目 录

[摘 要 1](#_Toc139297665)

[Abstract 2](#_Toc139297666)

[1 引言 3](#_Toc139297667)

[1.1 研究背景 3](#_Toc139297668)

[1.2 研究目的 4](#_Toc139297669)

[1.3 研究实际意义 5](#_Toc139297670)

[1.4 国内外研究现状 6](#_Toc139297671)

[1.4.1 手术目标跟踪研究现状 6](#_Toc139297672)

[1.4.2 超像素技术在图像识别领域的应用现状 7](#_Toc139297673)

[1.4.3 研究现状综述 9](#_Toc139297674)

[1.5 研究方法 9](#_Toc139297675)

[1.6 论文组成 10](#_Toc139297676)

[2 基于TLST的多模型目标跟踪算法 11](#_Toc139297677)

[2.1 本章概述 11](#_Toc139297678)

[2.2 两级超像素追踪算法 11](#_Toc139297679)

[2.2.1 算法原理 11](#_Toc139297680)

[2.2.2 初级外貌模型 12](#_Toc139297681)

[2.2.3 高级外貌模型 13](#_Toc139297682)

[2.2.4 目标跟踪过程说明 14](#_Toc139297683)

[2.2.5 TLST 算法不足 15](#_Toc139297684)

[2.3 模型一，TLST 改进模型 16](#_Toc139297685)

[2.3.1 改进一，更好的初级模型轮廓 16](#_Toc139297686)

[2.3.2 改进二，更快的高级模型建立 17](#_Toc139297687)

[2.4 模型二，图像相似度加权模型 18](#_Toc139297688)

[2.5 模型三，DSST 尺度滤波模型 21](#_Toc139297689)

[2.6 HOG、FHOG 特征提取方式说明 22](#_Toc139297690)

[2.6.1 HOG 特征提取 22](#_Toc139297691)

[2.6.2 FHOG 特征提取 23](#_Toc139297692)

[2.7 算法框架和整体流程图 23](#_Toc139297693)

[3 实验结果与分析 27](#_Toc139297694)

[3.1 本章概述 27](#_Toc139297695)

[3.2 实验环境 27](#_Toc139297696)

[3.3 实验数据 27](#_Toc139297697)

[3.4 实验方法 28](#_Toc139297698)

[3.5 实验结果 29](#_Toc139297699)

[3.5.1 不同模型融合结果对比 29](#_Toc139297700)

[3.5.2 不同超像素分割方法效果对比 30](#_Toc139297701)

[3.5.3 算法时间复杂度对比 32](#_Toc139297702)

[3.5.4 跟踪效果展示 32](#_Toc139297703)

[4 商业模式 36](#_Toc139297704)

[4.1 竞争优势 36](#_Toc139297705)

[4.2 目标客户群体 36](#_Toc139297706)

[4.2.1 目标客户细分 36](#_Toc139297707)

[4.3 推广模式 37](#_Toc139297708)

[4.3.1 线下推广 37](#_Toc139297709)

[4.3.2 线上推广 38](#_Toc139297710)

[4.4 营销策略 38](#_Toc139297711)

[4.4.1 销售策略 38](#_Toc139297712)

[4.4.2 产品策略 38](#_Toc139297713)

[4.3.3 营收策略 39](#_Toc139297714)

[4 结论 40](#_Toc139297715)

[参考文献 42](#_Toc139297716)

**基于TLST 的多模型目标跟踪算法**

# 摘 要

人们的生活需求和精神需求随着科技的发展和进步而同步增加，因为传统的医疗模式已经不能满足医疗行业的现状。而内窥镜是直接与患者的皮肤、粘膜密切接触的医疗器械，由于其结构复杂，对医生的技术也提出了很高的要求。其次，购置成本高，反复清洗、烘干、消毒等，导致医院运营成本高；以及使用、清洁、消毒都易损伤镜子，从而导致维修成本高。基于以上情况，目前已经有研究将目标跟踪算法应用到内窥镜手术中，但很多研究是通过追踪手术中医疗器械指引手术目标。目前大部分的目标追踪算法在手术场景中的应用效果并不良好。此外，不同于其他追踪算法，实时性要求会是手术场景下应用的算法的一个刚性需求。针对以上问题，本文提出了一种基于多模型融合的目标追踪算法，该算法融合了多个模型，先在原超像素跟踪算法的基础上提出了基于 HOG 特征的改进，改善了其追踪效果，同时基于图像相似度，提出了一种简单的跟踪方法，最后结合 DSST 尺度滤波器，对目标尺度做了调整和修改。最后基于人工标注的数据，我们对不同模型组合之间的结果做了交并比的结果检验，并且比较了各种情况下的算法时间，最终结果显示本文的算法在手术场景下的应用要比其他算法优秀，并且时间复杂度上处于可接受范围。

**关键词：**目标跟踪；特征描述；微创手术；DSST

**Multi Model Target Tracking Algorithm Based On TLST**

# Abstract

People's living needs and spiritual needs increase with the development and progress of science and technology, because the traditional medical model can no longer meet the status quo of the medical industry. The endoscope is a medical device that is in close contact with the patient's skin and mucous membranes, and because of its complex structure, it also puts forward high requirements for the doctor's technology. Secondly, the purchase cost is high, and repeated cleaning, drying, disinfection, etc., resulting in high operating costs of the hospital; And use, cleaning, disinfection are easy to damage the mirror, resulting in high maintenance costs. Based on the above situation, there have been studies to apply the goal tracking algorithm to endoscopic surgery, but many studies are to guide surgical goals by tracking medical devices during surgery. At present, most of the object tracking algorithms are not well applied in surgical scenarios. In addition, unlike other tracking algorithms, real-time requirements will be a rigid requirement for algorithms applied in surgical scenarios. In view of the above problems, this paper proposes a target tracking algorithm based on multi-model fusion, which integrates multiple models, first proposes an improvement based on HOG features on the basis of the original superpixel tracking algorithm, improves its tracking effect, and proposes a simple tracking method based on image similarity, and finally combines the DSST scale filter to adjust and modify the target scale. Finally, based on the manually labeled data, we test the results of the intersection and comparison between different model combinations, and compare the algorithm time in various situations, and the final results show that the application of the algorithm in this paper is better than other algorithms in the surgical scenario, and the time complexity is within an acceptable range.

**Key words:** Object Tracking; Feature Descriptor; Minimally Invasive Surgery; DSST

# 1 引言

近年来，随着计算机技术的蓬勃发展，各项技术与医疗健康领域的融合在不断加深。基于自然语言处理、语音识别的电子病历正在逐渐走向流行；基于大数据分析技术的医疗数据库能够全方位记录病人的信息，为医疗决策提供更丰富的数据；基于计算机视觉、图像识别技术的医疗影像诊断，减少了医生对于病情的判断复杂工作，可让其腾出更多的时间聚焦在需要更多解读或判断的内容审阅上。

计算机技术正在快速改变医疗健康领域，本文所研究的是计算机视觉技术中的目标追踪技术在手术中特别是微创手术中的应用。

## 1.1 研究背景

随着居民生活水平的提高，我国的内窥镜检查、手术需求旺盛。代表性的有腹腔镜，专科的，如妇科的宫腔镜，泌尿外科膀胱镜、输尿管镜，骨科的膝关节镜、椎间盘镜，胸外科的胸腔镜，呼吸内科的纤支镜，等等。2020年我国开展了929万例腹腔镜手术。近4000万例消化内镜手术。输尿管镜碎石取石和经皮肾镜碎石取石术更是广泛应用于尿结石的治疗。经尿道前列腺电切微创术已取代开放性前列腺切除术。随着电子技术发展，医用手术摄像系统进入高清、超高清。图像分辨率更高，影像更清晰。同时，机器人手术系统开始广泛应用于临床手术，手术操作更加灵活，效率更高。进口的达芬奇机器人手术系统由外科医生控制台、床旁机械臂系统及成像系统三部分组成，2020年已经在全国多所大型医院开展4万多例腹腔镜手术。国产手术机器人系统也正在进行临床试用阶段。推动着医用内窥镜行业的创新发展。随着微创手术扩展到更多的临床科室，内窥镜诊疗技术能应用到更广泛的临床场景。

中国经济稳步发展，政府对公共卫生的重视，国家对包括内窥镜在内的本土创新医疗器械的鼓励，居民健康卫生意识增强、对医疗消费升级的需求增加，将推动我国内窥镜行业持续发展。

## 1.2 研究目的

由于内窥镜是直接与患者的皮肤、粘膜密切接触的医疗器械，由于其结构复杂，消毒和清洗不彻底，容易造成交叉感染；其次，购置成本高，反复清洗、烘干、消毒等，导致医院运营成本高；以及使用、清洁、消毒都易损伤镜子，从而导致维修成本高。培养一名合格的内窥镜手术医生时间和金钱成本很高。

手术中对手术目标跟踪，常用的方式是借助于术中超声，或X线透视。同于气体的干扰，体位的影响，术者的影像学技术差异等，跟踪效果欠满意。由于内窥镜手术性质，可以直接录制图像。目前已经有研究将目标跟踪算法应用到内窥镜手术中，但很多研究是通过追踪手术中医疗器械指引手术目标。目前大部分的目标追踪算法在手术场景中的应用效果并不良好。原因一是因为手术情况复杂多变，器官的出血，镜头模糊，器官变形，手术器械遮挡等，均可能影响到目标的追踪。二是要求对手术目标的追踪需要实时性。这对于很多算法提出了很高的要求。

## 1.3 研究实际意义

腹腔镜手术的定义是在冷光源照明下，把腹腔镜插入腹腔中，通过数字成像，实时显示在监视器进行的手术。其中两孔法后腹腔镜手术具有其创伤小，安全性高，疼痛小，恢复快等特点。因此，目前已经在多种器官系统应用，常见的手术有：阑尾切除手术，胆囊切除术，胰腺部分切除术，肿瘤切除术，肾切除术等等。

然而腹腔镜手术因为开口小，专业性强，因此对医生的熟练度有着很高的要求，与此同时，部分腹腔镜手术，例如腹腔镜肠梗阻手术因为其手术难度大，风险高，仍未普及。首先，既往手术史等因素可能会导致腹腔内广泛粘连，其次，腹胀、肠管扩张等因素使得可操作空间变小，器械受限制，再次，肠壁变薄弱从而容易受到损伤。有数据显示，在肠梗阻手术中，有6.6%的几率因为手术器械原因导致肠壁穿孔，并且有近乎五分之一的损伤是因为严重术后症状才被发现。

因此，无论是术前培训时提供一种辅佐培训，还是在术时基于实时成像为医生供实时信息反馈，对手术中目标进行检测和追踪都十分有必要。这种追踪即可以在仿真培训时给新手医生提供良好的反馈建立经验模型，同时又可以给手术中的医生提供更准确的信息，避免手术器械对内部脆弱器官的损伤。除此之外，手术目标精准、鲁邦的检测与跟踪，能够为手术导航中术前 3D 模型的非刚性配准提供依据和特征述，减少手术导航中抖动、漂移等现象。

## 1.4 国内外研究现状

### 1.4.1 手术目标跟踪研究现状

目前，国内外学者基于微创手术辅佐技术的广泛研究提出了多种目标追踪算法。陈兆瑞[1]在一篇算法综述中将目前计算机辅佐的微创手术工具跟踪算法分为四大类。第一类是基于机器学习的算法，通过特征工程对不同条件下的手术工具进行特征提取，基于海量的特征数据通过 SVM 或者模式识别的方法训练出有效的分类器，最后通过分类器进行手术工具的检测和相应的三维重建。第二类是基于立体图像分析的算法，伴随深度摄影机的出现，三维建模更加方便准确，通过深度摄影机获取的三维特征，结合目标的颜色和深度特征和其他信息，可以提取出目标相应的 3D 特征。第三类，基于普通视频图像分析的算法[18，19，20]。该类算法指的是，在普通视频中，难以对整个工具进行检测追踪，一般是对目标尖端进行识别。例如通过Sobel算子对尖端轮廓进行特征提取，或者是利用 CN 特征获取目标颜色特征等等。第四类，基于辅助设备定位的算法。该类算法通过物理的方式，利用附加的导航设备对工具进行精确定位，例如葛鑫[2] 2011年提出的基于六自由度电磁的跟踪方法。Unmil[3]在 2014 年提出了一种基于单相机的微创手术检测算法，该算法运用了Camshift 算法和 Kalman 滤波，在非遮挡情况对于多特征点的目标追踪有着很好的效果。Du [4] 在 2016 年提出了一种结合了 2D 和 3D 的微创手术追踪算法，该算法先在 2D模型下利用广义霍夫变换和 SIFT 特征构建一种追踪器，并在每一帧中用该 2D 追踪器初始化3D 场景，甚至在目标遮挡的情况下也取得了良好的效果。Zhao [5]在 2017 年提出了一种基于深度学习和卷积神经网络的追踪算法，该算法通过对工具轴端和末端执行端分开进行检测，前者通过RANSAC 方案中的线性特征进行描述，后者通过卷积神经网络提取特征。除了对手术目标的检测与追踪算法的研究以外，Shaharan [6]等人还提出了一套对于微创手术培训技能评估指标。用于非观察性培训系统的设计和使用。

### 1.4.2 超像素技术在图像识别领域的应用现状

目前大部分的追踪算法大致可以分为两类，一类是生成类方法[24,25,26,27]，一类是判别类方法[28,29,30,36]。生成类方法是基于外貌轮廓模型在下一帧图像中寻找一个最佳的目标，通常是定义一个损失函数，使得该函数值最小。而判别类方法将目标追踪视为二分类问题，通过提取图像的特征，对分类器进行训练，达到前景背景的区分从而达到追踪的目的。

超像素分割因为自然考虑了物体的轮廓边界，因此在目标追踪中有着良好的应用前景。在应用到图像识别领域中，有的[33,34,35]只依靠超像素的特征就可以对物体的轮廓特征和颜色特征进行提取和追踪,有的[16,17,21,22,23]还要结合物体本身的关键 点或者其他方式，才能进行有效的跟踪。Ren [7]提出了一种基于二分类的算法，每一帧的图像通过约束 Delaunay 三角网被分成几个部分，通过条件随机场来匹配帧与帧之间的部分。然而这种方法的追踪效果并不是很好，并且在时间成本上也很高。Yang [8]提出了一种基于外貌轮廓模型生成的方法，通过获取建立一个置信图表示出超像素块与追踪目标之间的关系，来确定下一帧目标的位置。不过这种方法需要一些训练过程，因此需要人为调整超参数，并且需要其他追踪算法事先获得前几帧的追踪结果来初始化训练过程。Wang[9]提出了一种基于多对多的超像素追踪算法，通常超像素算法都是在新一帧中的超像素和之前的超像素之间建立一种联系，不同于一对一的匹配，Wang基于KNN 搜索，提出了一种多对多的超像素匹配追踪算法，不过这种匹配过程很复杂，因此在时间复杂度上效果不佳。

不同于其他超像素追踪算法，Cai[10]等人提出了一种基于动态图结构的算法。该算法通过利用图像本身的位置信息，通过超像素块构建图节点，而边是通过图中超像素节点之间的信息提取得到，再利用马科夫随机场区分前景和背景，通过权值投票机制选出属于目标的超像素块从而得到相应的追踪结果。然而这种方法时间成本较高。基于这种思想，Yeo[11]提出了一种基于吸收马科夫链的超像素追踪算法。该算法将标签为背景的超像素块视为吸收态顶点，其他超像素块到达吸收态顶点的时间就是它到背景的距离，时间越短，距离越短，属于背景的概率越大，反之，则可能属于目标。因此该算法先对t帧的超像素进行分割，建立吸收态马科夫链图，然后计算每个超像素块吸收时间，再结合 t-1 帧的吸收态马科夫链图和平均吸收时间，对t帧的超像素进行分类，以此递归。该算法的好处是，因为超像素的原因，不仅可以追踪到目标的位置，还可以体现出目标的轮廓，不过构建吸收态马科夫链图会花费大量时间，因此在时间成本上需要进行权衡。

### 1.4.3 研究现状综述

虽然目前已经有很多研究开始尝试将目标跟踪算法应用到手术中，但是很多研究都是针对手术中的医疗器械的追踪，并且很多典型的目标追踪算法在手术场景中的应用效果并不良好，一是因为手术情况更加复杂，器官出血，镜头模糊，器官变形，手术器械遮挡等等问题比较复杂，二是因为对于手术目标的追踪需要实时性要求，因此在这方面对于很多算法提出了很严格的要求。

而目前基于超像素的追踪算法也大致和其他原理的追踪算法相同，一是基于外貌模型的建立选出最有可能的目标，二是基于分类器的前景背景区分。超像素分割虽然对于目标追踪提供了更为良好的轮廓信息，但是很多情况也加大了算法的时间复杂度，很多基于超像素的追踪算法应用通过构建图结构来完成目标提取，虽然可以做到对目标进行相应轮廓的匹配，但是时间成本上需要进行更多考虑。

## 1.5 研究方法

本文研究方法主要分为三部分。

第一部分是基于一种超像素跟踪算法的改进。通过结合目标的HOG特征，改善了目标的外貌模型，并且修改了算法细节使得总时间时间成本减少。

第二部分是基于图像相似度度量提出了一种简单的跟踪方法。对原图像进行降采样，在低分辨率的情况下通过图像相似度，获得目标的位置。

第三部分是基于DSST尺度滤波的应用。基于尺度滤波器，提取出不同尺度的目标区域，对目标的尺度进行训练的调整。

## 1.6 论文组成

本论文主要由三部分组成。第一部分引言，主要对课题背景，相关的研究情况以及论文组成进行介绍。第二部分技术创新点和描述，主要介绍了本文算法应用的模型所涉及的知识，包括图像相似度度量、特征描述子、DSST 尺度滤波，以及本文的创新点及核心算法，包括本文算法涉及的三个模型。第三部分实验结果分析，主要采用IoU评估方式在跟踪效果和时间复杂度上对本文算法进行了多角度对比实验。

# 2 基于TLST的多模型目标跟踪算法

## 2.1 本章概述

本章主要介绍了 TLST 模型，并且在 TLST 的基础上针对其不足提出了融合模型，该模型主要由三个模型组成，第一个模型为对 TLST 算法改进的模型（TLST-i），第二个模型为基于图像相似度的加权模型，第三个为 DSST 尺度模型。

## 2.2 两级超像素追踪算法

2017 年 Jun Wang 提出了一种基于超像素分割的目标追踪的算法，名为两级超像素追踪，Two-Level Superpixel Tracking (TLST),该算法采用 SLIC 超像素分割方法，通过对图像在两种等级之间进行超像素分割并且建立联系，来达到对目标物体的追踪。下面是对该算法的简要介绍。

### 2.2.1 算法原理

假设代表在第 t 帧时的追踪结果，代表第 t 帧的观测结果，那么基于贝叶斯模型的追踪算法的原理就是，已知，寻找最佳的。用{x，y，w，h}即目标框的位置和大小来表示。追踪的结果可以通过最大后验概率来计算。

由贝叶斯定理可知，

又有，

基于(2-2)、(2-3)可知是一个叠积分，只需要计算出和即可求得。

前者称为观察模型(observation model)，后者称为运动模型(motion model)。在观察模型中,认为为第 t 帧图像目标框内超像素点的置信度之和。同时认为运动模型符合高斯分布。

### 2.2.2 初级外貌模型

初级外貌模型的建立主要分为三步。第一步，确定 ROI，第二步，对 ROI 进行超像素分割，第三步，计算超像素点的特征。下面将简要介绍每个步骤。

第一步，因为考虑到物体的移动是连续的，短时间内物体不会发生剧烈的移动，因此只需要考虑物体附近的区域即可。因此根据确定的目标框，同心扩大倍，即是 ROI。

第二步，根据 ROI 区域的大小，确定需要分割的超像素块大小，然后对 ROI 进行超像素分割。

第三步，根据(2-4)建立初级外貌模型

其中表示超像素块位置，公式为，

表示超像素块HIS特征，公式为，

表示超像素度权值，公式为，

代表超像素的数量。代表每个超像素的位置信息是组成该超像素的像素点的位置信息的平均。对于 HSI 特征同理。代表超像素块在目标区域内的面积。代表该超像素块的权值的正负，当像素块的面积超过某一定阈值时认为其权值为正。

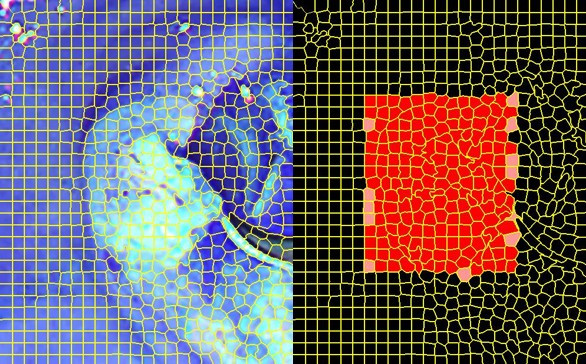


图 2-1 超像素分割手术目标效果和相应的权值示意图

### 2.2.3 高级外貌模型

当新的一帧到来时，基于物体运动不会太快的假设，目标仍然处于 ROI 中，此时对 ROI 做进一步超像素分割，分割的块数是初级模型的倍，这样做的目的是考虑目标形变的情况。如果目标发生形变，那么则需要更多的超像素块来保证对目标抽象的质量。

设 sps 表示新的一帧下的超像素集合，sp 表示之前帧的超像素集合。利用位置和 HSI 特征在当前帧的超像素块和之前帧的超像素块之间建立联系，求得新的超像素的权值。新的一帧下的超像素块和之前帧的超像素块之间的关系如下。

其中定义如(2-11)，表示是 HIS 特征和位置特征的综合结果

其中定义为，

定义为，

dfmax和dpmax是归一化因子。对于任意新的超像素块 r，求得超像素块 k，使得最大，则 r 的权值是 k 的权值和的乘积。有了权值则可以构建新的外貌模型。

### 2.2.4 目标跟踪过程说明

新的一帧到来，通过构建高级外貌轮廓模型，得到超像素新的权值，从而整个 ROI 区域得到相应的置信图。由于认为运动模型是高斯分布的，因此选取 N 个目标框，计算相应的高斯概率以及框内的置信度，选取结果最好的框作为下一帧的目标位置。

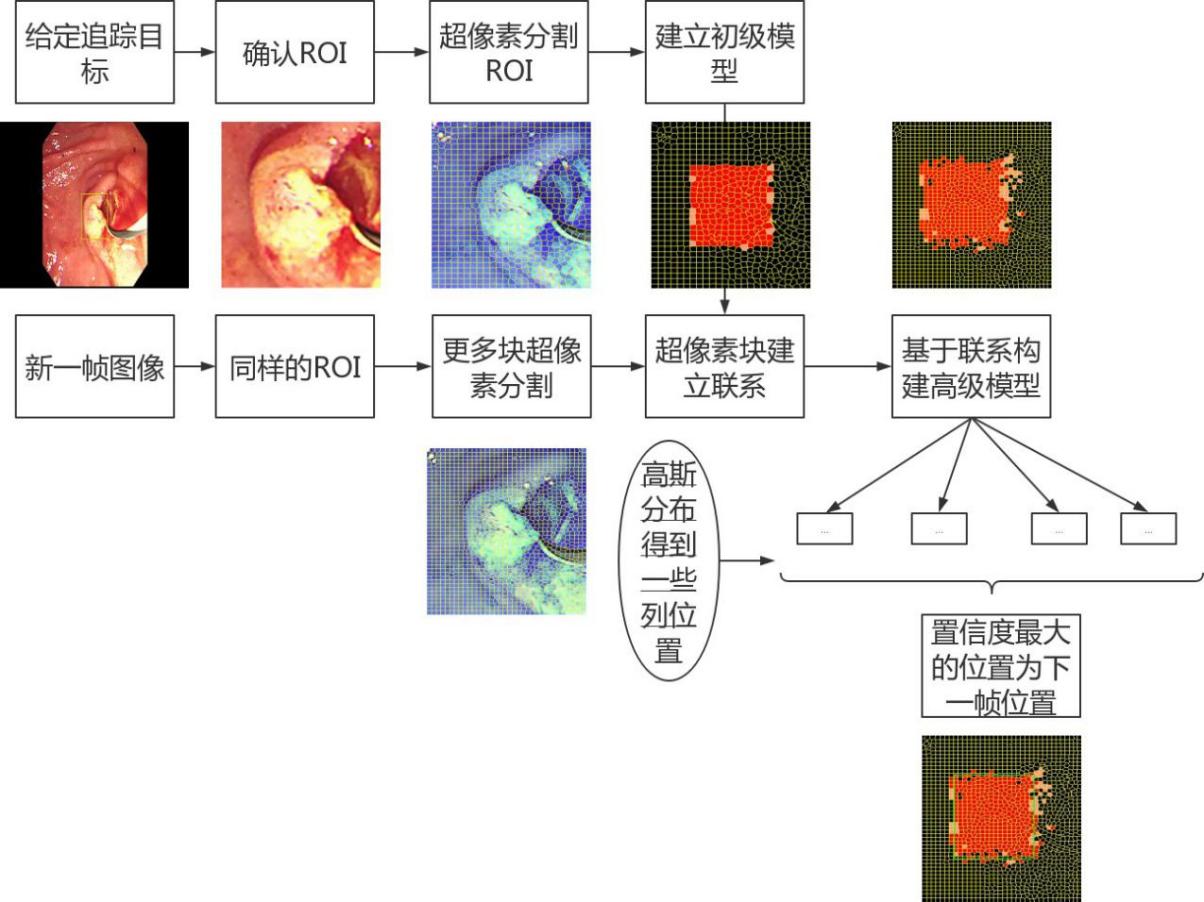


图 2-2 TLST 算法流程示意

### 2.2.5 TLST 算法不足

TLST 算法的基本思想是在高级模型和初级模型之间建立一种关系，从而通过初级模型的中超像素的权值来估计高级模型中超像素的权值。但是算法的问题有两个：

一、关于初级模型中超像素的权值仅仅和面积有关，导致目标区域（通常是一个长方形）内的所有超像素块权值均为 1，这显然有待改进，因为真正目标的形状不是标准的长方形，因此选定长方形内的超像素块有很多是属于背景而不属于目标。

二、对于高级模型的超像素 k，需要选择初级模型的超像素 j 与之对应，TSLT 的方法是，对于每一个初级模型的超像素块计算其和 k 的关系，选择关系指数最好的 j 与 k 建立对应关系。这会导致计算量是超像素数量的平方，时间复杂度剧增。

## 2.3 模型一，TLST 改进模型

### 2.3.1 改进一，更好的初级模型轮廓

初级模型的权值赋值不通过在目标内的面积决定。而是按照以下规则设定初级模型中的超像素权值。

1、将目标区域分成 m\*n 个 p\*q 的方块。

2、遍历所有方块，每遍历一个块，求该块的 HOG 特征，如果其 HOG 特征的最大值大于某个阈值，则认为该方块是边界方块，反之则不是。

3、去除可能的奇异方块，即边界方块之间应该相连，如果该方块是孤立的，即该方块周围的 8 个方块均不是边界方块，则将其排除。

4、对于超像素 k，其所在区域覆盖了 s 个方块，将 s 个方块分成 3 类，一类是边界方块，一类是边界内方块，一类是边界外方块。（边界内外的区分可以通过边界方块的 HOG 特征的主成分方向以及方块的位置来确定）。若边界内方块数量与 s 的比值超过一定阈值，则认为超像素块 k 权值为 1，否则权值为 0。

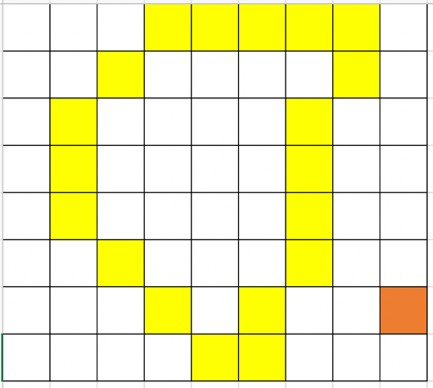


图 2-3 边界方块和奇异方块示意

### 2.3.2 改进二，更快的高级模型建立

高级模型的超像素权值的确定不再遍历整个初级模型的超像素集合。对于高级模型的超像素 k，其位置中心为(x,y)，以此为中心的 D\*D 范围内初级模型超像素有个，只需要遍历这部分的超像素块即可。通过控制窗口 D 的大小，则可以控制窗口内超像素的个数在常数级别。从而整个运算是超像素个数的一次方复杂度。

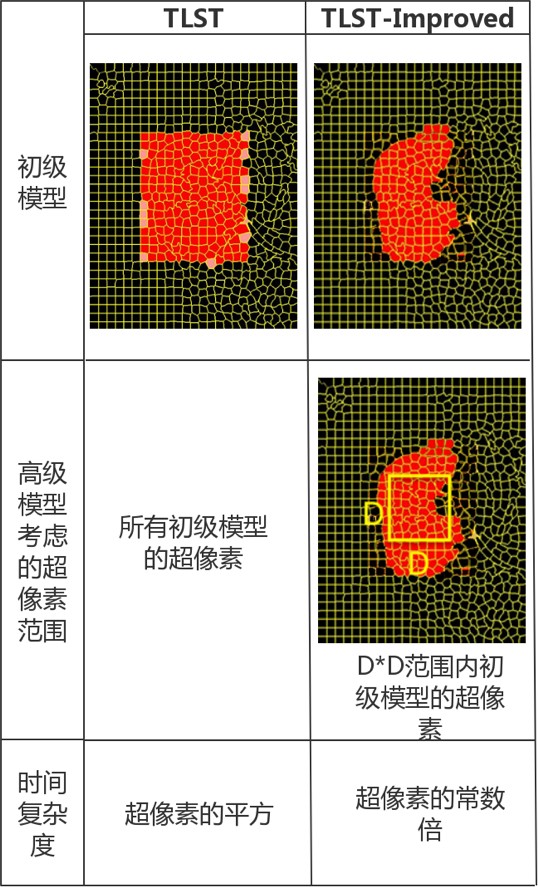


图 2-4 TLST 和其改进的对比

## 2.4 模型二，图像相似度加权模型

TLST 算法通过新的一帧超像素分割的超像素块与之前超像素块建立权值关系，从而建立置信图求得最佳位置。而权值关系的建立基于超像素块的 HSI 特征以及位置信息。因此根本上，TLST 算法实际上是对超像素块颜色以及位置信息的综合利用，权值只是一种中间载体。为了更加直接的应用超像素块的颜色以及位置信息，本小节提出了一种基于降采样金字塔和图像相似度的目标追踪方法。

简单来将，通过获取对 ROI 区域进行降采样，类似于 SIFT 特征中的步骤，即对 ROI 区域进行图像压缩，获取不同分辨率的图像，然后以目标框初始大小为窗口，遍历整个 ROI 区域，通过图像对比，寻找概率最大的位置，并通过相应的分辨率比例，找到原图中最有可能的位置。以下是算法简介：

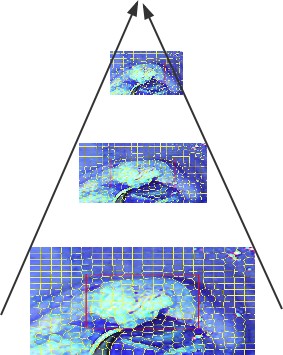


图 2-5 降采样金字塔示意图

第一步，对每一帧图像添加双边滤波预处理。

对于手术中的场景，因为灯光，光照，等原因，会有一些离群值，也就是一些暗点或者亮点出现在视野之中，这对于超像素技术来说需要事先剔除，否则这种噪声有可能在之后形成一个专门的块，影响结果。但是简单的高斯滤波只考虑像素的位置信息，并没有考虑像素之间的颜色差异，因此会模糊物体轮廓边缘，这对于后续的超像素处理也是不好的。因此需要一种既可以滤除噪声，又可以保障物体边界的滤波来进行预处理，而双边滤波正好满足条件。

第二步，对原 ROI 区域进行降采样。

降采样的目的是为了降低后续计算图像相似度的时间复杂度，如果在原像素基础上对目标进行比较，因为分辨率级别过高，会使得处理时间过长，无法满足实时要求。降采样的思想和 SIFT 特征提取中降采样相同，不过不同点是，层与层之间的降采比例是不同的，假设原 ROI 是 m\*n 的图像，只要分辨率小于 即可以满足实时要求，那么降采样得到的一系列图像的分辨率比例系数为,其中 >>>> ⋯ >。

第三步，对每一层降采样图像求最佳跟踪位置。

构建相应压缩比例的目标框图像，以压缩后的目标框图像为窗口，遍历该层降采样的 ROI 区域，求得对比结果最佳的位置。例如，对于第 k 层降采样结果，ROI 区域大小为，目标框原本大小为 w\*h，那么与之对应的压缩目标大小为。从压缩的 ROI 图像左上角开始，确定一个的窗口，然后先向右水平滑动窗口，再向下垂直滑动，以此遍历整个图像。每滑动一次计算一次窗口内的图像和目标压缩图像的相似度比较结果，从而得到结果最佳的位置处。

第四步，按比例加权。

由于图像经压缩之后信息丢失严重，因此压缩比例越高的降采样层的结果的权值应该越低，因此假设为降采样层从下到上的分辨率比例系数，那么与之对应的权值为,其中。

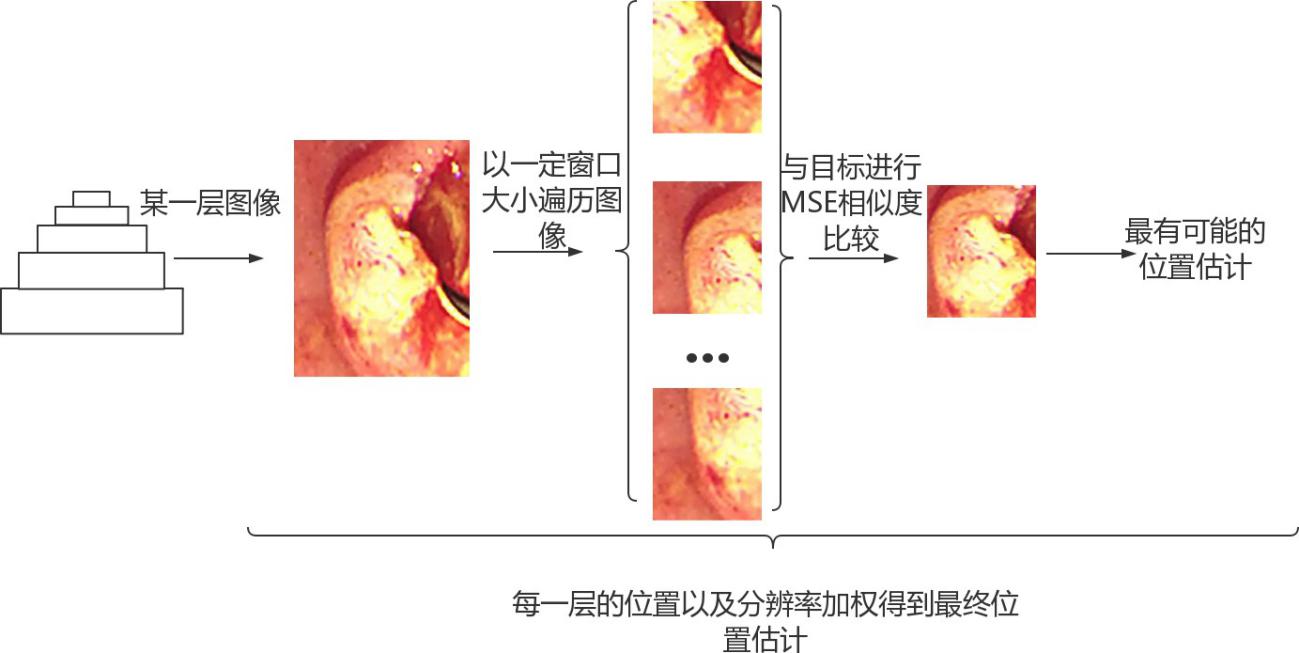


图 2-6 基于图像相似度的加权模型示意图

## 2.5 模型三，DSST 尺度滤波模型

根据之前 DSST 的描述，本文通过前两种模型获得目标的位置，然后在该位置中心，应用尺度滤波器，调整其尺度。关于尺度滤波器，主要分为尺度模型的初始化，尺度模型预测，和尺度模型更新三个部分，以下将依次简介其应用过程。

尺度模型初始化。

按照 DSST 的尺度提取方法，提取出以第一帧初始结果为中心的 33 个尺度样本，然后将其 resize 到同样大小。每个样本分别提取 FHOG 特征记为 sf。将特征通过快速傅立叶变换转换到频域进行计算，记为 sff。将 sff 和其复共轭点乘记为 B。将高斯输出相应的快速傅立叶变换结果记为 ysf，将 ysf 和 sff 的复共轭点乘记为 A。A 和 B 即为初始化尺度滤波的分子和分母。

尺度模型预测。

新一帧到来，获得目标位置之后，同样在目标中心获得 33 个尺度的照片并且 resize 到同样大小，获取相应图像的 FHOG 特征并进行快速傅立叶变换，记为 xsf，然后将 xsf 和 A 点乘，然后将结果点除 B+，再将结果进行快速傅立叶逆变换，然后取其实部，则得到新一帧图像在尺度滤波下的相应。相应中最大的结果就是最佳的尺度。

尺度模型的更新。

已经检测到最佳尺度之后，提取最佳尺度的 FHOG 特征，记其快速傅立叶变换结果为 xsf，记其相应的高斯相应的快速傅立叶变换结果为 ysf，则,

。其中表示学习率。即，新的尺度滤波是在检测结果的基础上更新 A，B 得到的。

## 2.6 HOG、FHOG 特征提取方式说明

### 2.6.1 HOG 特征提取

由于 HOG 在本算法中的应用是为了检测边界方块，因此并不需要特别区分梯度的方向，因此本文中涉及的 HOG 特征是无符号的。本文中涉及的 HOG 特征维度为 9 维，即把 180 度分成了 0-20，20-40，40-60，60-80，80-100，100-120，120-140，140-160，160-180 九个区间。本文选择的直方图统计的单元格大小是 8\*8，因此对于所处理的目标区域的宽高不是 8 的倍数，会将相应区域左右或者上下扩大一定像素区域，使得其满足 8 的倍数。HOG 特征是归一化的特征，归一的过程是由几个网格的特征组合在一起归一，本文选择的归一的大小是 2\*2 网格。

### 2.6.2 FHOG 特征提取

由于 FHOG 的是 31 维特征，也就是 9 维无方向 HOG 特征和 18 维有方向特征和 4 维邻域特征，本文只利用了其前 27 维特征，即 9 维无方向部分和 18 维有方向部分。对于无方向部分，可以通过相应的 HOG 特征获得，因为关于 FHOG 的提取，本文选择的网格大小也是 8\*8，归一窗口的大小也是 2\*2 网格。对于有方向部分，需要将 360 度每 20 度一个 bin 分成 18 部分，重新进行直方统计。

## 2.7 算法框架和整体流程图

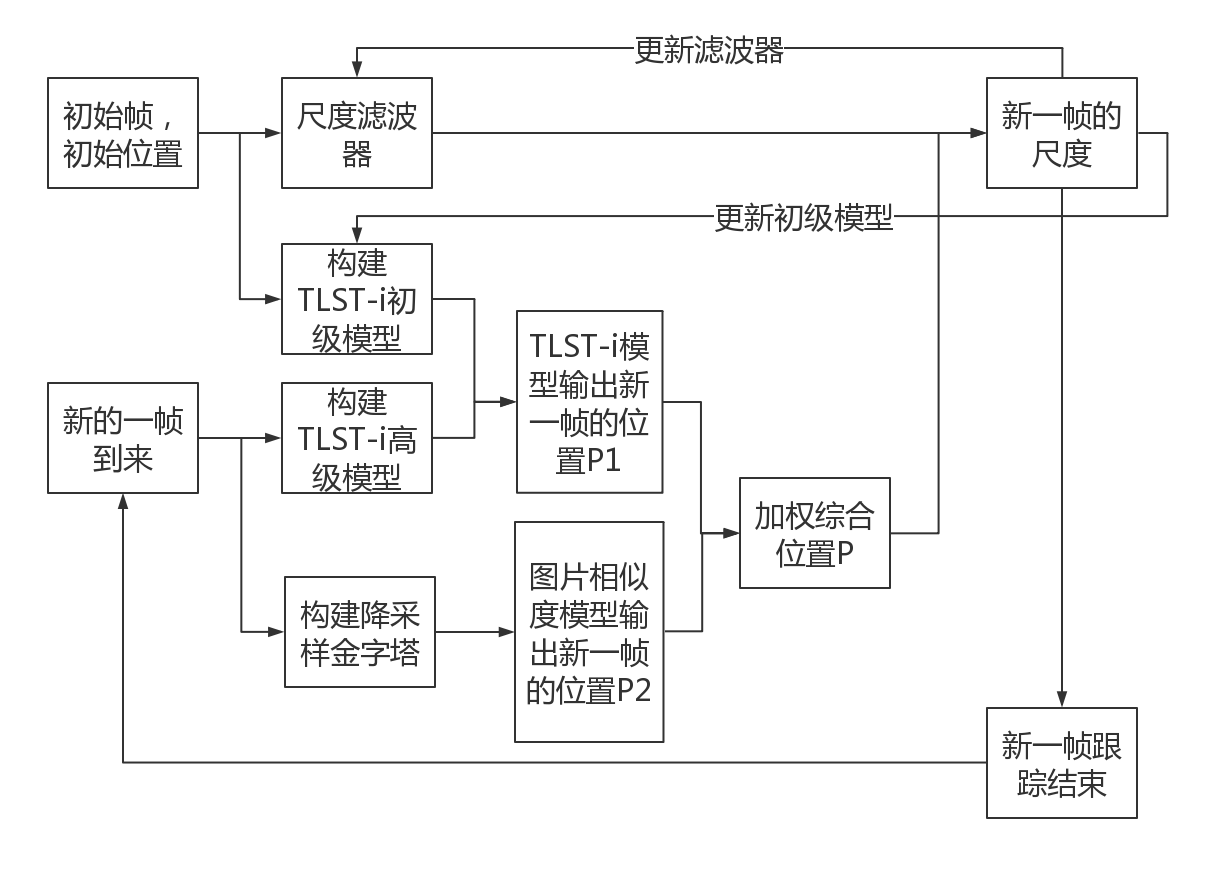


图 2-7 算法整体框架

整体算法框架如图2-7。

1. 用最开始的位置初始化 DSST 尺度滤波，构建 TLST 初级模型
2. 当新一帧到来时，TLST 改进模型构建高级模型输出位置 P1，降采样金字塔模型构建降采样金字塔输出位置 P2
3. 加权计算得到估计位置 P
4. 基于 P 和 DSST 尺度滤波，得到新一帧的尺度 S
5. 基于 P，S 构建新一帧的 TLST 初级模型，更新尺度滤波
6. 输出 P，S
7. 重复 2-6 直到任务结束

降采样金字塔模型。

1. 对每一帧图像利用双边滤波进行预处理
2. 对 ROI 区域按照比例系数，进行降采样获得降采样金字塔
3. 每一层的比较结果 S
4. For 每一层采样 M in 金字塔
5. 计算出目标相应比例的图像 I
6. 以 I 为大小遍历 M，移动一次计算相应的 MSE 值
7. 选择 MSE 最大的位置作为 M 下的估计位置
8. 根据和 S 加权计算最终的目标位置 P
9. 输出 P

DSST 尺度模型。

1. Input：输入图像，上一帧尺度，尺度模型,，当前位置
2. Output：估计目标尺度，更新的尺度模型,
3. 以目标当前新位置为中心，提取 33 种不同尺度的样本
4. 利用,,,根据公式计算出y
5. 计算 max(y)得到目标准备尺度
6. 基于当前位置和估计的尺度，提取样本特征。
7. 基于更新
8. 基于更新

TLST 改进模型。

1. 基于 TLST 算法过程构建初步模型（除了超像素块的权值特征）
2. 将目标区域按一定方块大小分成 m\*n 块方块
3. 提取目标区域内的 HOG 特征并按照每分割的方块进行统计
4. For 每个方块 i
5. If 方块 i 的 HOG 特征最大值>sigma:
6. L（i）=1
7. Else：
8. L(i)=0
9. 去除奇异方块
10. For 每个超像素 k
11. 统计其边界内方块个数 m 和其覆盖的方块总数 n
12. If m/n>beta:
13. Weight(k)=1
14. Else:
15. Weight(k)=0
16. 当新一帧到来，构建高级模型
17. For 高级模型中的每个超像素块 k
18. 其中心位置为 p，以 p 为中心展开一个 D\*D 的窗口
19. 统计在 D\*D 窗口内出现的初级模型的超像素块集合 S
20. For 超像素块 t in S
21. 按照 TLST 算法计算
22. 选取 S 中最大的超像素块作为 T
23. Weight（k）=Weight（T）
24. 剩余算法和 TLST 模型相同

# 3 实验结果与分析

## 3.1 本章概述

本章主要介绍了本文的实验过程，主要包括实验的环境说明、实验数据来源和信息 说明、实验的评估方法以及不同角度的实验结果和分析。

## 3.2 实验环境

本文实验均在一台 MacBook Pro（15-inch，2018）上运行，其配置如下。

表 3-1 实验配置表

|  |  |
| --- | --- |
| 操作系统 | macOS High Sierra (version 10.13.6) |
| 处理器 | 2.2 GHz Intel Core i7 |
| 内存 | 16GB 2400 MHz DDR4 |
| 图形卡 | Intel UHD Graphics 630 1536 MB |
| 运行环境 | Python3 |
| 环境依赖包 | Numpy、cv2、skimage 等 |

## 3.3 实验数据

本文实验数据均来自于真实医院手术视频，其视频信息数据如下，关于视频1，2，3为对十二指肠乳头的跟踪，其中视频 1，手术目标没有明显位移和形变，视频 2，手术目标有明显的位移以及形变，视频 3，手术目标有微量位移和形变，视频 4-7 是对其他手术器官的跟踪。

表 3-2 实验配置表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 视频序号 | 时长/s | 输入帧率 | 宽 | 高 | 帧数 |
| 1 | 7 | 25 | 1920 | 1080 | 181 |
| 2 | 15 | 25 | 1920 | 1080 | 376 |
| 3 | 12 | 25 | 1920 | 1080 | 301 |
| 4 | 15 | 25 | 720 | 576 | 374 |
| 5 | 19 | 25 | 1920 | 1080 | 375 |
| 6 | 30 | 25 | 352 | 288 | 750 |
| 7 | 33 | 25 | 1280 | 720 | 821 |

## 3.4 实验方法

本文实验采用交并比来对结果进行评价。即，每一帧，真实的目标位置和大小为一个长方形框，算法追踪结果为一个长方形框，计算两框之间的交集的面积和并集的面积的比值。其计算公式如下。

因为本文的数据集没有官方的标准结果，因此本文采用人工标注的方式获取每一帧的真实位置，并且多次标注选取平均值作为参照依据。

## 3.5 实验结果

### 3.5.1 不同模型融合结果对比

对于不同的模型组合情况，本文实验了只有 TLST 模型的单独表现情况，以及 TLSTi 和基于图像相似度的加权模型的组合表现情况，以及 TLST-i、图像相似度加权模型和 DSST 尺度滤波模型的组合表现情况。其结果如下。

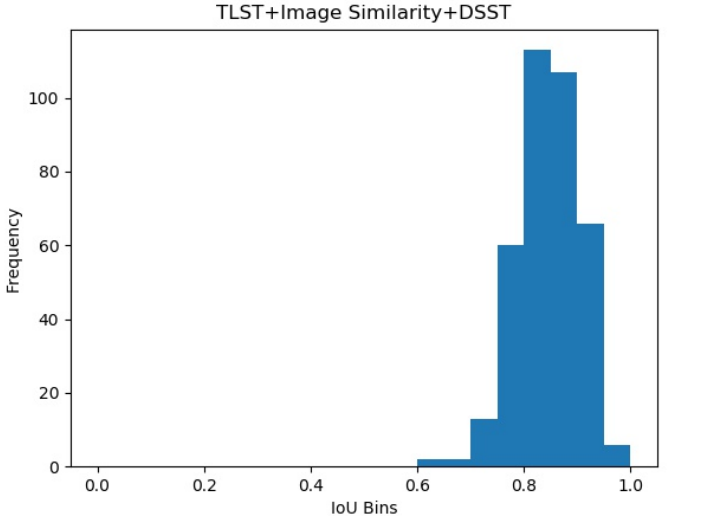


图 3-1 本文模型 IoU 统计情况

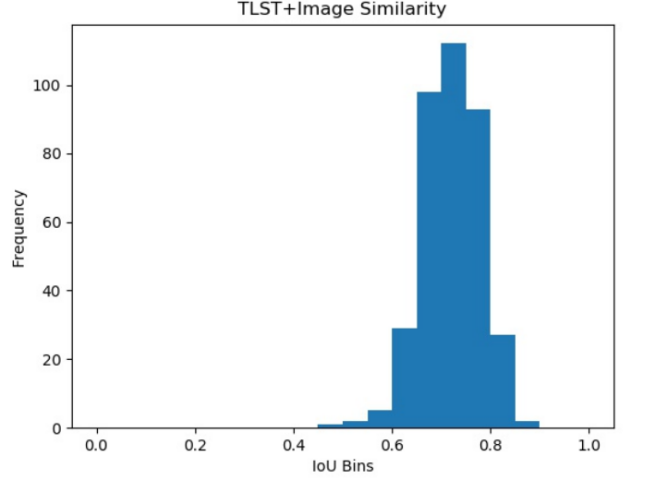


图 3-2 TLST-i 模型和图像相似度模型融合的 IoU 统计情况

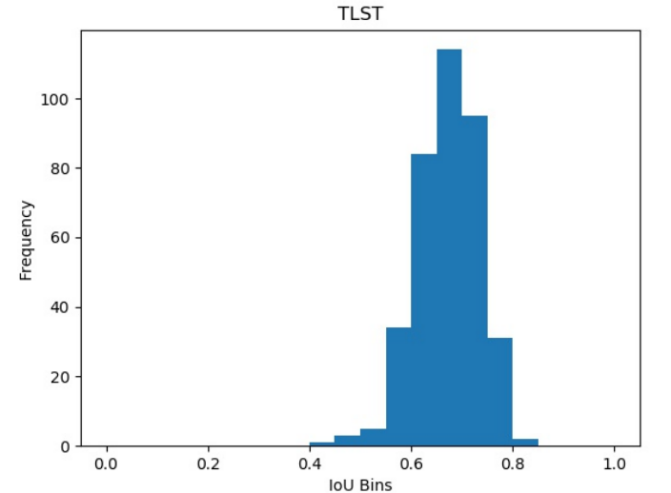


图 3-3 TLST 算法的 IoU 统计情况

从图 3-1 图 3-2 图 3-3 可知，本文的融合模型 IoU 大致集中在 0.8 到 0.9 之间，TLST-i 和图像相似度融合的模型的结果大致集中在 0.65-0.8 之间，而 TLST 算法的结果大致在 0.6-0.75 之间。由此可见，本文的算法在 IoU 尺度下效果要好很多。

### 3.5.2 不同超像素分割方法效果对比

除了不同模型的组合对比，本文还对不同的超像素分割方法做了对比实验。

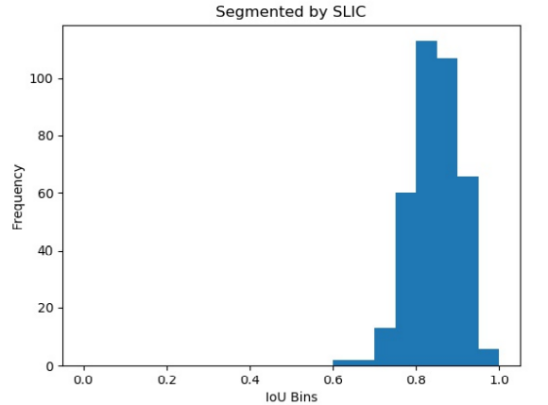


图 3-4 本文算法运用 SLIC 超像素分割的 IoU 统计结果

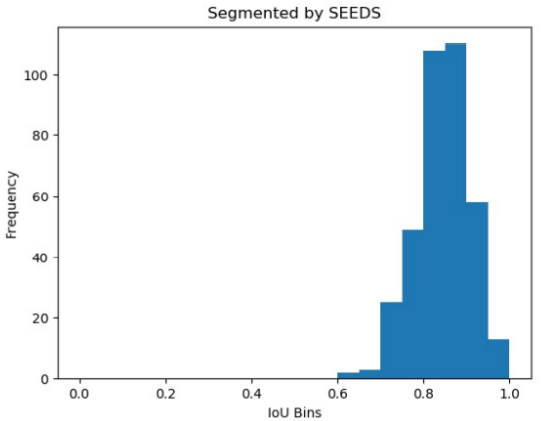


图 3-5 本文算法运用 SEEDS 超像素分割的 IoU 结果

由图 3-4 图 4.5 可知，两者对于本文的模型影响较小。虽然 SEEDS 论文里到其算法速度快于 SLIC，但是考虑其分割结果对于参数很敏感，并且两种算法在本文数据集上的时间表现并无太大差异，因此本文算法所用超像素分割算法为 SLIC。

### 3.5.3 算法时间复杂度对比

考虑到本文的应用场景是手术应用，因此本文对每个视频的追踪计时，部分结果显示如下表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 视频编号 | 时长/s | 输入 FPS | 本文算法耗时/s | 输出 FPS | TLST 算法耗时/s |
| 1 | 7 | 25 | 7 | 24 | 32 |
| 2 | 15 | 25 | 15 | 22 | 60 |
| 3 | 12 | 25 | 12 | 23 | 54 |

通过对比本文算法耗时以及 TLST 算法耗时可知，本文对其时间复杂度的改进起了明显的作用，并且对比视频的输入 fps 和输出 fps，可以看出本算法可以满足实时性需求。

### 3.5.4 跟踪效果展示

图3-6和图3-7分别展示了本文算法和TLST算法对于十二指肠乳头的跟踪效果，通过对比可以明显看到，TLST 算法的跟踪效果存在明显漂移的情况，例如图 3-7 的 1-3 图中，目标框的中心位置不是目标的中心位置，除此之外，TLST 算法追踪的目标框大小并没有随着目标的改变而改变，而本文的算法则做了相应的尺度的调整。

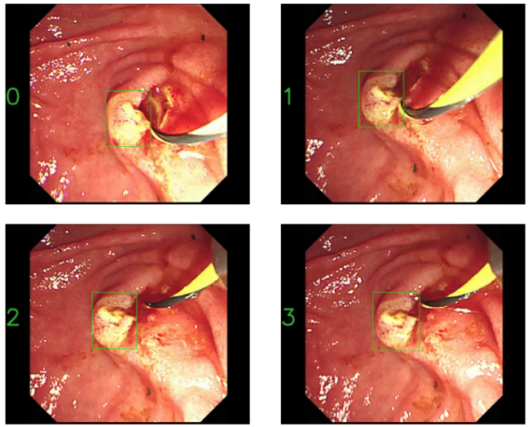


图 3-6 本文算法跟踪十二指肠乳头效果展示

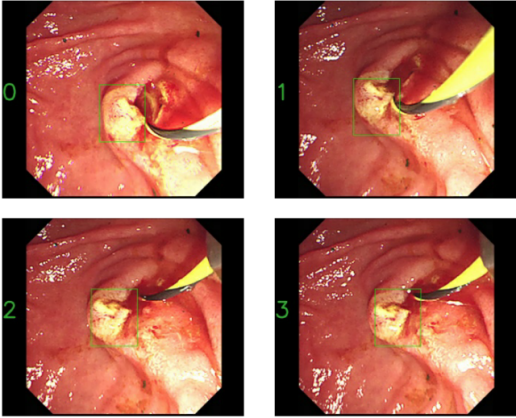


图 3-7 TLST 算法跟踪十二指肠乳头效果展示

图3-9、图 3-10、图 3-11、展示了本文算法在其他手术视频下对手术器官的追踪情况。对于不同手术场景，本文算法均取得了较为良好的表现，可见本文出的算法能够鲁棒、持续地对手术目标进行跟踪。

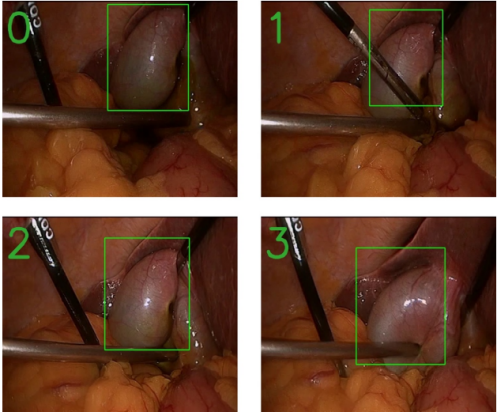


图 3-8 本文算法跟踪其他手术目标效果展示（1）

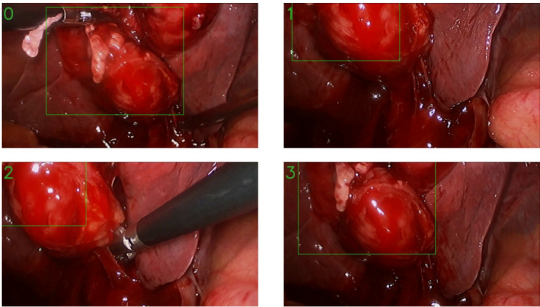


图 3-9 本文算法跟踪其他手术目标效果展示（2）

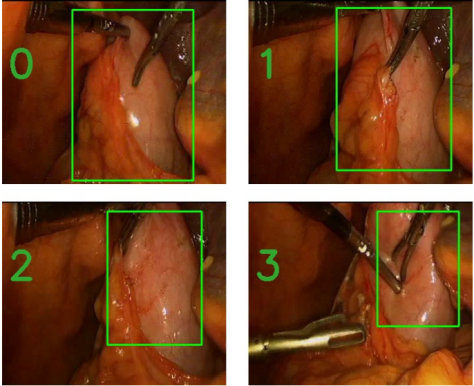


图 3-10 本文算法跟踪其他手术目标效果展示（3）

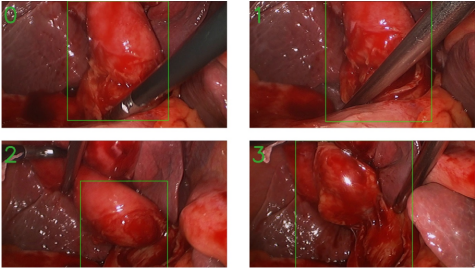


图 3-11 本文算法跟踪其他手术目标效果展示（4）

# 4 商业模式

## 4.1 竞争优势

本产品核心技术相比于其他产品针对手术中的医疗器械的追踪，采用针对手术中手术目标的追踪，兼顾实时性的要求。相比传统方式更为有效的解决了因器官出血、镜头模糊、器官变形、手术器械被遮挡等复杂的手术情况下出现的问题。

## 4.2 目标客户群体

客户目标为与医疗相关的企业，医院，医学院等

### 4.2.1 目标客户细分

(1) 外科手术需求

本产品帮助医生更精准完成手术，提高手术成功率。医生在手术过程中可能会面临较为复杂的情况。当下针对于手术中医疗器械的追踪，在手术过程中进行实时反馈较少，且对于不同原因造成的无法准确判断真实情况的处理有待提高。本产品可以兼容实时性的要求，并且在处理复杂问题上有较为良好的表现。

(2) 医疗器械需求

人工智能+机器人手术在不断发展，在未来会成为外科医生的得力助手，但目前为止其自主判断能力和精确度还可以更进一步。在手术过程中机械因器械自身反光或镜头模糊等原因可能造成判断失误，导致手术导航中抖动、漂移等现象发生。本产品更为有效解决复杂问题导致追踪出现问题这一情况。

(3) 教育领域需求

当下医学生培养周期长，培养成本较高，原因之一在于不熟练的临床手术会大幅提高病人的手术风险。风险之一是医学生或实习医生在手术过程中的复杂情况下难以快速且准确的找到最佳手术路径，需要医生有丰富的经验。本产品可为手术培训提供仿真模拟及教学资料，帮助医学生或实习医生更快熟悉手术过程，了解更多复杂情况的处理方式。

## 4.3 推广模式

本产品以线下推广为主，线上推广为辅。

### 4.3.1 线下推广

(1)与医院合作进行试点

和多个医院进行合作，在风险较小的微创手术上进行试用，让医生体验在缺少本产品辅助到使用本产品辅助带来的手术风险的降低和手术过程中的便利。通过产品收集手术数据，并反馈给试点医院，帮助医生更好复盘手术过程。

(2)与医学院进行合作，提供教学材料

医学院出资购买本产品，将本产品的使用带到实习课堂，并将其提供给有实习需求的医学生。同时医学院出资购买本产品所收集的手术资料，并利用本产品模拟手术下的复杂场景，帮助医学生积累经验。医学生在实习过程中将产品推荐给他人，提高产品知名度。

(3)与医疗器械制造商合作

找到相关医疗器械企业，例如生产内窥镜的厂家，希望与之达成合作，将本产品与其他产品相结合，拓展销售路径。

### 4.3.2 线上推广

创立相关网站，在网站中详细介绍产品优势，产品应用场景，并提供合作方式，

产品通过在医疗领域相关的论坛推广自己的产品，详细阐述产品优势及应用场景，并解答相关疑问，推荐本产品网站。

通过创立微信公众号分享产品相关信息，同时发布相关的医疗事件及进展。在微信公众号中提供网站快捷入口及合作方式，方便客户联系以及了解产品相关信息和近况。同时可以与试点医院的公众号进行合作，借助试点医院公众号推广产品。

## 4.4 营销策略

### 4.4.1 销售策略

前期主要进行产品的推广，寻求合作对象。与医院合作进行产品试点，提供设备以及维修服务，收集产品在实际手术中的数据，对产品进行改进增强，提供给更多医院。

中后期与医学院合作，提供产品以及数据资料，并提供相关人员进行技术指导和维修售后，收集相关数据。同时寻求与医疗设备制造商的合作，以售卖技术或合作研发的方式与相关企业合作，拓宽产品销售路径。

### 4.4.2 产品策略

1.完善全服务体系，做好售前、售中、售后服务，并可为客户提供附加服务，例如数据整理，器材培训等服务。

2.坚持研发，通过在使用中收集到的数据不断改进产品，优化产品性能，加强客户体验。后期加强研发依托本产品的新产品。

3.提供附加服务，例如辅助手术数据的解析，培训使用本产品的人员，利用本产品数据进行模型重建等服务。

### 4.3.3 营收策略

1.前期对医院提供设备以低价进行试点，以产品品质和性价比切入市场，占领市场份额；后期发展成熟后由价格优势转为品质优势，并向医学院出售产品以及教学资料。

2.提供“智选+”服务，通过收费为有需求的客户提供附加服务，满足不同客户的需求。

# 4 结论

腹腔镜手术其中两孔法后腹腔镜手术具有其创伤小，安全性高，疼痛小，恢复快等特点在医疗领域逐渐取代传统开放手术。同时因为其“微创”的手术性质，对医生的技术也提出了很高的要求。因此，无论是术前培训时提供一种辅佐培训，还是在术时基于实时成像为医生提供实时信息反馈，对手术中目标进行检测和追踪都十分有必要。实时性要求是手术场景下应用的算法的一个刚性需求。

本文提出的技术理论主要基于 TLST 算法，SLIC 超像素分割技术，以及 HOG、FHOG 特征，描述子和 DSST 尺度滤波器，提出了一种多模型融合的目标追踪算法，该算法融合了多个模型，先在原超像素跟踪算法的基础上提出了基于 HOG 特征的改进，改善了其追踪效果，同时基于图像相似度，提出了一种简单的跟踪方法，最后结合 DSST 尺度滤波器，对目标尺度做了调整和修改。最后基于人工标注的数据，我们对不同模型组合之间的结果做了交并比的结果检验，并且比较了各种情况下的算法时间，最终结果显示本文的算法在手术场景下的应用要比其他算法优秀，并且时间复杂度上处于可接受范围。本文提出了多模型融合的算法，并且该算法在 IoU 尺度下的检测表现良好，同时其时间性能上，由于我们针对 TLST 做了优化，因此处于可以接受的范围。因此不论是在跟踪效果还是时间复杂度上，本文的算法都比 TLST 算法优秀。在手术场景中对于手术目标反光、变形情况下，能够做到对目标的鲁棒性检测，实现对手术目标实时的监测与跟踪，能减少手术导航中抖动、漂移等现象，将会使腹腔镜在手术场景的应用效果更好。

本研究提出来相应的理论，使得研究结果具有较好的实用性和推广应用价值。

# 参考文献

[1] 陈兆瑞,赵子健.计算机辅助的微创手术工具跟踪算法综述[J].中国数字医学, 2016, 11(9):83-87.

[2] 葛鑫,邬小玫,汪源源,et al.用于微创手术导航的六自由度电磁跟踪方法[J].航天医学与医学工程, 2011, 24(3).

[3] Hulke U., Gupta A. Single camera based motion tracking forminimallyinvasive surgery[A]. 22nd Mediterranean Conference on Control and Automation[C]., 2014:356–361.

[4] Du X., Allan M., Dore A., et al. Combined 2D and 3D tracking of surgical instruments for minimally invasive and robotic-assisted surgery[J]. International journal of computer assisted radiology and surgery, 2016, 11(6):1109–1119.

[5] Zhao Z., Voros S., Weng Y., et al. Tracking-by-detection of surgical instruments in minimally invasive surgery via the convolutional neural network deep learning-based method[J]. Computer Assisted Surgery, 2017, 22(sup1):26–35.

[6] Shaharan S., Ryan D. M., Neary P. C. Motion Tracking System in Surgical Training[M]. , Travieso-Gonzalez C. M., Motion Tracking and Gesture Recognition[C]., 2017, https: //doi.org/10.5772/intechopen.68850.

[7] Ren X., Malik J. Tracking as Repeated Figure/Ground Segmentation.[A]. CVPR[C]., 2007，1:7.

[8] Yang F., Lu H., Yang M.-H. Robust superpixel tracking[J]. IEEE Transactions on Image Processing, 2014, 23(4):1639–1651.

[9] Wang J., Yagi Y. Many-to-many superpixel matching for robust tracking[J]. IEEE transactions on cybernetics, 2014, 44(7):1237–1248.

[10] Cai Z., Wen L., Lei Z., et al. Robust deformable and occluded object tracking with dynamic graph[J]. IEEE Transactions on Image Processing, 2014, 2323(12):5497–5509.

[11] Yeo D., Son J., Han B., et al. Superpixel-based tracking-by-segmentation using markov chains[A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition[C]., 2017:1812–1821.