**第一页**

各位评审老师，下午好。我叫甄勇，研究生期间跟随洪锋教授研究渔业船舶轨迹数据。我的毕业论文题目是“基于渔业船舶轨迹数据的海洋捕捞规律挖掘研究与实现”。

**第二页**

我将从以下几个方面介绍论文的研究情况，主要内容是设计了两个算法，并且对它们进行了系统实现。

**第三页**

渔业船舶监控系统又称为VMS系统，它通过安装在渔业船舶上的终端设备采集位置、速度、船艏向等信息，经由通讯卫星发送至地面基站进行数据收集与分析。它最早用于管理船舶航行安全，现在，研究人员利用这些数据分析渔船的捕捞努力量、渔业资源的时空变化以及渔业生态环境质量等问题。这些研究有助于渔业部门了解渔业生产现状，并制定科学的计划。这么多的研究都要面对一个共同的问题——识别渔船的捕捞行为。

**第四页**

普遍的研究方法可以分为两种：基于统计的方法和基于模型的方法。前者主要利用渔船在捕捞时速度较低这一条件，根据速度阈值识别捕捞行为。这种方法简单，但识别准确率低，且不具备海域及渔业船舶类别的通用性。后者使用分类模型对VMS数据进行训练和分类。这种方法的准确率较高，但是需要如航海日志之类的辅助数据对数据的状态进行标记，且时间开销较大。由于没有与VMS数据相配的辅助信息，于是，我们以单船拖网船舶的捕捞行为特点作为切入点，探索使用图像处理方法识别渔业船舶轨迹中的捕捞行为。

**第五页**

我们的数据集包括31艘单拖渔船约260万条记录，包含了经纬度、速度、船艏向等字段。右图是其中一条渔船两年的完整轨迹，但是由于轨迹大量重叠，尤其是在港口附近渔船频繁进出，对识别捕捞行为造成了干扰。当我们人工选取三个航次进行观察时，发现轨迹叠加区域不再靠近港口，如图上红色圆圈标识的区域，实际上是由于渔船在捕捞作业时往返拖曳渔网运动造成的。我们认为该区域标示着渔业船舶的捕捞区域，位于该区域的轨迹对应着渔业船舶的捕捞轨迹。这就要求我们对VMS数据按照航次分段，而分段的依据是港口的位置，这些信息都无法直接获得。因此，为了能够在缺少航海日志等辅助数据的条件下完成捕捞行为的识别，我们依次进行了港口定位，航次划分，以及捕捞行为识别。

**第六页**

我们称相邻两条记录的经纬度坐标值相同的情况为“坐标驻留”。这种现象表明在这段时间内船舶近似静止，把它用于定位港口的位置。图上所示的例子就是某艘渔船的VMS数据统计到的全部48次“坐标驻留”，港口占38次。所有识别得到的港口汇总如图所示，共得到86个不同的港口。

**第七页**

利用港口定位的结果，我们从VMS数据中找出港口对应的记录，它们之间的记录就是航次记录，表示渔船从离开港口出海捕捞直至返回港口的完整过程。

**第八页**

从航次轨迹中可以比较明显的观察到渔船捕捞行为造成的轨迹重叠现象，如图上红色区域所示。但是仍然存在干扰因素，包括由于进出港口和轨迹交叉，如图上蓝色区域所示。为了避免这种干扰。我们采用Douglas-Peucker算法进行轨迹压缩。图（a）压缩前共7,183条记录，图（b）压缩后参数取0.03得到的折线图，保留了126条记录。通过图（c）压缩后的散点图可以看出，干扰因素基本得到了解决。

**第九页**

我们利用数学形态学这一图像处理方法识别渔船的捕捞行为。通过“腐蚀”和“扩张”两种基本运算保留轨迹重叠区域，作为渔船的捕捞区域。图（a）将轨迹数据映射到图像上，图（b）到图（d）依次进行了腐蚀、扩张、腐蚀三次操作最终得到了捕捞区域。

**第十页**

捕捞区域内的轨迹对应着渔船的捕捞行为，如左图点A到点B之间的线段。至此，我们完成了捕捞行为的识别。由于缺少辅助数据进行验证，我们按月份统计捕捞时间和捕捞距离，并与渔业产量数据计算相关度。如右图捕捞作业时间与渔业产量相关度为89.80%。间接证明了算法的准确性。

**第十一页**

通过分析前人的研究成果，我们发现捕捞区域之间的航行行为常被忽略。这种轨迹连接了两个捕捞区域，可以被认为是两者间的航道。借鉴于城市计算研究中主干道路对划分城市区域的贡献，渔船航道理应有助于捕捞区域的划分。如图所示cd两点就是所谓的捕捞区域之间的轨迹。

由于轨迹的复杂性将极大的影响计算时间复杂度，我们仅保留轨迹的端点并连线，如右图所示，称线段CD为航行线段，它表示线段两端点应处于不同的捕捞区域。

将一个季度统计的全部航行线段与对应的捕捞热度图叠加，可以发现，热度区域内部存在着大量线段，比如红色勾勒区域。根据航行线段的定义：线段两端应处于不同区域，证明了它们可以作为捕捞区域分割的依据。并将季度捕捞热度图作为分类的对象。

**第十二页**

对于实际生产而言，如果两个捕捞区域相邻，渔民往往把它们视为一个整体。因此，将捕捞区域划分成相互独立的个体对渔民捕捞作业更具有指导意义。首先，根据热度值是否为0进行粗粒度的区分。其次，根据热度值的大小选择相邻两层进行比较。左图热度值较小，范围较大；右图热度值较大，范围较小。针对图上这种在同一区域一对多的关系，我们讨论右图中的多个区块该如何组合。