方式下，姿态还原均较为准确；
2. 在有较大平动加速度或者外磁场干扰时，通过设定阈值过滤数据，使得姿态解算可以尽量保持准确。

4.2 位移融合滤波实验

4.2.1 实验方案

在本节的实验中，采取以下方案：

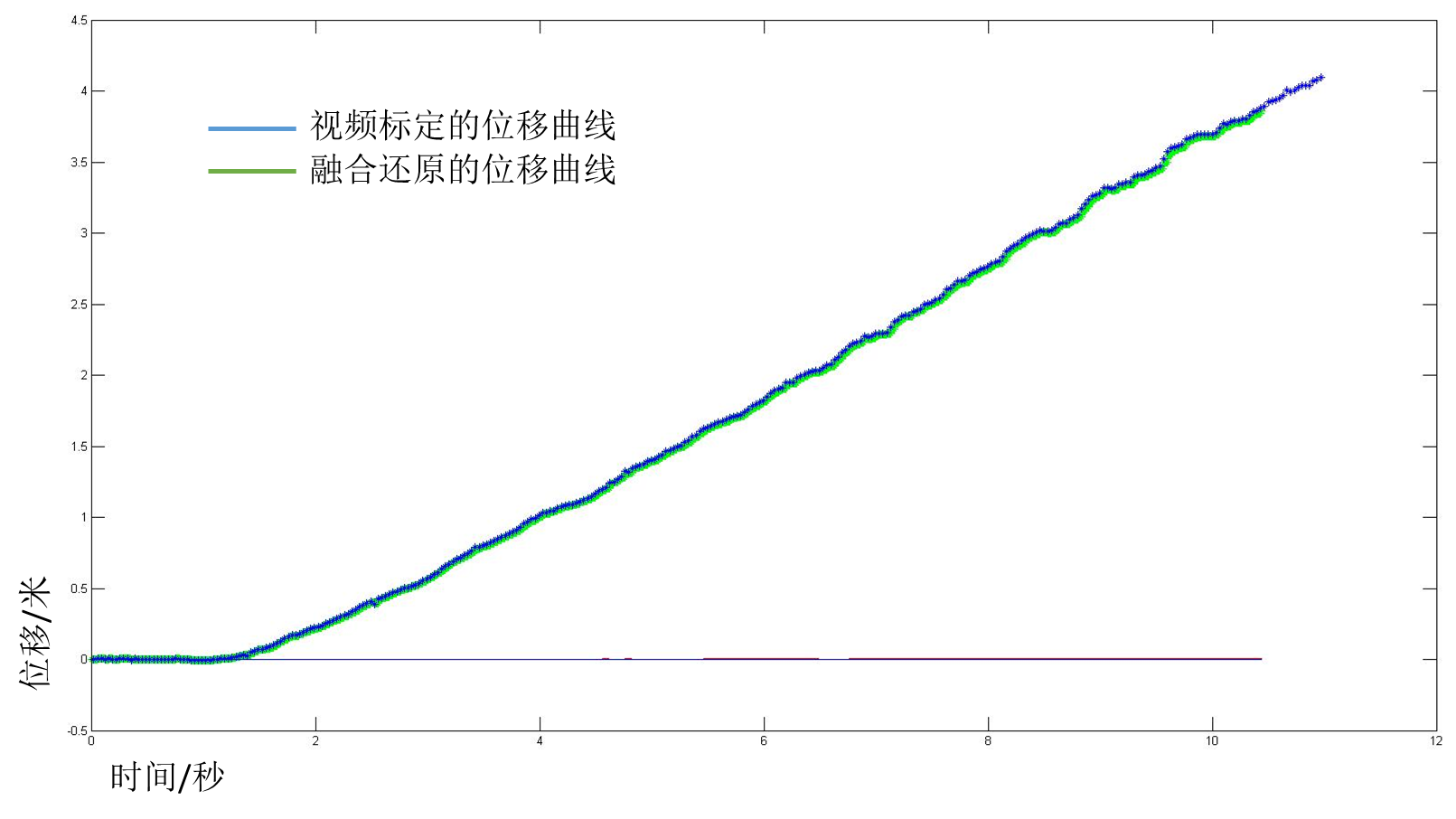
1. 人背对镜头站在机器人前方，手机持在手中或者放在口袋中，静止2秒，以采集数据完成手机数据、摄像头数据和机器人数据的初始化；
2. 人开始以事先规划好的路线正常行走，机器人跟随，同时保存摄像头拍摄的录像；
3. 跟随过程中使用手机姿态测算和位移融合实时还原人的运动轨迹；
4. 输出比较还原的运动轨迹和从视频中手动标定的运动轨迹；
5. 所用测试用手机为HUAWEI P9；
6. 使用的测试用机器人为美国iRobot公司的轮式机器人平台iRobot Create 2（如图\*\*\*）, 图\*\*\* 为iRobot 机器人的基本结构，其主要由左右轮的差分驱动进行控制, 可以实现原地转向，或按指定转弯半径前进、后退等功能。iRobot提供了公开的串口编程接口, 包括电机底层控制、左右轮速度控制、半径–速度控制等[\*\*\*]；
7. 为了实现视觉跟踪，在机器人上方安装了ASUS Xtion Pro Live 摄像头并调整其高度至正常拍摄包含人物的场景。此摄像头可以同时采集配准、同步的RGB与深度图像。

在本节中，一共完成如下三组实验：

1. 简单直线运动运动情况，人沿镜头方向直线行进一段时间后静止，机器人无需转向只需前进，对手机持于手上（每一步震动较小）和手机放在口袋中（每一步震动较大）分别测试一组数据，比较还原轨迹和标定的真实运动轨迹；
2. 直线运动情况，人与镜头方向有一定夹角直线行进一段时间后静止，机器人需转向，手机放在口袋中，比较还原轨迹和标定的真实运动轨迹；
3. 复杂运动情况，人与镜头方向有一定夹角开始运动，完成较为复杂的轨迹后停止，机器人需持续跟随，手机放在口袋中，比较还原轨迹和标定的真实运动轨迹。

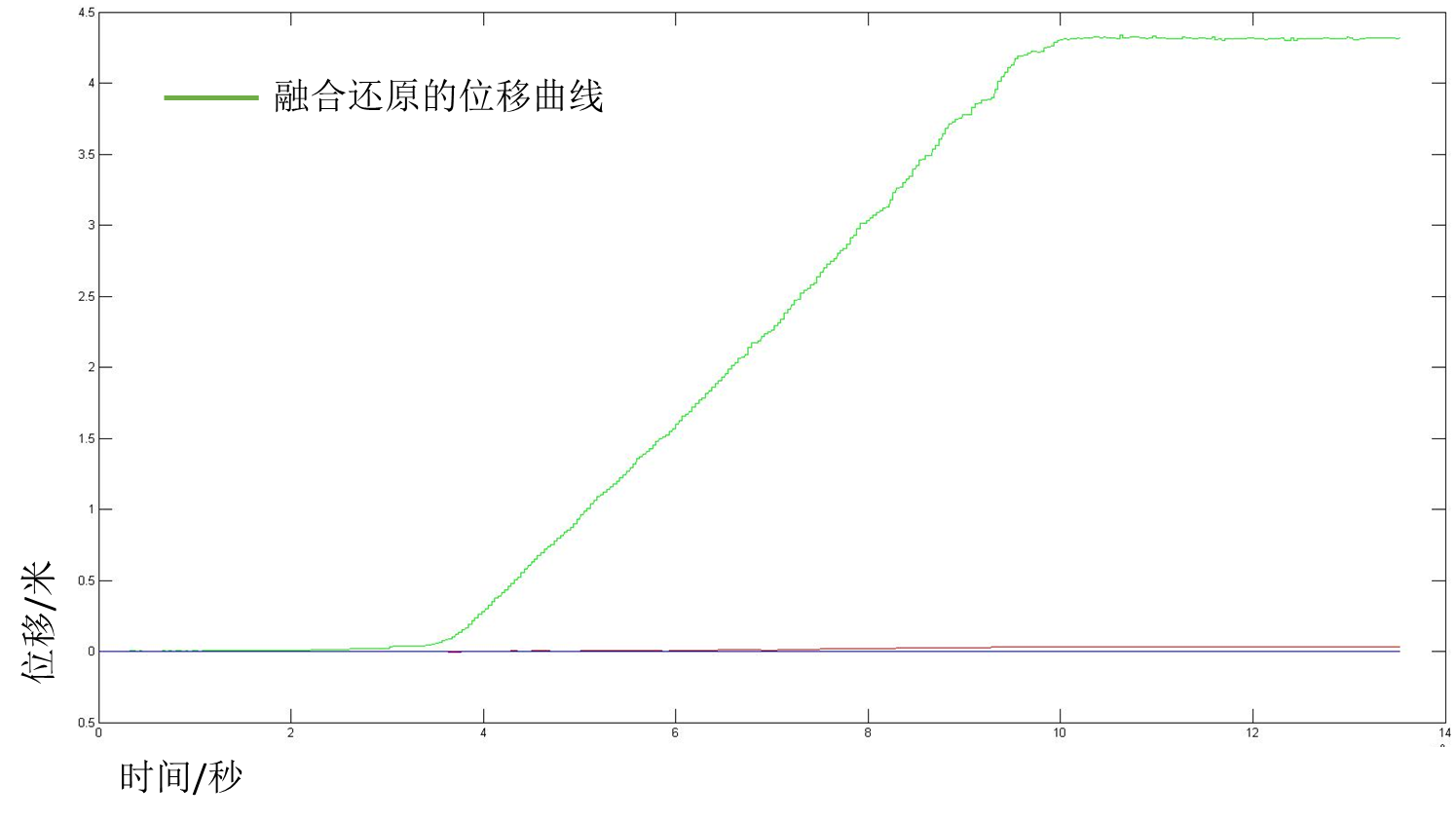
4.2.2 实验结果与实验结果讨论

1. 简单直线运动运动情况



图\*\*\* 简单直线运动情况（手机持于手上）还原的位移曲线

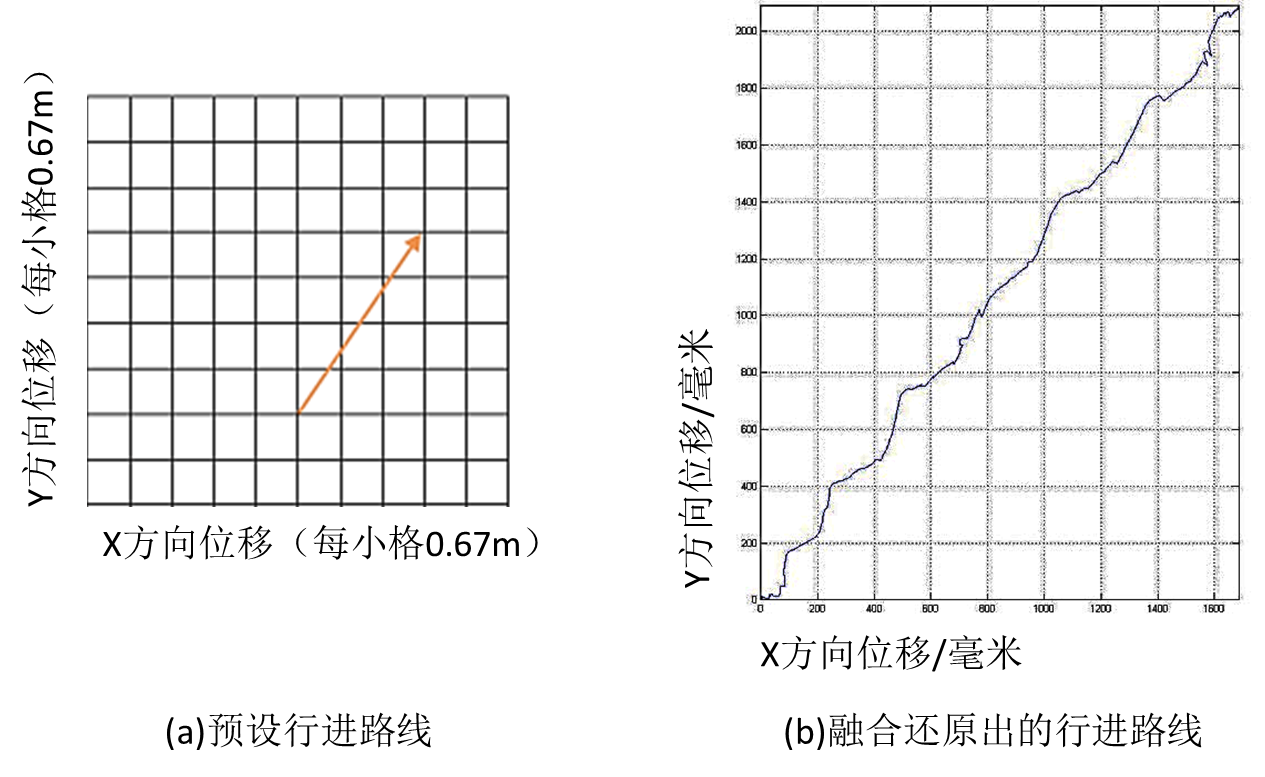
图\*\*\* 展示了简单直线运动运动情况，其中人沿镜头方向直线行进一段时间后静止，机器人无需转向只需前进，手机持于手上（每一步震动较小），可以看出，还原出的运动路径与真实运动路径相当吻合



图\*\*\* 简单直线运动情况（手机放在口袋中）还原的位移曲线

图\*\*\* 展示了手机放于口袋中（每一步震动较大）时，进行简单直线运动的效果，可以看出，还原出的运动路径与真实运动路径相当吻合。

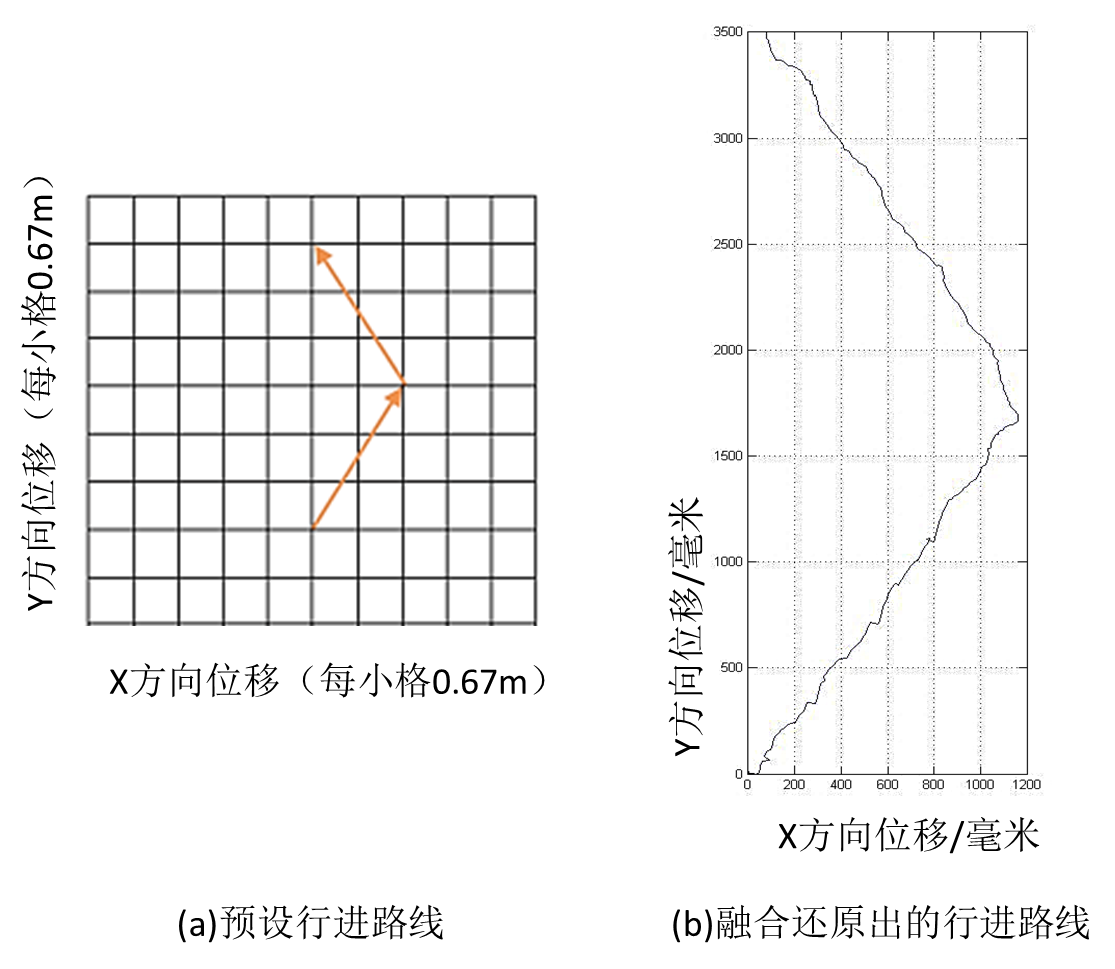
1. 直线运动情况



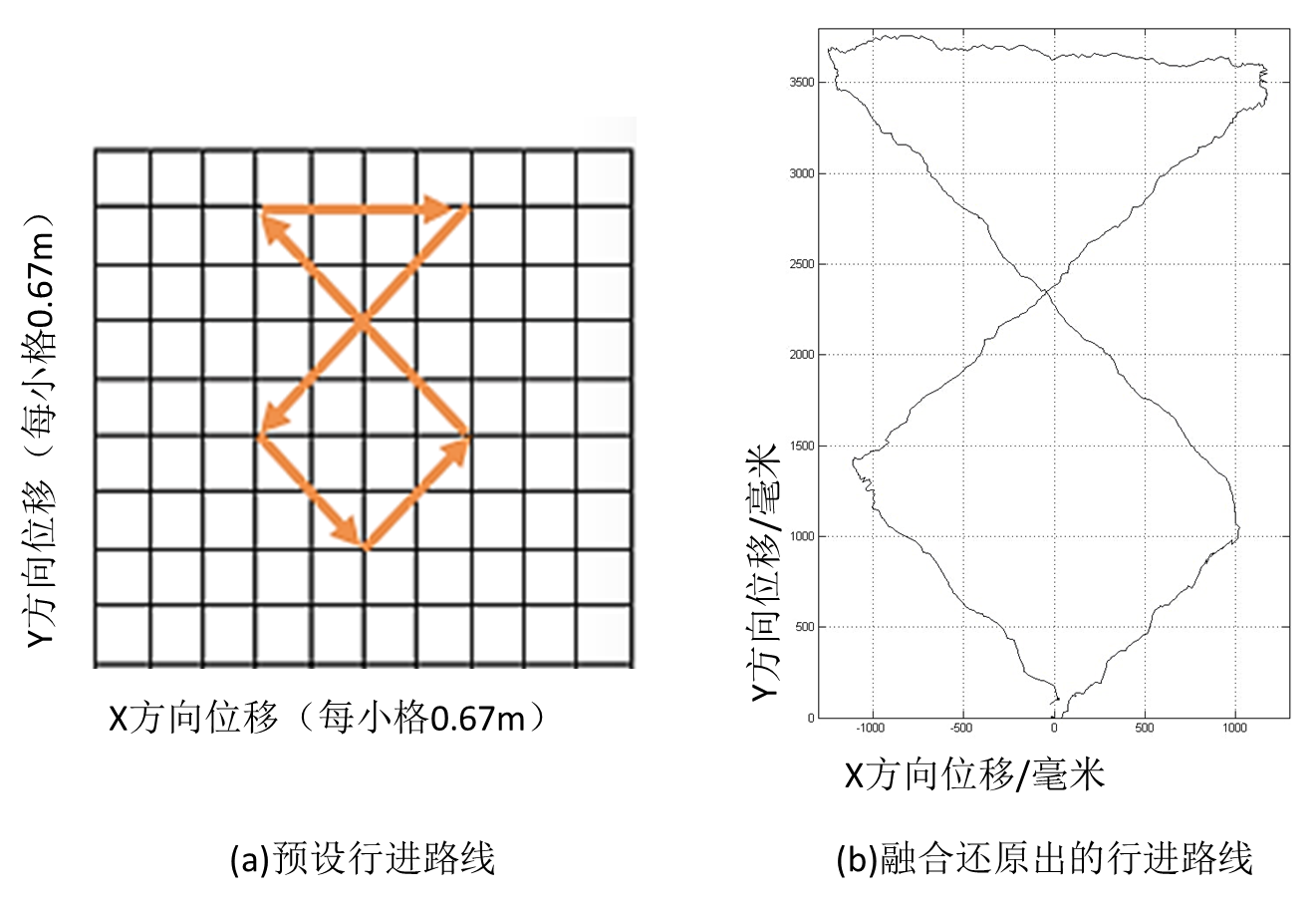
图\*\*\* 直线运动情况下还原的位移曲线

图\*\*\* 展示了直线运动情况下还原的位移曲线情况，运动过程为为人与镜头方向有一定夹角直线行进一段距离后静止，在跟踪过程中机器人需转向，手机放在口袋中。左图（a）为预设的行进路线，即被跟踪目标（人）会按照此路径行走，右图（b）为通过位移融合系统还原出的人的行进轨迹。通过比较可以看出，还原出的行进路线符合实际运动情况。

1. 复杂运动路径情况



图\*\*\* 较为复杂运动路径下还原的位移曲线（组一）



图\*\*\* 较为复杂运动路径下还原的位移曲线（组二）

图\*\*\*和\*\*\*展示了运动路径比较复杂时的实验情况。初始时，人与镜头方向有一定夹角开始运动，完成较为复杂的轨迹后停止，机器人需持续跟随，手机放在口袋中。对于这两组实验，左图（a）为预设的行进路线，即被跟踪目标（人）会按照此路径行走，右图（b）为通过位移融合系统还原出的人的行进轨迹。对比左右两组可以看出，对于比较复杂的运动路径，本系统仍能较好地还原运动路径。

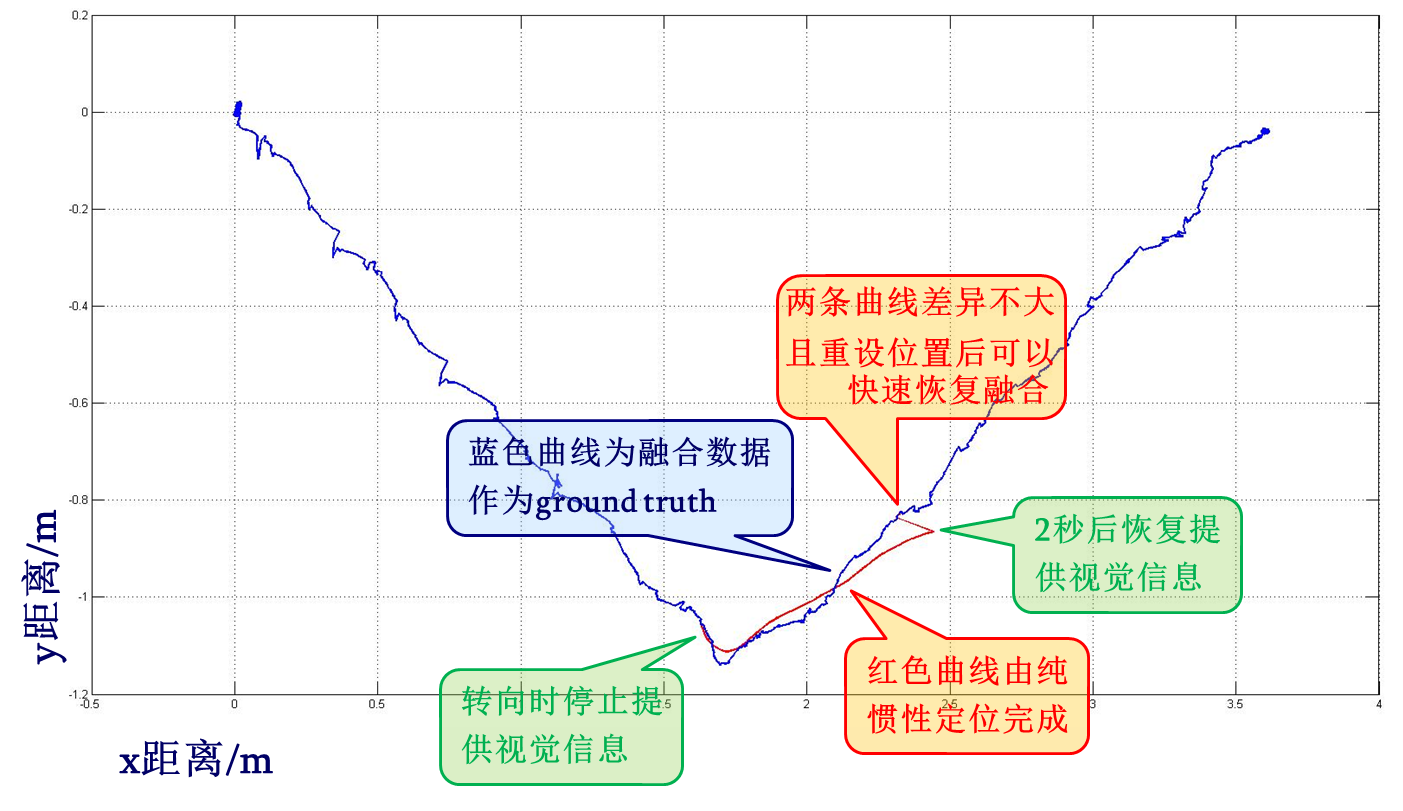
4.3 视觉定位漏检实验

4.3.1 实验方案

在本节中展示了一组实验以验证本系统可以解决视觉定位中发生的漏检问题。

实验设计与\*\*\*中类似，均为机器人实际的跟踪实验。本节中的实验模拟了如下常见的漏检情形：目标移动并在某处转向，转向时摄像头视线被障碍物（如墙壁等）遮挡，因此若只用视觉定位，则机器人无法获得正确的目标位置，进而无法正确地选择移动路径从而丢失目标的跟踪。因此，本实验将移动路径设定为一个简单的转向路径，在转向前停止提供视觉信息，在两秒后（即转向后过一段时间）恢复提供视觉信息。输出对比持续提供视觉信息还原出的运动曲线和模拟漏检时系统还原出的运动曲线。

4.3.2 实验结果与实验结果讨论



图\*\*\* 视觉定位漏检实验

如图\*\*\*所示，蓝色曲线为不发生漏检时，融合视觉定位和惯性定位还原出的运动曲线，作为对照。实验组在转向前停止提供视觉信息，2秒后恢复提供视觉信息（即机器人转向后摄像头又能重新检测到目标），期间运动曲线由纯惯性定位完成，绘制为红色曲线。对比两条曲线和恢复提供视觉信息后的曲线可以得出：

1. 在转向过程中，红蓝两条曲线差异不大，在两秒后差距不到0.5米，使得机器人转向后摄像头可以重新拍摄到目标。
2. 在恢复提供视觉信息后，触发了位置重设，然后在后续运动过程中进行融合，可以看到在较短时间后两条曲线能恢复一致，即可以快速恢复融合。

综上所述，在视觉定位发生短时间漏检时，本系统仍然可以实时给机器人提供相对准确的目标位置信息，使之追踪目标并恢复检测。

## 第五章 工作总结与展望

5.1 本文工作总结

本文的主要工作包括：

* 分析了当前基于计算机视觉和惯性定位的目标跟踪方法的局限性。
* 设计并实现了一种视觉和惯性定位结合的传感器融合系统。该系统通过获取目标携带的具有惯性测量单元的装置（如手机等）的传感器信息和深度摄像头检测的视觉信息，首先使用扩展卡尔曼滤波器融合校正惯性装置姿态，然后对目标进行实时的惯性定位，并利用视觉信息进行融合校正，最后将校正后的位置信息反馈给视觉信息提供者。
* 通过以人作为追踪目标，手机作为惯性数据提供者，带有深度摄像头的机器人作为视觉信息提供者和输出接收者，在多种情况下进行跟踪实验，验证了本文的传感器融合系统的有效性，能解决惯性定位的累计误差和视觉定位的漏检、轨迹切换等问题。

5.2 未来工作展望

在本研究的基础之上，未来可以从以下几点继续进行研究开发：

1. 在机器人上添加传感器，以增强对目标和机器人绝对定位的准确性

目前而言，本系统尽量保证获得目标相对于机器人尽可能精确的位置。与此同时，由于机器人自身轮子打滑，或者由于制导不够精确，可能会导致由制导系统计算出的机器人相对于运动初始点的绝对定位和实际值有一定偏差。比如急动、急停、原地转向等均会影响机器人计算自身位置的准确性。

如果可以给机器人上也可以放置惯性传感器等，则可以融合校正机器人自身的绝对位置。比如使用角速度计和地磁计校正转向角度，使用加速度计校正急动、急停等。

1. 结合室内地图，通过视觉识别空间内物体特征，在地图上精确标定目标位置。

通过结合室内地图，或者采用SLAM算法[]的类似过程重建地图，可以结合机器人和目标相对于运动起始点的绝对位置，在地图中标定目标和机器人。比如在商场中，可以结合商家的位置信息和目标的位置信息，准确地反馈当前位置附近商家的各项信息提供给使用者。

## 参考文献

1. 李一龙 基于RGB-D的机器人人物识别跟踪系统，2017
2. N. Trawny, S.I. Roumeliotis: Indirect Kalman Filter for 3D Attitude Estimation.
3. HUAWEI P9:http://consumer.huawei.com/cn/mobile-phones/features/p9-cn.htm
4. Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 1998,86(11):2278–2324.
5. Hubel, D. H. & T. N. Wiesel, Receptive fields of single neurons in the cat's striate cortex, Journal of Physiology, 1959, 148: 574-591
6. R-CNN: Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C], CVPR, 2014.
7. SPPNET: He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[C], ECCV, 2014.
8. Fast-RCNN: Girshick R. Fast R-CNN[C]. ICCV, 2015.
9. Fater-RCNN: Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[C]. NIPS, 2015.
10. YOLO: Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection[J]. arXiv preprint arXiv:1506.02640, 2015.
11. ImageNet: <http://www.image-net.org/>
12. PASCAL VOC: <http://pascallin2.ecs.soton.ac.uk/>

## 致 谢

感谢杨芙清院士、梅宏教授和北京大学软件工程研究所，为我提供了宽松自由的学习与研究环境，使我有机会接触并参与软件工程领域的研究项目，让我学习到很多书本上学不到的内容，为我未来的工作打下了坚实的理论基础，并积累了极富价值的工程经验。

感谢网构中间件小组的黄罡老师和刘譞哲老师。在黄罡老师开设的软件工程实验班上，我参与设计并实现了第一版缓存优化系统，在此过程中，两位老师给了我悉心的指导，使我对这个项目产生了浓厚的新区与深厚的感情。非常感谢他们给了我机会，让我可以在毕业设计中继续进行移动浏览器改进的项目。特别感谢我的导师刘譞哲老师，在百忙之中为我的研究指引方向，关心我的工作进度，在我感到懈怠时督促并鼓励我。在撰写论文与准备答辩的过程中，刘老师为我提供了大量有价值的写作建议，并亲自为我修改论文，付出了大量的时间与精力。

感谢网构中间件小组的师兄师姐，为我提供了良好的学术氛围与研究环境。特别感谢马郓师兄，他之前所做的研究是本文工作的重要前提与依据。在我毕业设计的工作中，他不遗余力地对我的工作进行指导，帮助我解决学习与生活上的各种问题，为我提供很多宝贵的研究思路和技术思路，每天花费大量时间和我讨论工作，时常忙碌到深夜。即使师兄远赴国外开会也一直挂念着我的工作，和我保持交流。在撰写论文的过程中，他帮我耐心修改论文的整体结构，逐字逐句地找问题。没有师兄的帮助，我的工作将很难达到现有的高度。感谢陈镇鹏师弟，在忙碌的期中考试周，他利用宝贵的时间帮我完成了部分实验。

感谢大学四年来知道的我的所有老师。老师们的言传身教，让我拓宽了眼界，使我在计算机领域打下了良好的基础。

最后，衷心感谢我的家人，尤其是我的父母。你们一直在背后默默支持我、鼓励我，给我无微不至的关怀，是我大学四年工作与生活中最坚实的后盾。