西安交通大学

硕士学位论文

基于隐形状模型的手掌识别

学位申请人：余静

指导教师：梅魁志 副教授

类别（领域）：工程硕士（控制工程）

2014 年 05 月

Hand Detecting Based on Implicit Shape Model

A thesis submitted to
Xi'an Jiaotong University
in partial fulfillment of the requirements
for the degree of
Master of Engineering

By
Jing Yu
Supervisor: A.P. Kuizhi Mei
(Control Engineering)
May 2014

论文题目：基于隐形状模型的手掌识别
类别（领域）：工程硕士（控制工程）
学位申请人：余静
指导教师：梅魁志 副教授

摘 要

智能家电正在改变人们的生活，其人机交互的方式越来越自然，而手势控制是热门的发展方向之一。手掌识别作为手势识别的基础，对其研究有应用价值。本文针对手掌检测的特点，研究并实现了一种基于局部特征的隐形状模型算法。主要工作内容如下：

首先，对隐形状模型的原理进行了阐述，重点分析了码本的组织结构和构建过程。在训练阶段，对局部特征所构成的特征池进行聚类，以得到码本的入口；将局部特征相对于目标中心的位置信息作为码本入口的空间分布记录。在检测阶段，从测试样本提取局部特征，查询、匹配码本入口和相应的空间分布记录；与当前特征的坐标结合，对目标坐标作出预测、投票，并采用均值漂移模型估计法统计票集得到最终的目标坐标值。

其次，设计了针对手掌检测的隐形状模型。在样本库的设计方面，将手掌的姿态按偏左、偏右、朝上划分为三个子类，从原始图片中分割各子类手掌并叠加到随机背景中，创建了不同姿态的训练集和测试集。三个子类的训练/测试集容量分别为：480/124 张，293/128 张，680/160 张。在特征的提取方面，结合隐形状模型规模适应的特点，采用具备规模不变性的 DoG 特征点提取算子，并根据手掌颜色稳定的特点，设计了灰度特征描述子。

最后，结合手掌多姿态的特点和肤色特征，对隐形状模型作了相应扩展。训练阶段对三个子类分别训练得到子类码本，检测阶段将子类码本进行了组合，采用组合码本完成多姿态手掌的检测。此外采用肤色掩模缩小检测范围，并对码本进行了降噪处理。对各个优化方法，测试了优化前后的性能，并作了对比分析。

关 键 词：手掌识别；特征；隐形状模型；肤色；码本组合

论文类型：应用研究

Title: Hand Detecting Based on Implicit Shape Model
Professional Fields: Engineering (Control Engineering)
Applicant: Jing Yu
Supervisor: A.P. Kuizhi Mei

ABSTRACT

Our daily life has been improved by intelligent household appliances, and the way human interacting with machines is becoming more and more natural. Interaction based on gesture recognition is one of the hottest and most promising area. As the basic step of gesture recognition, the research on hand dectction is important. In view of characteristics of hand detection, a local shape information based Implicit Shape Model is proposed in this thesis. The main work is as following:

Firstly, the principle of Implicit Shape Model (ISM) was illustrated. The structure and constructing procedure of codebook were primarily analyzed. In training step, the local features in the feature pool were clustered to generate entries of the codebook; the coordinates of local features relative to object centers were recorded as spacial information of codebook entries. In detecting step, local features were extracted from the test sample, then codebook entries and corresponding spacial information were queried and matched; matched information were combined with the coordinates of local features to predicit and cast votes for object positions. The Mean-Shift Mode Estimate method was also used to count votes and generate the final position of the object.

Secondly, the Implicit Shape Model for hand detection was designed. For sample libraries, hand gestures were divided into three subclasses: leftward, rightward and upright ones. Hands were segmented from original pictures and put onto random backgrounds. Thus training set and test set of different subclasses were created. The sizes of training sets and test sets of three subclasses were: 480/124,293/128,680/160. During the stage of feature extracting, considering the scale adaptability of ISM, the DoG feature point extractor was selected. Besides, as color of hands is stable, the patch feature descriptor of gray value was used.

Finally, corresponding extensions were added to ISM according to the feature of skin color and various shapes of hands. In training stage, the codebooks of three subclasses were obtained respectively. Then in detecting stage, these codebooks of subclasses were combined together to accomplish the detection on various gestures of hands. Besides, skin mask was used for decreasing the detection area, while denoising was done for the codebooks. At last, for each optimization method, the performances of ISM before and after optimizations were compared and analyzed.

KEY WORDS: Hand detection; Feature; Implicit shape model; Skin color; Codebook combination

TYPE OF THESIS: Application Research

目　录

CONTENTS

1 绪论

1.1 课题背景

如今智能家电的发展正在使人们的生活越来越便利和有趣，但随着这类设备的功能变得强大和丰富，人们对其易用性和交互的自然性也提出了更高的要求。因此，高级人机接口有着广泛的应用前景。其中，手势控制成为人机交互的热门方向[1,2]。对于智能电视而言，手势控制即人通过各种手势发出指令，电视机通过摄像头捕捉手势并获取指令。从而人可以摆脱摇控器或其它硬件的束缚，获得更加轻松的娱乐体验。

在整个手势识别系统及其上层应用[3]中，手掌的检测是实现的基础。只有实现了准确定位手掌这个前提，才能有效地进行手势的识别，从而获取用户的意图。本文以手掌检测作为研究重点，以识别和定位手掌为目标。

手掌检测的任务是对于一张原始图像，能够判断图像中是否存在手掌，若存在则标注手掌区域。在智能控制发展的早期，数据手套由于有精准实时的传感器进行信息传递，在手掌检测或人手控制方面应用比较广泛。2009 年微软推出的 XBOX360 配套设备 Kinect，除了可依靠相机捕捉三维空间中用户的运动信息外，还附加了其它功能，但价格较昂贵。近年，利用单目摄像头检测手掌的研究开始兴起[4]，此方法价格较低廉，若能达到较高的性能，则可大大减少硬件设备的成本。

使用摄像头检测手掌，首先需要对应用场景的图像进行分析。由于手掌的灰度分布比较均匀，因此某些在人脸识别上取得成功的特征不能表现手掌的特点；再如手掌是非刚性的，它的形态比较多样，在统计学习中用表现形态的特征（如 Haar 和 HOG）也达不到理想的检测率，或者会导致较大的误检率；另外，其应用场景比较多样，光照变化大，背景也比较复杂，并且手掌自身的尺度范围大。本文结合这些特点，经过相关文献调研，尝试基于局部信息的检测或分类方法，提出用隐形状模型作为手掌检测的训练和检测算法。

基于统计学习的窗口检测算法因其海量学习的特点，在某些目标识别方面获得了较大的成功。如 Adaboost 算法，是基于样本权重更新和弱分类器投票的统计学习算法。该算法在 Haar 特征和级联的扩展之后，在人脸识别领域取得巨大成功，现已应用于许多实际系统中。

应用于人脸的 Haar 特征是典型的统计特征。可以利用积分图快速计算指定 Haar 特征的特征值。学习时，需要对不同的坐标和规模的特征进行尝试，即从海量的特征池中挑选最具分辨能力的特征。因此，训练的过程非常漫长。

检测时，由于统计学习只能学习大小一致的样本，故需要用与训练样本同样大小的扫描窗口对测试图进行扫描，扫描窗口的重叠造成了冗余的计算。另一方面，由于目标的大小未知，需要对测试图进行多次缩放，每次缩放后需要重新计算积分图，以便计算特征值。为了方便积分图和特征值计算，往往将样本置于矩形框中，使得框内

除了样本的空隙包含背景。当样本为手掌时，由于手指的间距变化大，造成空隙较大，即样本在框内不是很“饱和”。多次出现的背景噪声会被学习成特征，即一些特征其实是背景的特征，而非手的特征。

另一方面，由于手掌的形态多样，难以找到一种统一的形态特征，将手掌与背景区分。要达到较高标准的检测效果时，训练得到的分类器将比较庞大。本文用基于 OpenCV 的 Adaboost 算法，采用标准 Haar 特征，用 2900 张正样本和 6000 张负样本进行训练，发现在达到指定级联数量时，误检率并没有训练到预期的水平。若不限制级联数量，则训练几乎将陷入无限循环。经测试，20 级的分类器，正检率为 93.2%时，误检为 21.1%，此时级联 Adaboost 的弱分类器达到 682 个。

鉴于手掌检测的特点，本文采取了新的思路。手掌虽然整体形态多样，但局部特征稳定。若能够从局部细节特征检测出整体目标，则可以绕开寻找整体特征，且具有形态鲁棒性。而隐形状模型（Implicit Shape Model，后文简称 ISM）模型在多角度、复杂背景和多类识别方面取得了成功，如行人、动物、汽车等，不仅对于非刚性目标表现出鲁棒性，而且可以同时进行多个目标的检测。它的优势与待解决的问题相匹配，很值得借鉴和尝试。

基于隐式形状模型的手掌检测方法把手掌看成由许许多多的图片块构成，这些图片块可能对应手指，也可能对应手掌，它们往往是具有一定特点的部位。

在训练阶段，主要任务是建立特征字典，即码本。一个码本有若干入口，每个入口有若干记录。码本的入口为通过聚类得到的目标特征集合，一个类表示一类特征，类中心作为该类特征的平均值，可称作该类特征的“代表”。入口后方的记录为该类特征在训练样本中出现时的相关信息记录，即把该特征每次出现时相对于目标中心的位置记录在该特征后。

码本在检测阶段两次发挥检索的功能。在检测阶段，第一步是检测兴趣点，提取兴趣点周围的图片块，再从码本中找到匹配项，此时码本的多个入口发挥了特征检索的功能，即对于测试图提取出的特征，查询其在码本中是否存在。第二步根据当前的特征位置和匹配的码本特征相对于目标中心的位置，推测出当前待检测图片中的目标中心位置，此时码本入口的记录发挥了目标位置参考的功能，即对于测试图提取出的特征，参考码本中记录的特征与中心的相对位置，推测出测试图中目标的位置。

隐式形状模型从训练手掌的图片块中提取特征，构建码本并将“代表”特征在训练样本中的每次出现作记录。检测阶段亦从检测图片的图片块中提取特征，对所有特征两次检索码本，可以得到多个对于目标中心位置的推测。后期进行均值漂移等数学运算，便可得到最终目标中心的假设。

1.2 国内外研究现状

目标检测是计算机视觉的一个重要的研究分支，有许多具有启发意义的研究成果。目标检测的研究对象以人脸[5-8]、行人[9-11]、车辆[12,13]居多，手掌检测常常作为手势检

测的一部分来研究[14-16]。这些研究从不同种类的分类器展开和深入。

在分类器的类型方面，有统计学习和基于局部信息两种不同的学习与检测方法。

基于统计学习的分类器，如 Adaboost，将若干具有弱分类能力的分类器组合起来，组合形成强分类器。1995 年，Freund 和 Schapire 提出了 AdaBoost 算法[17]，该算法源自 Boosting 思想，将寻找强分类特征的难题分解，用较容易的寻找弱分类特征的方式来实现，并在数学上证明了其可靠性。每训练一个弱分类器时，在目标的整体范围内，选择对正、负训练样本集最具分辨性能的特征。对于有稳定特征的目标，Adaboost 算法的表现良好，Adaboost 算法在人脸识别等领域获得了广泛的应用。

基于局部信息的模型[18,19]，如隐形状模型（Implicit Shape Model[20]，简称 ISM），认为目标是由细节组合而成。该模型不是直接寻找对正负样本有分辨性的特征，而是从目标中抽取局部特征，通过聚类的方式使反复出现的典型特征作为代表特征。另一方面，由于该方法只利用局部特征，识别目标时只需要对局部特征进行匹配，来自不同的训练样本的局部信息可以被结合到一起，以概率投票的形式对目标位置进行预测。因此对整体形态有变化的目标具有鲁棒性。此外，ISM 算法可以利用测试样本与训练样本的匹配信息将目标从背景中分割出来。

隐形状模型的雏形[21]于 2003 年提出，当时没有明确的定义，且方法被限制于只能用于检测最多包含一个目标的场景。Bastian Leibe 最早于 04 年提出隐形状模型的定义[20]，此后多位学者对该模型进行了研究和扩展。该算法起初的研究目标是检测训练样本中未出现的目标类型。隐形状模型实质上是局部形状与整体关系的概率模型。它以码本记录局部结构特征，隐形状模型记录码本入口出现的位置。

该方法与采用目标原型[22]或者相似的目标[23]的组合建立模型的思想相似，主要区别在于组合不是发生在整体样本目标之间，而是发生在局部图片块之间，因此本方法更加灵活，且 ISM 用更少的训练样本即可学到可能的目标形状。例如对于车辆识别，该方法不需要在训练样本中出现所有的实例，而可以将一幅训练样本中的车头信息和另一幅训练样本中的车尾信息结合起来。由该车头和车尾实例对同一个目标位置的假设投票，模型可以识别训练样本中没有出现过的车头与车尾组合的“新车”。

Leibe 的实验中对汽车、行人和动物均进行了测试[9]。结果表明，ISM 算法对有遮挡和多姿态的目标有较好的检测能力[24]。图 1-1 展示了对普通场景下汽车的检测性能，可以看出，基本的 ISM 算法（标注了“Without Verification”的黑色 ROC 曲线）已经表现优良。

图 1-2 展示了 ISM 算法对于被遮挡目标的检测性能，实验目标为牛。横坐标为目标的可见比例，纵坐标为检测率（标注了“Recall”的蓝线）和精确率（标注了“Precision”的红线）。可以看出，对于少量的遮挡，ISM 可以保持较高的检测性能。

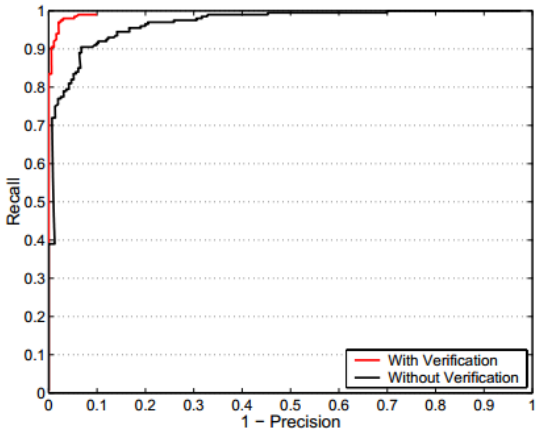


图 1-1 ISM 算法检测 UIUC 汽车样本库的结果

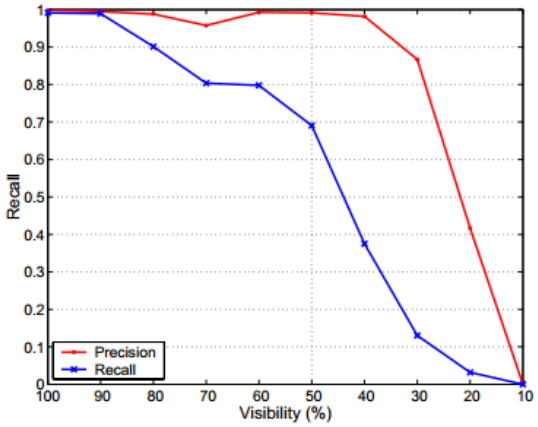


图 1-2 ISM 算法对于遮挡目标的检测性能

ISM 算法将目标识别和前景-背景分割视作两个相互促进的过程，最终目标都是使识别结果尽可能精确。Seemann[25]用 MDL 检验方法，根据分割目标的结果，对检测结果进一步筛选，挑选描述信息量最小的假设组合，从而消除误检。

1.3 论文的主要研究内容

本文的研究工作是探索一种适用于手掌的检测方法，以适应用单目摄像头进行智能家电手势控制的使用需求。本文通过分析目标的特点，再对应 ISM 算法的性质，对隐形状模型识别手掌的性能展开了研究。除了利用 ISM 算法的形态鲁棒性，后期还结合手掌的肤色特征缩小检测范围。此外，将不同参数下的检测性能进行对比。

第一章为绪论。介绍了手掌检测的研究背景和意义，对国内外研究现状作了概括和分析，并对论文的内容安排作了陈述。

第二章对隐形状模型的算法流程和原理进行了介绍。ISM 算法步骤分为训练和检测两部分。本文将训练和检测的流程进一步划分，将码本的产生细化为码本入口和码本记录的生成两步，将目标搜索细化为概率投票和假设搜索两步。

第三章介绍了用隐形状模型检测手掌的实现细节。首先介绍训练与测试样本，随后说明了评价算法的指标。然后对采用的特征和算法流程及主要数据结构进行介绍。

第四章介绍了根据手掌特点对模型的应用及优化。首先根据肤色特点，提出用肤色模型进行预处理。然后进行码本的组合，以完成多姿态检测。

第五章对实验结果进行了分析。对有无肤色预处理和组合码本与单个码本的实验结果，进行了对比分析。此外，对于码本降噪的效果和目标分割的效果也作了分析。本文主要以 ROC 曲线呈现实验结果。

第六章为总结与展望。从实验结果分析、总结，对整体优化结果得出结论。并通过分析隐形状模型的优缺点，结合手掌检测的需求，展望后期工作的改进方向。

2 隐形状模型介绍

本章分训练和检测两个阶段，阐述了隐形状模型理论。本章的核心概念为码本。训练即码本的构建过程，检测即码本的使用方法。具体来说，码本的构成分码本入口的生成和特征记录两个阶段，相应地，码本的使用分匹配入口和参考记录两个步骤。最后利用码本记录的相关信息，推测目标中心的位置，并在三维投票空间投票，经过统计搜索算法，得到目标中心位置的假设。

2.1 特征点及特征

特征值为整个隐形状模型的基础数据。无论是训练还是检测，都将从图像中提取特征作为第一步。从单幅图像提取特征信息，需先提取特征点，以确定特征区域，再对特征区域的像素值进行相关运算，得到特征值。

特征点是在局部区域有较大强度的数学属性的像素点。对于目标检测而言，需要提取一些带有较大区分度和富含信息的区域，作为训练和检测的特征。而特征点邻域往往包含了轮廓的关键信息，与目标特征区域有高度的重合性。采用特征点检测子来检测特征点，减少了要处理的数据量，并使得在不同实例中的相似区域可以被提取出来。此外，特征点具有旋转不变性和不随光照条件改变而改变的优点。

手掌的特征点往往出现在手指和指间，常用的特征点算子有 DoG、Harris[26]、Laplace 等。使用 Hessian-Laplace 特征点检测简单背景下的手掌的结果如图 2-1。

图 2-1 手部特征点检测示意图

图 2-1 中，十字形的中心表示特征点，圆圈为特征点提取的规模。可以看出，特征点提取的规模并不一致。提取特征点时，可在多个规模上搜索，以局部区域数学属性最大时的区域半径作为特征点的提取规模。

特征描述子，是描述图像块特征的算子。特征点周围的局部区域被提取出来后，可以对该区域进行计算，得到其特征描述值。Leibe 和 Schiele[27]使用原始的灰度值作为特征值描述子。而对行人进行不同特征的检测结果表明，基于形状上下文的特征描

述子，对行人的检测更有效[25]。计算特征值的步骤如下：

（1）根据提取特征点的范围，提取相应大小的图片块；

（2）对图片块区域的像素矩阵进行计算，得到特征值。

对手部提取特征点后，相应的图片块示例如图 2-2 所示。

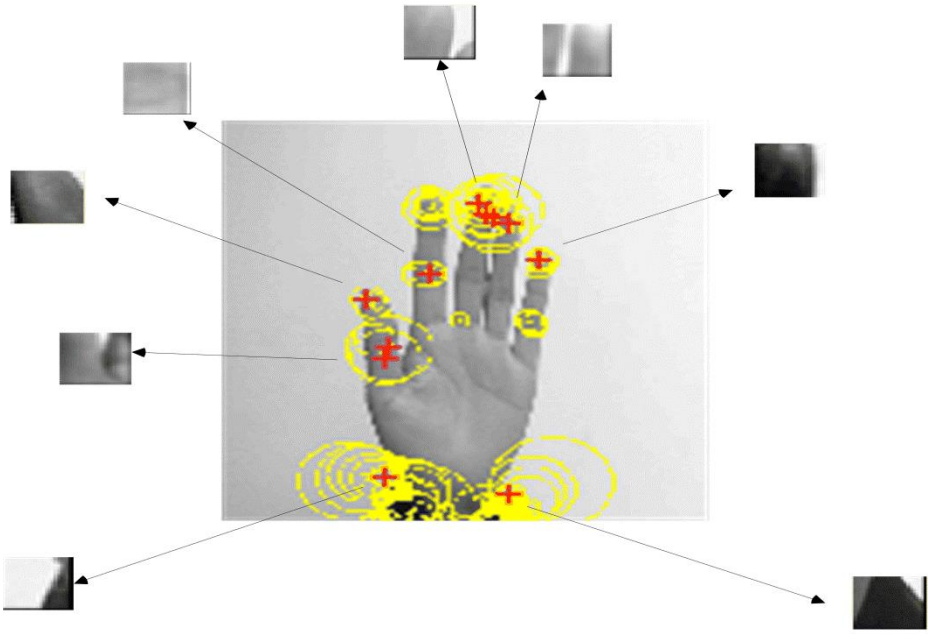


图 2-2　手部提取图片块示意图（图片块已缩放到相同规模）

图 2-2 中已将图片块通过插值法缩放到一致的尺寸。图片块是特征值的形式化表示，可以用图片块直观地展示特征之间的关系。

2.2 码本的结构

隐形状模型的核心概念为码本。训练阶段得到码本，检测阶段码本作为检测器。码本的结构示意如图 2-3，其主要内容分为两部分：码本入口和码本记录。

一个码本有多个入口，每个入口有若干记录。码本入口代表了该码本所表达的对象的局部特征，而入口后方的记录表示该局部特征每次在训练样本中出现的位置信息。一条码本记录表示该码本入口所代表的特征出现一次时相对于目标中心的坐标。设目标中心坐标为(c_x,c_y)，特征坐标及尺寸为(l_x,l_y,l_s)，则相对坐标为(c_x-l_x,c_y-l_y,l_s)。

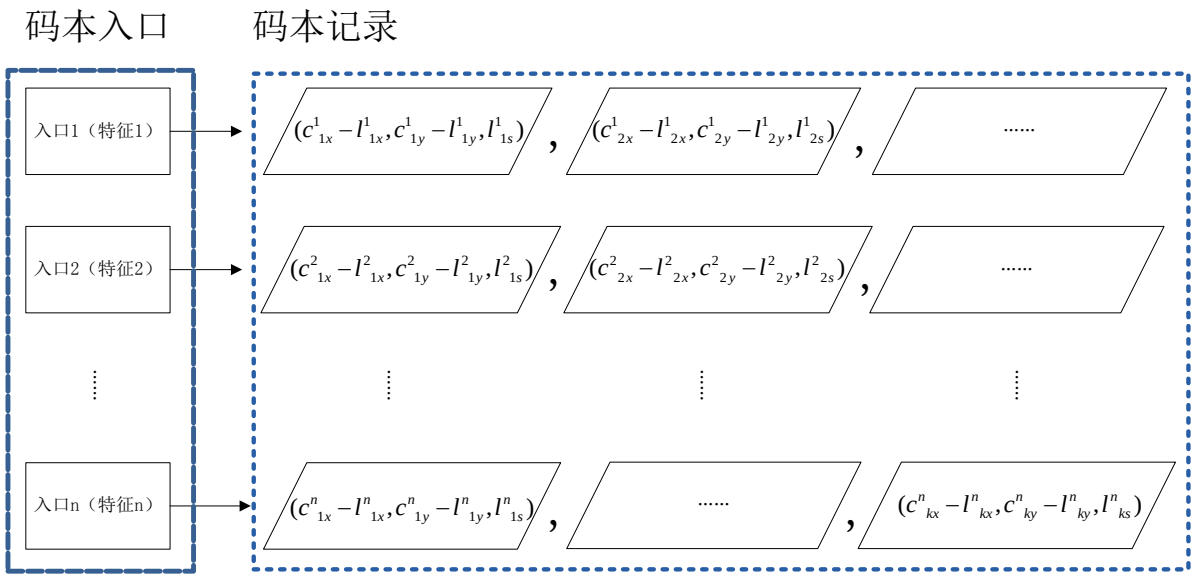


图 2-3 码本的结构

2.3 码本入口的生成

从训练集中提取特征点和特征后，把所有特征值都放入特征池。要从这个庞大的特征池中提取关键的和重复率高的信息，需要进行聚类。许多特征来自目标的同一部位，它们被认为是相似的特征。本文采用非监督的方式进行聚类，使视觉上相似的图片块对应的特征聚到一起。Leibe 和 Schiele[27]评估了一系列聚类算法，认为就目标检测而言，合成聚类的性能最优。

合成聚类法一开始把每个特征视为一个类，然后两两计算特征的相似度。每次迭代将最相似的类合并。当达到指定条件时，终止迭代。

Leibe 指出，合成聚类由于需要计算元素之间的相似度矩阵，运行时间的要求是比较高的[21]。这限制了可以处理的特征数量。本文采用该文献中使用的一种更高效的合成聚类法[21]。该算法基于互近邻对（Reciprocal Nearest Neighbor pairs，RNN pairs），即若 p 和 q 是一对互近邻点，那么 p 是 q 的最近邻点，q 亦是 p 的最近邻点。

数学上已证明，将 RNN 对的两个元素合成为一类，比 RNN 对中的任意一个元素与其它元素合成的效果要好。全局最近邻对是 RNN 对集合中的一个元素，寻找一个 RNN 对，显然比搜索全局最近邻对容易得多。因此，Leibe 用寻找 RNN 对并聚类的策略，代替了寻找全局最近邻对再聚类的传统策略。

RNN 聚类算法以构建 NN 链的方式寻找 RNN 对。NN 链即最近邻元素链，其构建方式是以随机点为起点并加入链尾，每次寻找链尾元素的最近邻元素，若最近邻元素已在链表中，则找到了一个 RNN 对。对于 RNN 链的构造是一个不断收紧的过程，当链尾元素已达到最紧时，停止构造，将链尾的两个元素作为一个 RNN 对。表 2-1 为基于 NN 链的 RNN 合成聚类算法。

表 2-1　RNN 聚类算法

初始化：设V为特征空间的点集，随机点$v \in V$。RNN 链名为L。R为聚类空间的临时类集。*last* 为 RNN 链中元素的标号，*lastsim* 为 RNN 链中元素的相似度。

$last \leftarrow 0; lastsim[0] \leftarrow 0; L[\text{last}] \leftarrow v \in V; R \leftarrow V \setminus v$

```
1   while R ≠ ∅ do
2       (s, sim) ← getNearestNeighbor(L[last], R)  //获取 L[last] 的最近邻点 s
3       if sim > lastsim[last] then
4           last ← last + 1
5           L[last] ← s; R ← R \ {s}; lastsim[last] ← sim
6       else
7           if lastsim[last] > t then
8               s ← agglomerate(L[last], L[last − 1])  //将 L[last]和 L[last-1]聚为一类
9               R ← R ∪ {s}; last ← last − 2
10          else
11              last ← −1
12          end if
13      end if
14      if last < 0 then
15          last ← last + 1
16          L[last] ← v ∈ R ;  R ← R \ {v}
17      end if
18  end while
```

表 2-1 中相似距离的计算方法取决于特征描述子。对于灰度值描述子，采用归一化灰度相关系数（Normalized Grayscale Correlation，NGC）；对于上下文描述子，采用欧氏距离。表 2-1 中最终将所有元素聚为一类，其过程是一个自底向上构建二叉树的过程，每次将两个结点作为子结点构成上层的父结点。而实际运用中，可以通过控制最大类的个数或迭代次数，使聚类停止在预期的阶段，即二叉树停止在预期的层次。

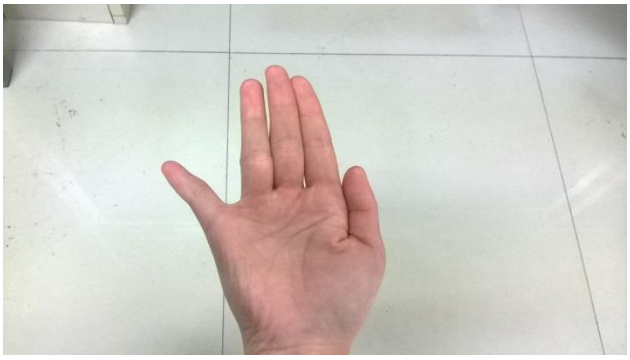
聚类的过程即对特征的抽象。在特征空间，特征值相近的元素聚为一类。而特征值相近意味着特征图片块相似，因此特征池经过聚类后，视觉上相似的特征聚集为类。类中心作为一个类的平均值，可以认为是该类的元素的代表，即该特征的代表。通过聚类，目标的特征以不重复的方式表现出来。类的规模越大，说明该特征出现得越频繁，则该特征是目标的典型特征。将聚类中心作为特征记录在码本中，作为码本的入口，则码本入口代表了该类目标的特征的集合。

假设最终得到的类的个数为 N，则这 N 个类的中心可以作为码本的 N 个入口。它们代表了训练样本集中多次出现的特征。这里只对特征值聚类，即图像块的坐标不影响聚类的结果，这对于样本的形态多样性有很好的鲁棒性。

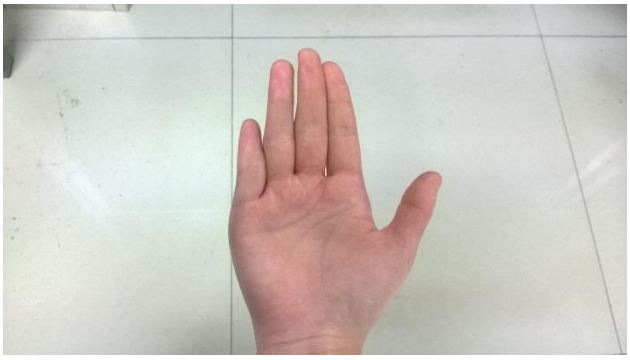
此外，对于聚类的结果，可以设定类规模阈值 $N_{\min}$，对于元素个数少于 $N_{\min}$ 的类，可认为是图像中的干扰特征而非目标特征，直接去除，而不作为码本入口。

2.4 空间分布学习

码本的入口，代表了一类目标的特征。而测试一张图时，若仅与入口匹配，没有其他信息，则无法判断目标的具体位置。因此，除了码本入口记录特征值外，还需要记录特征在目标中所处的位置。



(a)训练样本 1：小指偏离



(b)训练样本 2：拇指偏离

(c)测试样本：小指和拇指偏离

图 2-4　ISM 算法对于形状的鲁棒性示意

实际上，对于可变形目标，特征部位相对目标中心的分布是不固定的。隐形状模型的方法不需要训练样本中出现所有特征分布的组合。如图 2-4(a)在一张样本中出现

小指偏离，图 2-4(b)在另一张样本中出现拇指偏离，那么对于一张小指和拇指均偏离的测试图片（图 2-4(c)），即使样本集中没有出现这种姿势的手，隐形状模型也可以结合样本 1 与样本 2 的信息，而具有对测试样本图 2-4(c)的识别能力。

以图 2-4 为例，聚类阶段将小指和拇指两个特征都记作码本的入口。在训练阶段将小指相对于手掌中心的坐标和小指的特征区域规模以$(c_x - l_x, c_y - l_y, l_s)$的形式记录在小指特征入口的后方，即记录小指每一次出现时的坐标信息，对拇指亦然。则在检测阶段，对图 2-4(c)，其提取的小指和拇指特征分别与码本入口匹配，并可参考码本记录推测出手掌中心的位置。

通过聚类得到的码本入口，是目标特征值的均值，而此时特征的空间信息并没有记录。训练阶段主要是记录特征在空间上的分布，以作为检测阶段推测目标中心位置的参考。为保证记录的一致性，训练样本中目标的规模必须相同，记作(w_{sample}, h_{sample})。对训练图片再一次进行目标区域特征点和特征值的提取，并与码本的入口进行匹配。对于相似度达到阈值的匹配项，将这个特征的特征值和与目标中心的相对坐标记录在匹配的入口后方，认为是该入口所代表的这类特征的一次出现（Occurrence，简记 Occ）。训练算法如表 2-2。

表 2-2　ISM 算法的训练流程

初始化：设码本名为 C。统计特征的出现 Occ。// Occ[i]表示第 i 种特征的所有出现记录的集合。

1　对于所有的码本入口c_i

2　令$Occ[i] = \varnothing$

3　对所有的训练图片

4　　设(c_x, c_y)为目标中心

5　　对于特征区域$l_k = (l_x, l_y, l_s)$及其特征值f_k

6　　对于所有的码本入口C_i

7　　如果$sim(c_i, f_k) \geq t$

8　　$Occ[i] \leftarrow Occ[i] \cup (c_x - l_x, c_y - l_y, l_s)$

经过训练，针对某一目标隐形状模型就产生了，记作$ISM(C) = (C, P_c)$。其中，C表示码本，P_c表示空间概率分布。$ISM(C)$记录了在训练过程中，码本的入口出现过的位置，即目标的特征及其可能出现的位置。码本作为目标的特征集合，应当全面地记录目标的特征及其可能出现的位置。在理想状况下，码本不仅全面地记录了目标的特征和空间信息，而且不包含与目标无关的噪声特征。

2.5 假设的产生

经过对样本的特征提取，对特征池聚类生成码本，样本重新匹配码本入口，训练阶段的结果——隐形状模型$ISM(C)$已经产生了。下文介绍利用该模型进行目标检测的

算法。

简单概括之，检测就是对测试图片提取的特征，查询码本中与其匹配的特征相对目标中心的位移，再根据数学统计算法得出关于目标中心位置的假设。

检测的过程如下：首先，用与训练阶段一致的特征点检测算子和特征描述子，对检测图片进行特征提取，如图 2-5 左上角示意图；然后，将提取到的特征与码本的入口进行匹配，此为第一次查询码本，如图 2-5 第 1 步，左侧中间的图为匹配的码本入口；参考匹配的特征的坐标记录，在哈夫投票空间[28-30]按相应的权重投票，此为第二次查询码本，如图 2-5 第 3 步，左侧下图为投票示意；最后，对投票空间的所有票进行均值漂移统计，将局部最大值作为目标位置的假设。

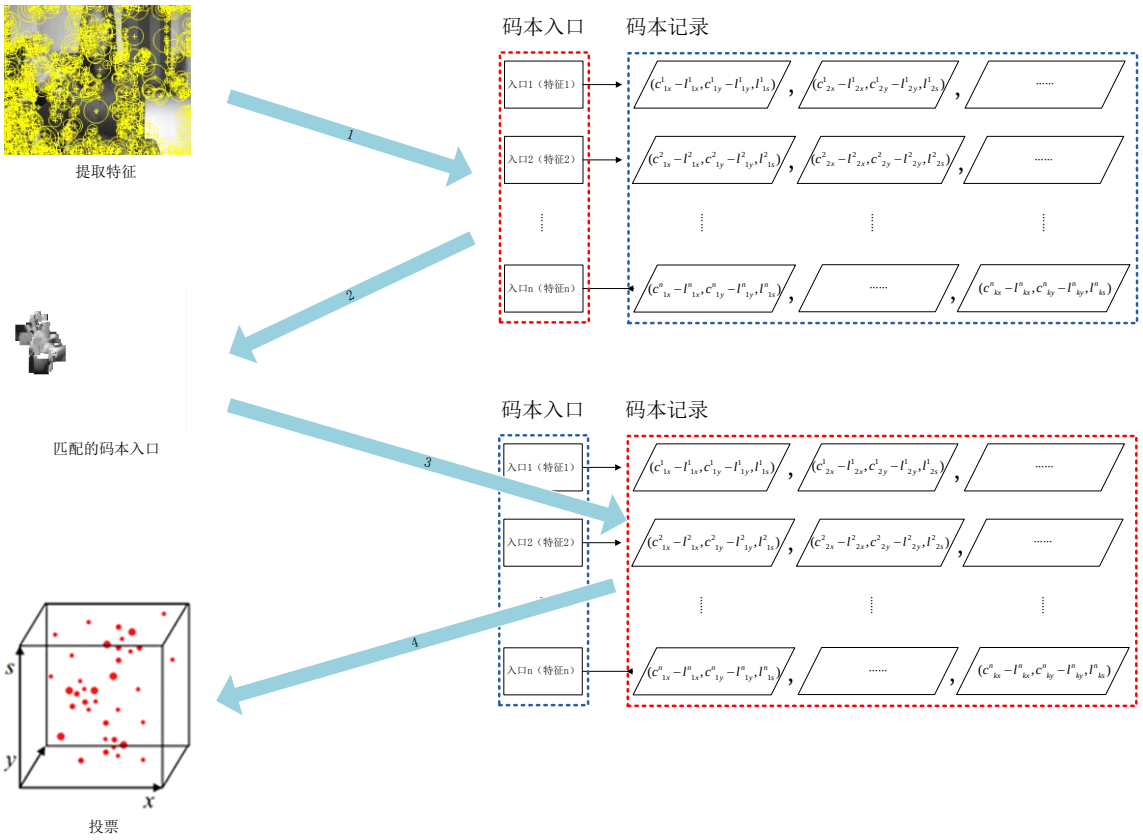


图 2-5　检测阶段两次参考码本

2.5.1 概率哈夫空间投票

投票程序是以概率模型[20,21]为基础的。设 f 为坐标 l 处观察到的特征值。将 f 与码本入口进行匹配后，有若干入口与之相似。设 c_i 为码本的第 i 个入口，则 f 与 c_i 匹配的概率为 $p(c_i \mid f,l)$。而对每个 c_i，都可以根据前面学习的空间概率分布 P_c 查询得到 $p(x|c_i,l)$，从而对位置 x 进行投票。该票的意义是认为目标中心有 $p(x|c_i,l)\ p(c_i \mid f,l)$ 的概率出现在 x 处。用 $p(x|c_i,l)$ 与 $p(c_i \mid f,l)$ 求取中心在坐标 x 处的概率的公式化表达为：

$$p(x|f,l)=\sum_i p(x|f,c_i,l)\,p(c_i\mid f,l) \tag{2-1}$$

由于中心坐标 x 的取值只参考当前特征坐标 l 与码本记录中的特征与中心的相对位置，故 x 只与码本入口 c_i 和坐标 l 有关，与特征值 f 无关，$p(x|f,c_i,l)$ 可以简化为 $p(x|c_i,l)$。同理，特征与码本入口 c_i 匹配的概率只与特征值 f 有关，与特征的坐标 l 无关，故 $p(c_i\mid f,l)$ 可以简化为 $p(c_i\mid f)$。公式（2-1）简化为

$$p(x|f,l)=\sum_i p(x|c_i,l)\,p(c_i\mid f) \tag{2-2}$$

式（2-2）等号右侧第一项是在指定特征 l 和其解释 c_i 的情况下，对目标可能出现位置的一个哈夫投票。第二项是特征 f 和码本入口 c_i 的匹配度。

对目标中心进行投票时，目标的相对尺寸作为投票空间的第三维[27]。具体到特征，则目标的尺寸体现为特征描述子的图片块尺寸，一般用特征点为中心的半径表示。若 (x_0,y_0,s_0) 为投票空间中的一个点，则 (x_0,y_0) 为图像中的二维坐标，s_0 为当前特征点的提取半径与参考特征点的提取半径的比值。

若码本的一条特征的记录相对中心的位移为 $(x_{occ},y_{occ},s_{occ})$，而测试图片的特征坐标为 $(x_{img},y_{img},s_{img})$，则投票的中心位置为

$$x_{vote}=x_{img}-x_{occ}(s_{img}/s_{occ}) \tag{2-3}$$

$$y_{vote}=y_{img}-y_{occ}(s_{img}/s_{occ}) \tag{2-4}$$

$$s_{vote}=(s_{img}/s_{occ}) \tag{2-5}$$

式（2-2）也可以理解为用学习的特征的空间分布 P_c 给目标中心投票 $p(x|c_i,l)$，$p(c_i\mid f)$ 为特征与中心的匹配度，即为票 $p(x|c_i,l)$ 的权重。这里将与特征 f 匹配成功的各码本入口视作相同，即 $p(c_i\mid f)=1/|c^*|$，c^* 为匹配的入口数。而对每一个码本入口的记录也视作以相同概率出现，即 $p(x|c_i,l)=\dfrac{1}{|Occ[\mathrm{i}]|}$，$|Occ[\mathrm{i}]|$ 为第 i 个入口的记录数量。

投票过程的算法如表 2-3。投票过程结束后，在 (x,y,s) 连续空间内形成了若干散点，见图 2-6(a)。这些点也称为票，票集即投票结果。票集合的分布庞大而复杂，后续需要采用统计方法从票集中搜索到目标的假设坐标 $h=(x_h,y_h,s_h)$。

表 2-3 ISM 投票流程

初始化：票集$V \leftarrow \varnothing$，对测试图提取特征点和特征。M 为与某一特征匹配的码本入口集。

对于所有的特征区域$l_k = (l_x, l_y, l_s)$及其特征值 f_k

$M \leftarrow \varnothing$

  对所有的码本入口c_i

    若$sim(f_k, c_i) \geq \mathrm{t}$

      则$M \leftarrow M \cup (\mathrm{i}, l_x, l_y, l_s)$

  对所有匹配的码本入口c_i^*

    设定匹配权重 $p(c_i^* \mid f_k) \leftarrow \dfrac{1}{|M|}$

  对于所有的匹配$(\mathrm{i}, l_x, l_y, l_s) \in M$

    对于入口 i 的所有记录$occ \in Occ[\mathrm{i}]$

      投票坐标 $x \leftarrow (l_x - \mathrm{occ}_x \dfrac{l_s}{\mathrm{occ}_s}, l_y - \mathrm{occ}_y \dfrac{l_s}{\mathrm{occ}_s}, \dfrac{l_s}{\mathrm{occ}_s})$

      记录投票权重 $p(x|c_i, l) \leftarrow \dfrac{1}{|Occ[\mathrm{i}]|}$

      计算最终的票权重 $w \leftarrow p(x|c_i, l)\, p(c_i \mid f_k)$

      对坐标 x 投出权重为 w 的一票：$V \leftarrow V \cup (x, w, occ, l)$

2.5.2 规模适应假设搜索

第 2.5.1 节陈述了将测试图提取的特征与码本匹配并在投票空间进行投票的算法流程。现在所有票已经统计在投票空间中，要从中找出概率最大值作为目标位置的假设。

概率密度估计法适用于这种已获得采样数据集、求取最大概率点的问题。而概率密度估计方法分参数法和非参数法两种。

参数法是首先假设数据分布服从某个密度函数，然后从采样的数据估计出函数的参数。该方法简单高效，在有一定的先验知识参考时，可以达到比较好的模拟效果。

非参数法不指定概率模型，由相临的若干采样点估计出该邻域中某一点的概率密度。这种方法对复杂和多变的数据集有更强的适应性。对于 ISM 投票的结果，要求取在连续空间的最大值，非参数法更加适用。这里采用均值漂移模型估计法（Mean-Shift Mode Estimation[31]，简称 MSME）。

Fukunaga 等人在一篇研究概率密度梯度函数的估计的文献中，提出了均值漂移的概念，当时的含义指的是偏移的均值向量[31]。此后的很长一段时间内，Mean-Shift 并没有引起人们的注意。1995 年，Yizong Cheng 对基本的均值漂移做了详细的定义，并对其应用领域作了推广[32]。首先，他定义了核函数，使样本点与被偏移点的距离，决定了其相对偏移量对于均值偏移向量的贡献；其次，对于样本点的重要程度，设定了

一个权重系数。Yizong Cheng 还指出均值漂移可以应用的领域，并给出了若干实例。此后，均值漂移在目标跟踪[33]、图像分割[34,35]等领域得到了广泛应用。

MSME 方法将局部最大值沿着邻域内点的方向漂移到概率密度最大值点。该方法认为最大值的邻域内的点服从核概率分布。核概率分布函数 $K(\mathrm{x})$ 有如下性质：

（1） 中心对称

（2） 非负

（3） 以原点为中心

（4） 核空间内积分为 1

搜索过程可以被解释为对目标中心的核概率密度估计，表达形式如式（2-6）。

$$p(o_n,x)=\frac{1}{V_b}\sum_k\sum_j p\left(x_j|f_k,l_k\right)K\left(\frac{x-x_j}{b}\right) \tag{2-6}$$

其中 b 为核带宽，V_b 为体积。实际运用中 b 和 V_b 往往与坐标 (x,y) 和尺寸 s 有关。这里采用高斯函数作为核函数。则式（2-6）形式为：

$$p(o_n,x)=\frac{1}{V_b(x)}\sum_k\sum_j p\left(x_j|f_k,l_k\right)K\left(\frac{x-x_j}{b(x)}\right) \tag{2-7}$$

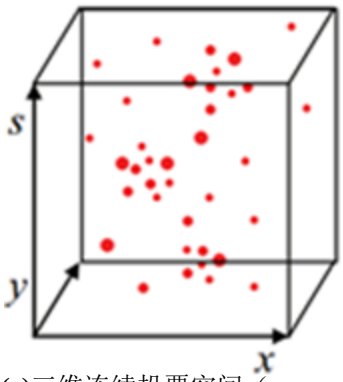
用 MSME 方法搜索最大值之前，先用阈值过滤缩小搜索范围。如图 2-6 所示。

图 2-6 展示了缩小假设搜索范围的步骤。在投票阶段结束后，将产生的票集 $V=\{(x,w,occ,l)_i\}$ 根据三维坐标 x 投射到投票空间 (x,y,s)，如图 2-6(a)，坐标系中的每一个圆点表示一票，圆点越大表示该票的权重 w 越大，即概率越大。投票的参考记录 occ 和当前特征坐标 l 作为参考信息，在后续的目标分割阶段将作为计算目标-背景概率的参数。

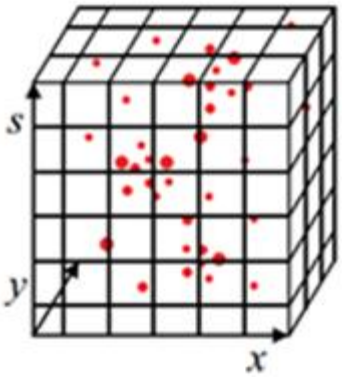
图 2-6(b)将三维投票空间进行立体划分，每一个尺寸为 $(\Delta x,\Delta y,\Delta s)$ 的分割单元 bin 限制了目标坐标 (x,y,s) 的范围。划分之后，每一票都有一个所属的 bin。

图 2-6(c)在各个 bin 的局部空间内搜索局部票权重最大的票。对局部最大值进行阈值过滤，保留的 bin_j 对应的局部最大票 max_j 作为均值漂移的起点。图中的小立方体表示保留的 bin。

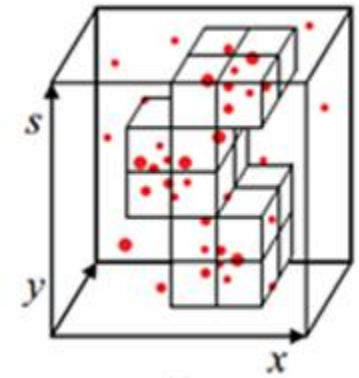
图 2-6(d)以局部最大票 max_j 为中心，以核空间的方式确定假设搜索范围。至此前期处理结束，在核空间内进行均值漂移，求取局部概率最大处坐标。在均值漂移过程中，根据式（2-7）对票值进行核函数加权求和，并作归一化，最终得到假设坐标。



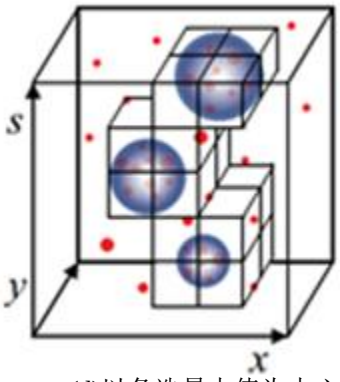
(a)三维连续投票空间（x，y：目标的x、y坐标，s：目标规模）及投票结果。



(b)对投票空间进行划分。



(c)在划分的子空间内找到局部最大值，并保留达到阈值的局部最大值，作为备选最大值。



(d)以备选最大值为中心的核空间，作为MSME搜索的范围。

图 2-6　规模适应最大值搜索的前期处理示意

用 MSME 方法进行规模适应假设搜索的算法流程如表 2-4。

表 2-4　MSME 假设搜索算法

初始化：对投票空间V 进行三维网格划分。X 为局部最大值的集合， $X \leftarrow \varnothing$ 。

for all grid_i **do**
  $x \leftarrow GetLocalMax(\mathrm{grid}_i)$
  if $x \geq \theta$ **then**
    $X \leftarrow X \cup \{\mathrm{x}\}$
  end if
end for
for all $x \in X$ **do**
  repeat
    $score \leftarrow 0, x_{new} \leftarrow (0,0,0), sum \leftarrow 0$

续表 2-4　MSME 假设搜索算法

10	**for all** $vote=(x_k,\mathrm{w}_k,\mathrm{occ}_k,\mathrm{l}_k)$ **do**
11	**if** x_k is inside $K(\mathrm{x})$ **then**
12	$score \leftarrow score + w_k K\left(\frac{x-x_k}{b(x)}\right)$
13	$x_{new} \leftarrow x_{new} + x_k K\left(\frac{x-x_k}{b(x)}\right)$
14	$sum \leftarrow sum + K\left(\frac{x-x_k}{b(x)}\right)$
15	**end if**
16	**end for**
17	$score \leftarrow \frac{1}{V_b(x)} score$, $x \leftarrow \frac{1}{sum} x_{new}$
18	**until** convergence
19	**if** $score \geq \theta_h$ **then**
20	create h for position x
21	**end if**
22	**end for**

对于创建成功的假设 h，对参与均值漂移的票作标记，在目标分割阶段票 (x,w,occ,l) 的 occ 将作为像素分割的参考。

在三维空间进行规模适应假设搜索，其规模适应体现在可依据当前特征的尺寸和码本记录的尺寸，得到当前目标相对于训练目标的相对规模系数。由于训练样本中目标的规模是已知的，故当前目标的规模也可以计算出来。对于不同规模的测试样本，都可以根据 $h=(x_h,y_h,s_h)$ 得到目标中心的二维坐标 (x_h,y_h) 和目标规模 $(s_h w_{sample}, s_h h_{sample})$。

2.6 前景背景分割

上一节介绍了目标中心的搜索过程。在定位目标中心后，关于目标的具体分割信息还没有获得。假设 P 为测试图中的一个像素点，P 属于目标的概率为 $p(P=figure)$，P 属于背景的概率为 $p(P=ground)$。在假设目标中心在 x 处时，P 属于目标的概率为 $p(P=figure\,|\,x)$，P 属于背景的概率为 $p(P=ground\,|\,x)$。

对于 P 是否属于目标，有如下似然度计算式可作为判别标准：

$$L=\frac{p(P=figure\,|\,x)}{p(P=ground\,|\,x)} \tag{2-8}$$

可以分别求取 $p(P=figure|x)$ 与 $p(P=ground|x)$，满足判断式 $L\geq\theta$ 即被认为是目标像素。

第 2.5.1 节着重介绍了测试图中的一个特征对假设的贡献 $p(x|f,l)$，第 2.5.2 节即是对多个特征的投票结果进行概率估计，从而得到目标位置的极大值假设。由贝叶斯公式可得特征对假设的影响为：

$$p(f,l|x)=\frac{p(x|f,l)\,p(f,l)}{p(x)} \tag{2-9}$$

由于图片块包含有特定的分割信息，虽然目前并不知道测试图中特征 (f,l) 的图片块具体的分割信息，可以用 $p(P=figure|x,f,l)$ 暂且表示。设 P 属于目标的概率为：

$$p(P=figure|x)=\sum_{p\in(f,l)}p(P=figure|x,f,l)p(f,l|x) \tag{2-10}$$

在已知图片块的情况下，若 P 为图片块中的目标像素，则 $p(P=figure|x,f,l)$ 为 1，否则为 0。此时 $p(P=figure|x,f,l)$ 显然未知，可以参考相似特征的图片块的分割信息获得。

由于特征 (f,l) 与码本匹配，故该特征对应的图片块内目标与背景的分割信息可以参考码本入口中的记录 occ。从假设搜索的过程来看，漂移得到假设 $h=(x_h,y_h,s_h)$ 的核空间中的票点是 h 的成因，则在 h 创建成功后，可以由这些票点得到 h 成立条件下的像素信息。用于计算似然度范围的票所在的核空间见图 2-6(d)示意。

将 $p(f,l|x)$ 分解为与码本匹配的形式，将式（2-2）代入式（2-9）得式（2-11）。

$$p(f,l|x)=\frac{\sum_i p(x|c_i,l)\,p(c_i|f)p(f,l)}{p(x)} \tag{2-11}$$

将式（2-11）代入式（2-10）得到：

$$\begin{aligned}p(P=figure|x)&=\sum_{p\in(f,l)}p(P=figure|x,f,l)\left(\frac{\sum_i p(x|c_i,l)\,p(c_i|f)p(f,l)}{p(x)}\right)\\&=\sum_{p\in(f,l)}\sum_i p(P=figure|x,c_i,l)\frac{p(x|c_i,l)\,p(c_i|f)p(f,l)}{p(x)}\end{aligned} \tag{2-12}$$

同理，P 为背景点的概率为：

$$\begin{aligned}p(P=ground|x)&=\sum_{p\in(f,l)}\sum_i p(P=ground|x,c_i,l)\frac{p(x|c_i,l)\,p(c_i|f)p(f,l)}{p(x)}\\&=\sum_{p\in(f,l)}\sum_i(1-p(P=figure|x,c_i,l))\frac{p(x|c_i,l)\,p(c_i|f)p(f,l)}{p(x)}\end{aligned} \tag{2-13}$$

式（2-12）和式（2-13）中，$p(x|c_i,l)\,p(c_i\mid f)$为统计投票时的票权重；$p(f,l)$为指示变量，表示特征(f,l)的存在性；$p(x)$为常量。因此可以将$p(f,l)$和$p(x)$提取公因式，如此计算似然度（2-8）的公式简化为：

$$L=\frac{\sum_{p\in(\mathrm{f,l})}\sum_{i}p(P=figure\mid x,c_i,l)p(x|c_i,l)\,p(c_i\mid f)}{\sum_{p\in(\mathrm{f,l})}\sum_{i}(1-p(P=figure\mid x,c_i,l))p(x|c_i,l)\,p(c_i\mid f)} \tag{2-14}$$

即计算某一像素点P的目标似然度时，对于所有包含该像素的特征图片块，在其投票阶段所有匹配的码本记录中，若P在对应位置属于目标，则为P的目标概率投票，若P在对应位置属于背景，则为P的背景概率投票，票权重即为目标中心投票权重。由于本过程是由中心位置假设，回溯到中心投票阶段，再参考投票记录中的前景背景分割情况，逐一对像素进行概率投票，是一个自顶向下的过程。该过程与 2.5.1 节及 2.5.2 节的由单个特征匹配码本、再由众多记录对中心位置投票的自底向上的过程刚好相反，因此也称为自顶向下分割。其算法流程如表 2-5。

表 2-5　自顶向下分割算法

初始化：假设h及其支持票集V_h（参与 MSME 过程的票集）。img_{mask}为方阵，1 表示像素点属于目标，0 表示像素点属于背景。

1　**for all** $v=(x,w,occ,l)\in V_h$ **do**
2　　img_{mask}为码本记录occ对应的图片块的目标背景分割 mask
3　　sz为l表示的图片块边长
4　　$Rescale(img_{mask},sz)$
5　　$x_0\leftarrow(l_x-\frac{1}{2}sz),y_0\leftarrow(l_y-\frac{1}{2}sz)$
6　　**for all** $x_p,\mathrm{y}_p\in[0,sz-1]$
7　　$P\leftarrow(x_p+x_0,y_p+y_0)$
8　　$p(P=\mathrm{figure})+=w*img_{mask}(x_p,y_p)$
9　　$p(P=ground)+=w*(1-img_{mask}(x_p,y_p))$
10　**end for**
11　**end for**

3 基于隐形状模型的手掌检测器的实现

本章以手掌作为检测目标，对基本隐形状模型的实验过程进行了详细的说明。首先，结合目标特点，设计了样本子类，并制作了训练样本集和测试样本集，并介绍了算法的评价方法。其次，结合尺度不变性的检测需求和颜色稳定的目标特点，设计了DoG 作为特征点提取算子，统一尺度灰度特征作为特征描述子。最后，结合各阶段算法的输入输出，对主要的数据结构进行了说明。

3.1 样本介绍

3.1.1 训练样本

1）训练集 1

本文为了实现多姿势手掌的识别，需要训练多种姿势的码本，因此选取了五个固定场景，并将手掌姿势分为闭合偏左、闭合偏右、闭合朝上、张开偏左、张开偏右、张开朝上六种。训练样本图片大小均为 720*576 像素的，手掌宽度范围为 100~130,高度范围为 180~250。训练码本时，对每种姿势的样本单独训练出一个码本。考虑到左右手的对称性，对左手和右手的训练和检测过程一致，本文所有的样本都只采集了右手。场景及姿势的划分见图 3-1。

(a)闭合偏左-场景1　(b)闭合偏右-场景2　(c)闭合朝上-场景3

(d)张开偏左-场景4　(e)张开偏右-场景5　(f)张开朝上-场景1

图 3-1　多姿势多场景样本示例图

另外，为了保证训练特征池中的特征都是手部特征，需要制作标定手部区域的掩

模（mask）。除了手动标注外，常见的标识运动目标的方法有高斯背景模型[36]等。由于手掌区域颜色较均匀，属于肤色范围，也可以采用肤色模型来标识，本文采用的就是这种办法。对原始样本图进行肤色提取后，往往会有部分背景和人脸部分也划分到肤色区域，需要人工过滤掉手部以外的部分。如图 3-2。对肤色模型的介绍详见 4.1.1 节。

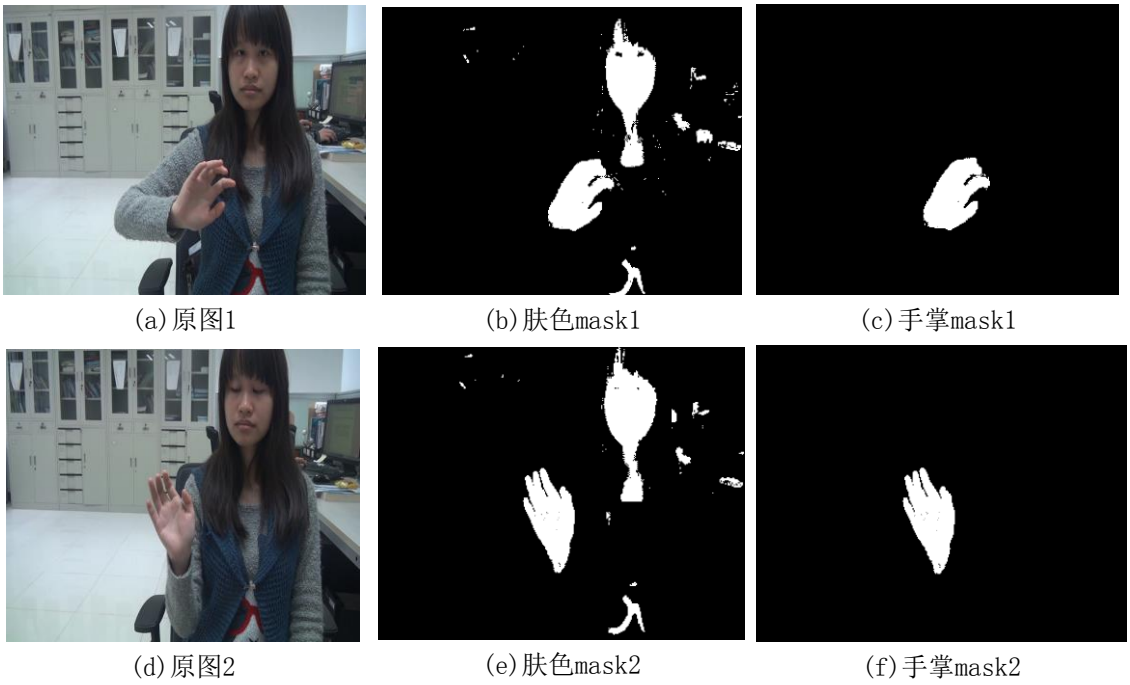


图 3-2　手掌 mask 制作流程

对每种姿势的每种场景取 10 张图，即每种姿势有 50 张训练样本及对应的 mask。6 种姿势总计 300 张样本原图及 300 张 mask。

2）训练集 2

训练集 2 为复杂背景中的目标。背景为来自于室内应用场景随机背景库。制作流程是首先将手掌从简单背景中分离，再放置到随机背景中。

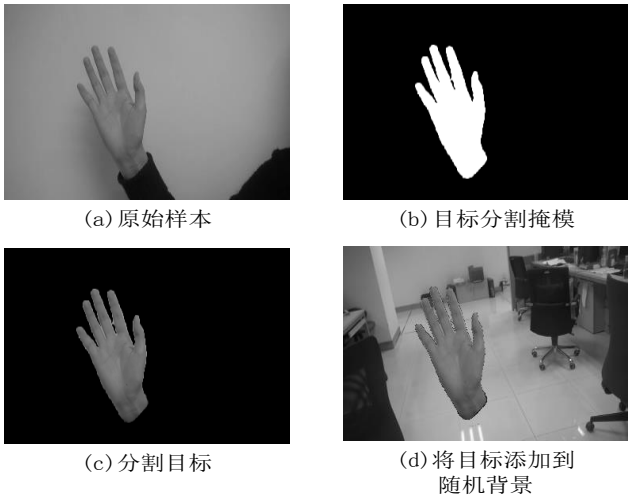


图 3-3　训练集 2 制作流程

本文采用背景建模法分离前景目标。即采用与肤色差别较大的背景，再由 YUV 空间的肤色模型即可判断像素是否属于肤色区域。YUV 肤色模型的详细介绍见 4.1.1 节。图 3-3 展示了训练集 2 的制作流程。首先将图像由 RGB 空间转换到 YUV 空间，再将像素值（Y，U，V）满足$U \in 130-170$，$V \in 99-130$的像素点作为目标掩模中的白色点，其余点作为掩模中的黑色点，最后将分割出的目标加入到随机背景中。

训练集 2 中手掌宽度范围为[80,140]，高度范围为[130,240]。分为偏左、偏右、朝上三个子集，其中偏左为 480，偏右 293，朝上 680 张。

3.1.2 测试样本

1）测试集 1

测试集 1 与训练集 1 取自相同的场景，同样来自 5 个场景，分为 6 种姿势。每种姿势的测试样本为 30 张，共计 180 张。为了评价检测结果，需要人工标注手掌的位置及尺寸，如图 3-4 所示。用$RGB=(255,0,0)$的红色矩形框标记手掌位置后，用程序确定矩形框的起止点坐标(x_1,y_1,x_2,y_2)。

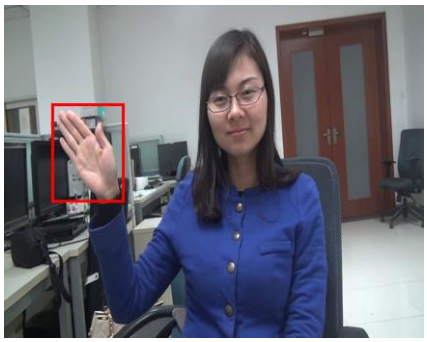


图 3-4　人工标注测试样本目标位置

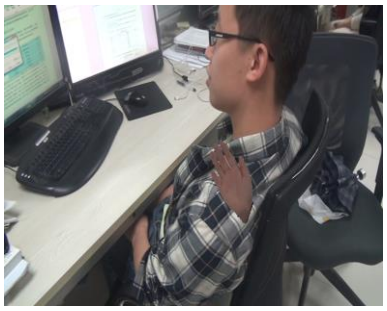
测试缩放性能时，按照缩放系数$scale \in [0.5,3.0]$对原测试图进行缩放，并对目标坐标(x_1,y_1,x_2,y_2)进行等比例缩放。这样得到缩放后的测试图和对应的目标坐标，可用于测试该缩放系数下的检测性能。

对于训练集 1 而言，测试集 1 用于显示算法的基本性能。该测试集比较简单，后文测试主要采用测试集 2。

2）测试集 2

测试集 2 中手掌的原始尺寸与训练集 2 中的尺寸相当，背景为随机室内场景，图片大小为 720*576 像素。测试集 2 分为偏左、偏右、朝上三个子集，其中偏左 124 张，偏右 128 张，朝上 160 张。每张图片中包含一个目标手掌。测试集 2 的各子集示例如图 3-5 所示。

测试集 2 用于测试 ISM 算法的一般性能。后文中如无特别注明，均为采用训练集 2 训练、测试集 2 测试的结果。


(a) 偏左测试样本


(b) 偏右测试样本


(c) 朝上测试样本

图 3-5　测试集 2

3.2 评价指标

3.2.1 评价单个检测结果

手动标注的测试样本中的目标区域为 (x,y,w,h)。其中 (x,y) 为矩形区域的中心坐标，(w,h) 为矩形的宽与高。检测结果同样以矩形区域 (x',y',w',h') 表示。如图 3-6 所示，Δx 和 Δy 为原始与假设的矩形中心在 x 和 y 方向上的距离。

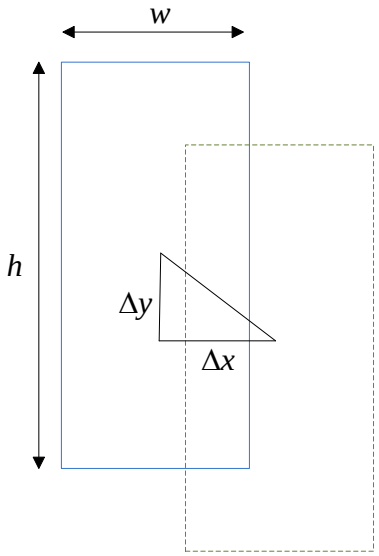


图 3-6　相关距离计算参数

用相关距离表示检测位置与原始位置的距离。见式（3-1）。

$$d_r = \sqrt{(\frac{2\cdot\Delta x}{w})^2 + (\frac{2\cdot\Delta y}{h})^2} \qquad (3\text{-}1)$$

本文对相关距离阈值取 0.5，即 $d_r \le 0.5$ 时，认为成功找到目标，否则认为检测失败。

3.2.2 评价算法性能

根据测试样本的正负性和预测结果的正负性，检测结果有四种类型，如表 3-1。

表 3-1　四种检测结果

样本类型	检测结果	
	正假设	负假设
正样本（Pos）	True Positive(TP)	False Negative(FN)
负样本（Neg）	False Positive(FP)	True Positive(TN)

对测试样本集的检测结果进行统计，得到 TP、FP、FN、TN 的频数。

检测率（recall）和准确率（precision）为两个常用的评价指标，计算公式为

$$recall = \frac{TP}{Pos} = \frac{TP}{TP + FN} \tag{3-2}$$

$$precision = \frac{TP}{TP + FP} \tag{3-3}$$

相应地，错误检测结果在所有结果中所占比例称为误检率（false alarm rate）。

$$fa = 1 - precision = \frac{FP}{TP + FP} \tag{3-4}$$

本文将 *recall* 和 fa 的组合转化为 ROC 曲线，以展示检测器性能。后文中也用检测率和误检率表示 *recall* 和 fa 。

3.3 特征介绍

3.3.1 DoG 特征点

DoG(Difference-of-Gaussian)特征点检测算子[30]是 Lowe 于 2004 年提出的，它可以检测不受图片规模影响的稳定的特征点。该特征点的检测方法是，对一个规模化的拉氏函数，寻找它的规模空间极值。Lowe 指出该拉氏函数可以用高斯差函数来精确估计。设 G 是方差为 σ^2 的高斯函数，$f(x,y)$ 为图像灰度值，则对原图进行参数为 σ 的高斯滤波，表示如下：

$$f(x,y) = G_\sigma(x,y) * f(x,y) \tag{3-5}$$

那么将图像进行不同参数的滤波再相减，表示如下：

$$\begin{aligned} & g_1(x,y) - g_2(x,y) = G_{\sigma 1}(x,y) * f(x,y) - G_{\sigma 2}(x,y) * f(x,y) \\ & = (G_{\sigma 1}(x,y) - G_{\sigma 2}(x,y)) * f(x,y) \\ & = DoG * f(x,y) \end{aligned} \tag{3-6}$$

若设置规模距离因子为 k ，则两个邻近的规模的高斯差函数公式为

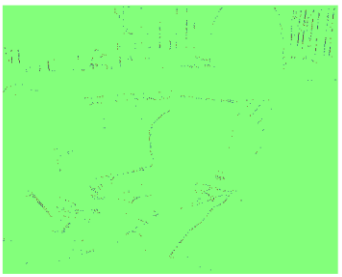
$$D(x,\sigma) = (G(x,k\sigma) - G(x,\sigma)) * f(x,y) \tag{3-7}$$

DoG 特征点为 $D(x,\sigma)$ 在图像坐标和规模坐标构成的三维空间中，取得局部最大值的点。具体到图像处理来讲，就是将一幅图像在不同参数下的高斯滤波结果相减，得

到 DoG 图，再在规模σ所在的维度上，计算多规模的 DoG 图的极值点。如图 3-7 所示，(a)、(b)、(c)为不同规模的双高斯差结果，(d)为对(a)、(b)、(c)取极值的点，(e)将取得的特征点在原图像中展示。



(a) DoG图像1 ($\sigma1=0.3,\sigma2=0.4$)　(b) DoG图像2 ($\sigma1=0.6,\sigma2=0.7$)　(c) DoG图像3 ($\sigma1=0.7,\sigma2=0.8$)




(d) 极值点　(e) 原图中的DoG角点

图 3-7　DoG 角点提取示意

当计算特征点所在图片块的特征值时，为了获取更多的局部形状信息，一般在半径$r=3\sigma$范围内进行特征值计算。

3.3.2 灰度特征

灰度特征即以特征区域的灰度值作为特征向量值。如特征点提取半径为 12，则在特征点为中心的 25*25 图片块内依次取灰度值，得到 625 维的特征值向量。灰度值描述子的特征计算方法是，计算向量元素的均值，用各元素减去均值再归一化，最终结果表达的是图片块内灰度值相对于灰度均值的分布情况。

为了计算不同规模的特征区域的相似度，将特征区域缩放到统一尺寸，实验中采用双线性插值法缩放成 25*25 的方阵。特征提取的原始范围σ则记录在特征点相关信息中。

对于灰度特征，采用相关系数作为相似度。计算当前 FeatureVector 类（见表 3-3）对象计算与另一 FeatureVector 类对象的相关系数相似度的代码如下：

表 3-2　相关系数计算代码

```
1        float corr = 0.0;
2        for( int i=0; i<nTotalBins; i++ )
3            corr += m_vBins[i] * other.m_vBins[i];
```

对向量元素乘加即可得到相关系数。相关系数越大，表示两个图片块越相似。相关系数越小，甚至小于零，表示两个图片块有较大差异。

用 $correlation(x, y)$ 表示灰度特征 x 和 y 的相似度。在聚类阶段，需要计算类间相似度，以确定是否合并类。设类 $C1$ 元素个数为 $n1$，类 $C2$ 元素个数为 $n2$，则类 $C1$ 和类 $C2$ 的相似度为

$$sim(C1, C2) = \frac{\sum_{p=1}^{n1}\sum_{q=1}^{n2} correlation(C1_p, C2_q)}{n1 * n2} \tag{3-8}$$

在空间分布学习和检测时特征匹配码本阶段，设码本入口为 Ci，当前待匹配的特征为 f，则匹配相似度为

$$sim(Ci, f) = correlation(Ci, f) \tag{3-9}$$

3.4 训练过程分析

3.4.1 训练流程

在样本集上训练的流程分为两步：

（1）对样本集进行特征提取，得到特征池。对特征池进行聚类，将聚类中心作为码本的入口；

（2）将特征池中的特征与码本入口逐一匹配，将匹配的特征的相关信息记录在入口后方。

这两个阶段都是在特征值所在的空间内操作的。假设特征池中的每个特征都是 nDims 维，则聚类和匹配都是在 $\mathbb{R}$ 上的 nDims 维向量空间 $\mathbb{R}^{nDims}$ 上进行。特征值的相似度与特征点在 $\mathbb{R}^{nDims}$ 上的距离是一致的。

图 3-8 展示了从样本提取特征并聚类的过程。为了直观，将特征用对应的图片块来展示。如图 3-8(a)所示，对样本集进行特征点提取后，用 mask 过滤，保留 mask 区域的特征点。再对保留的特征点进行图片块和特征值计算，将特征值加入到特征池（图 3-8(b)）。最后对特征池进行聚类，聚类中心作为码本的入口（图 3-8(c)）。至此，码本的基本框架就产生了。

从图 3-8(c)的类成员及类中心可见，形状相似的部位往往聚为一类。类中心作为该类特征的平均特征值，可以代表该类特征。本文认为小规模的类（成员数≤1）是偶发型噪点，非典型目标特征，可以直接删除，以减小码本容量。

码本的入口就绪以后，进入训练流程的第 2 步。将特征池中的特征与码本入口逐个进行匹配，若匹配成功，则插入以该码本入口为起始的链表。图 3-9 展示了码本记录的产生过程，其中步骤(a)和(b)与训练产生码本入口的图 3-8 的(a)和(b)阶段过程类似。

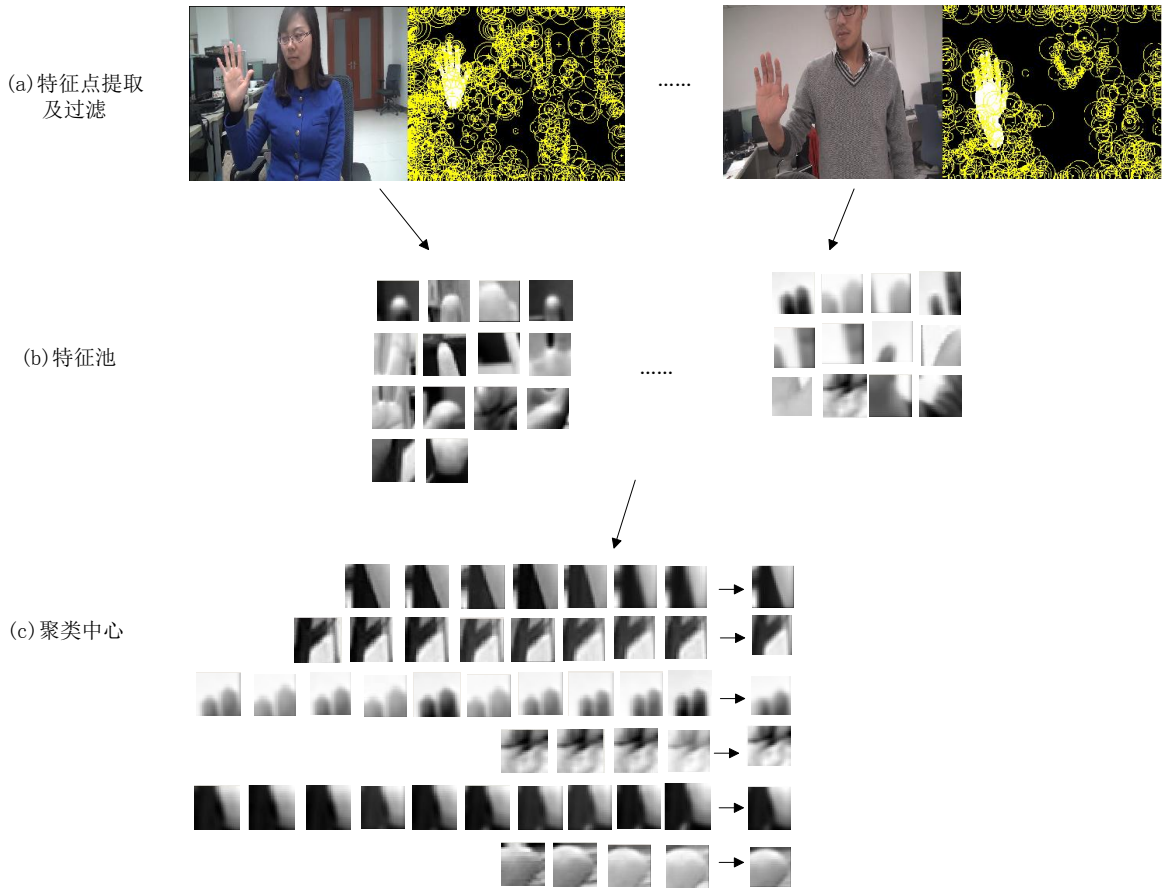


图 3-8　提取特征并聚类

对于较大的样本集，在聚类阶段可以选择少量样本进行聚类，在匹配阶段用大量的样本去匹配。只要保证码本的入口包括了目标的所有局部特征，则少量样本聚类即足够。匹配阶段应尽量选取使局部特征的分布更全面的样本集，以确保所有特征与目标中心可能的位置关系都被记录在码本中。采用少量聚类、大量匹配的方法可以压缩训练时间。图 3-8 的特征池和图 3-9 的特征池可以不一致。

图 3-10 展示了多个训练样本提取特征后，匹配成功的码本入口的相应图片块在原特征坐标处显示的结果。可以直观地看出，匹配的图片块基本上覆盖了目标区域。即对训练样本而言，码本中已包含足够还原目标的特征集。

特征池重新匹配类中心，使得一个特征值可与多个中心匹配，也有特征值与各中心均距离较远而不被记录。该步骤相比于直接将聚类结果存入码本，将类成员作为码本入口，更能表达数据的分布规律，使较强的特征可以多次被记录，且进一步剔除了噪点。

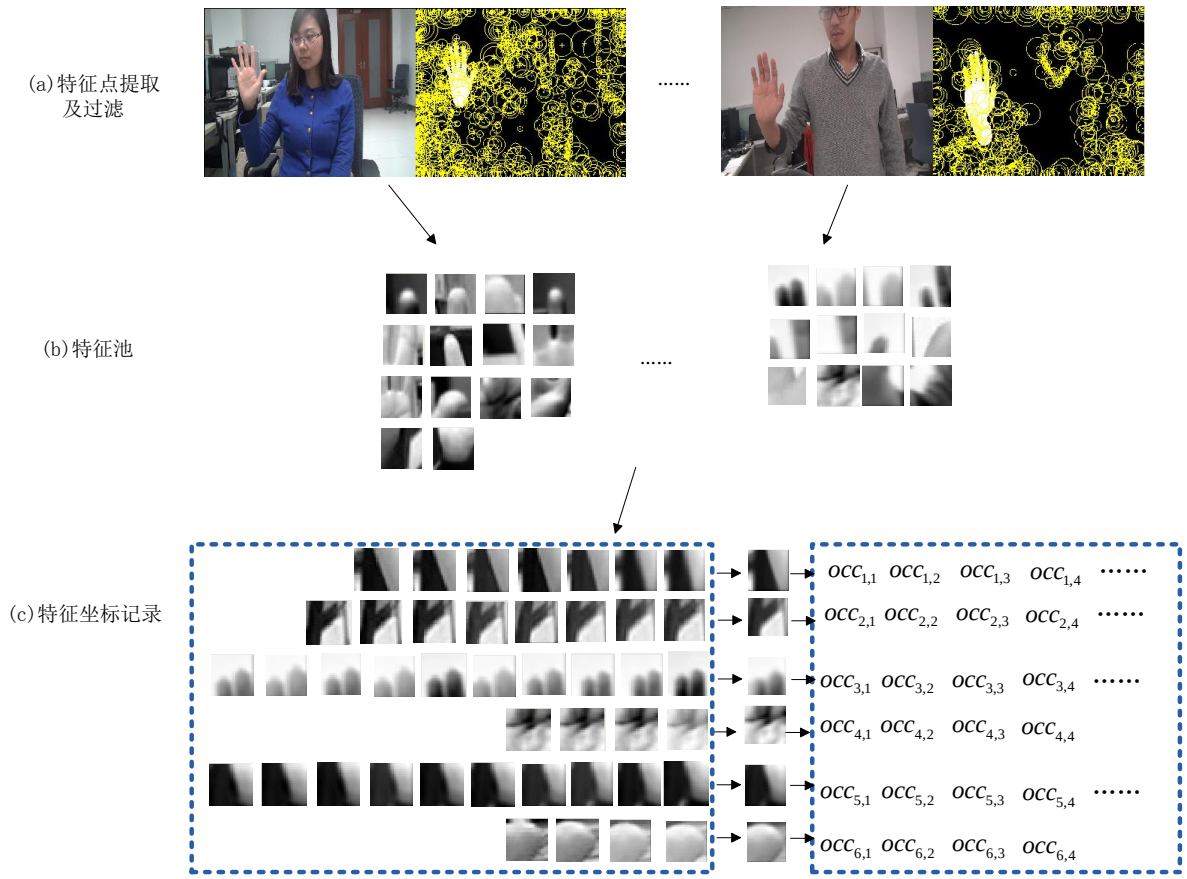


图 3-9　码本记录的产生过程

图 3-10　匹配成功的特征图片块

3.4.2 主要数据结构

特征值向量是数据集的元素，在训练阶段和检测阶段的特征表示方式应该完全一致。表 3-3 为特征向量值的通用数据结构。

表 3-3　FeatureVector 类的重要字段

类型	字段	备注
int	m_nDims	特征值维数
vector<float>	m_vBins	特征值向量

训练样本和测试样本在加载后，首先都需要提取特征。表 3-4 为样本信息类 FeatureCue，它保存了一幅样本的相关信息。其中 OpGrayImage 为自定义图像类型，PointVector 为坐标点集。由于提取特征的规模非统一，Point 类中还包含有特征描述子

计算规模 s 字段。另外，由 m_imgSrcMap 和 m_vPtIdzs 可以得知保留的图片块对应的 mask 图片块。m_vPointsInside 和 m_vPatches 存在对应关系。而保留的特征的特征值可以通过读取 m_vPtIdz 作为序号，从 m_vFeatures 读取相应元素获得。 此外，用一个 vector<OpGrayImage>类型数据记录所有样本的保留的 mask 图片块，在进行空间分布学习阶段，该局部特征 mask 与码本记录同步保存。

表 3-4　FeatureCue 类的重要字段

类型	字段	备注
int	m_nFeatureType	特征类型
QImage	m_qimgSrc	原始 RGB 样本
OpGrayImage	m_imgSrc	样本灰度图
OpGrayImage	m_imgSrcMap	样本的二进制 mask 图
PointVector	m_vPoints	样本中的特征点集
PointVector	m_vPointsInside	mask 过滤后的特征点集
vector<OpGrayImage>	m_vPatches	过滤后的特征点集对应的图片块集
vector<FeatureVector>	m_vFeatures	图片的所有特征
vector<int>	m_vPtIdzs	保留的特征点在 m_vPoints 里的序号

由于码本记录中需要特征与目标中心的相对位移信息，见表 3-7。每个特征除了需要一个 FeatureVector 记录其特征值外，还需要一个 FeatureVector 记录其相对坐标 (x, y, s)。在 FeatureCue 中的 m_imgSrcMap 加载成功后，计算 mask 区域的中心作为目标中心 $(\boldsymbol{C}_x, \boldsymbol{C}_y)$。计算 m_vPatches 的每个元素的特征值的同时，计算其相对中心的位移及特征尺寸 $(\boldsymbol{C}_x - l_x, \boldsymbol{C}_y - l_y, l_s)$，并以 FeatureVector 记录。由一个 vector< vector<FeatureVector> >类型二维向量记录了特征池的所有元素的相对位移。

ClusterParams 类包含了训练的参数。合成聚类阶段要对聚类的层次进行控制，m_nMaxNodeSize 控制了最大类的规模。m_dSimilarity 字段既是类合并阈值，也是特征匹配码本入口的阈值。

表 3-5　ClusterParams 类的重要字段

类型	字段	备注
double	m_dSimilarity	类合并的相似度阈值
int	m_nMaxNodeSize	最大的类规模

码本入口的相关信息由 Codebook 类记录，包括类中心的特征值 m_vClusters 及类成员图片块的均值图片块 m_vClusterPatches。m_vClusterPatches 可以直观地展示聚类得到的目标特征。

表 3-6　Codebook 类的重要字段

类型	字段	备注
vector<FeatureVector>	m_vClusters	聚类中心
vector<OpGrayImage>	m_vClusterPatches	聚类中心对应图片块

与码本入口匹配的特征作为一条 Occ（Occurrence，码本记录的简写）记录在其后方，每条 Occ 的记录由一个_ClusterCooccurrence 类的对象表示。具体参考表 3-7。（dPosX，dPosY，dScale）来自于前面提到的记录特征相对中心位移的 vector< vector<FeatureVector> >二维向量。

表 3-7　_ClusterCooccurrence 类的重要字段

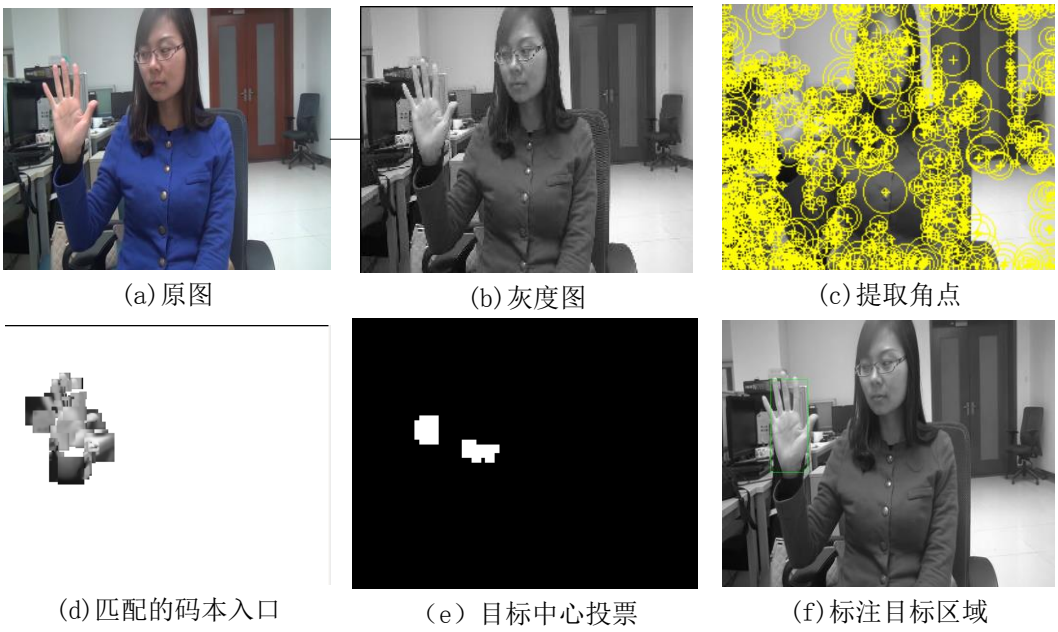
类型	字段	备注
int	nPose	Occ 所属目标的子类编号
float	dScale	Occ 的规模系数
float	dPosX	Occ 与目标中心的相对坐标 x
float	dPosY	Occ 与目标中心的相对坐标 y
float	dDistance	Occ 与码本入口的相似距离

码本中的记录不在 Codebook 类对象中，单独用 vector< vector<_ClusterOccurrence> >表示。相应地，vector< vector<int> >型数据记录了码本记录对应的 mask 在保留 mask 图片块集 vector<OpGrayImage>中的序号，该记录在分割目标阶段可用作像素前背景概率投票的参考数据。

3.5 检测流程分析

3.5.1 检测流程

检测流程可参考图 3-11。对一幅 RGB 测试图（图 3-11(a)），首先将其转化为灰度图（图 3-11(b)），再提取与训练阶段类型一致的特征（图 3-11(c)）。将每个特征与码本入口进行相似度匹配，图 3-11(d)为匹配的码本入口的图片块展示。随后由投票和搜索算法（表 2-3 和表 2-4）得到目标中心位置的假设。图 3-11(e)显示了所有被投票的中心位置，图 3-11(f)中绿色方框的中心为对图 3-11(e)中的票进行均值漂移搜索，得到的假设平面位置(x, y)，绿色方框的尺寸为假设尺寸与训练样本的尺寸规模比s计算而得。

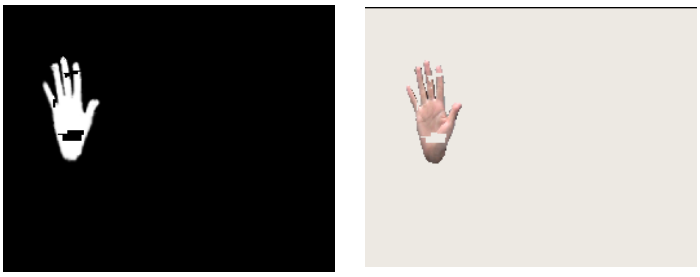


(a) 原图　(b) 灰度图　(c) 提取角点

(d) 匹配的码本入口　(e) 目标中心投票　(f) 标注目标区域

图 3-11　检测流程（顺序：从(a)到(f)）

若要分割目标，根据图 3-11(e)中参与形成假设的核空间中票的记录 Occ，参考分割掩模，对每个像素计算前景似然度。计算前景似然度时，用一张前景概率图片记录前景概率，另一张背景概率图片记录背景概率，根据分割算法（表 2-5）逐像素作前景和背景概率累加，最后将用前景概率图除以背景概率图即可得到似然度图。图 3-12(a)为分割的前景像素标定，图 3-12(b)为从原图像中分割出的目标。



（a）自顶向下分割　(b) 在原图的分割效果

图 3-12　目标分割

3.5.2 主要数据结构

检测阶段由于要将特征值与码本匹配，并根据特征的坐标推断目标中心的坐标，故与训练样本相似（3.4.2 节），特征需要两个 FeatureVector 对象（表 3-3）记录特征值和推测的中心位置信息。

依据 ISM 投票算法（表 2-3）投票时，需要有一个数据结构表示单个票信息 (x, w, occ, l) 。HoughVote 类记录了投票空间中的一个投票点的相关信息，具体参考表 3-8。m_nOccMapId 为 3.4.2 节末尾记录的对应的 mask 图片块的序号。

表 3-8　HoughVote 类的重要字段

类型	字段	备注
FeatureVector	m_fvCoords	目标中心的坐标和规模
float	m_dValue	投票权重
int	m_nImgPointId	特征在测试图特征集中的序号
int	m_nClusterId	特征匹配的码本入口序号
Int	m_nOccNumber	投票的Occ在该码本入口的记录中的序号
int	m_nOccMapId	投票的 Occ 在所有 mask 图片块中的序号
int	m_nCueId	目标子类号

将每个特征与码本匹配后，得到了大量由 HoughVote 类对象构成的票集，需要在投票空间中进行假设搜索，如使用均值漂移模型估计（表 2-4）。VotingSpace 类包含了投票空间的属性及数据。VotingSpace 类的 m_vlVotes 字段用于储存所有的 HoughVote 类型的票，m_vBinMeans 储存每个 bin 内的搜索结果，即是假设的目标位置，m_vBinScores 为票权重$\{w_k\}$在核空间$K(x)$上的加权概率和 *score*。m_fvWindowSize 为投票空间上划分 bin 的参数，用以控制寻找局部最大值的搜索范围。m_vlVotes、m_vBinScores 和 m_vBinMeans 维数相同，均为非空 bin 的个数。m_vlVotes 的每个元素记录了一个 bin 中的所有票。

表 3-9　VotingSpace 类的重要字段

类型	字段	备注
vector< list<HoughVote> >	m_vlVotes	投票空间的票集
vector<float>	m_vBinScores	所有 bin 中的局部最大 score
vector<FeatureVector>	m_vBinMeans	bin 中的票的漂移结果
FeatureVector	m_fvWindowSize	一个 bin 的 size 参数

假设搜索结束后，对于 VotingSpace 中的 m_vBinScores 中大于阈值θ的元素，产生目标中心的假设（参考算法表 2-4）。假设由_Hypothesis 类对象记录，具体字段见表 3-10。

表 3-10　_Hypothesis 类的重要字段

类型	字段	备注
int	x	目标中心坐标 x
int	y	目标中心坐标 y
float	dScale	目标规模 s
int	nBoxX1	目标区域起始坐标 x
int	nBoxY1	目标区域起始坐标 y
int	nBoxWidth	目标区域宽度

续表 3-10　_Hypothesis 类的重要字段

类型	字段	备注
int	nBoxHeight	目标区域高度
float	dScore	假设搜索结果分数
float	dScoreMDL	MDL 检验分数
int	nPose	目标所属的子类

若要求实现目标分割，在使用均值漂移法构造假设_Hypothesis 类对象时，对位于局部最大值核空间内的票作记录，即每个_Hypothesis 类对象都有一个核空间票子集记录。分割阶段，根据该记录参考 VotingSpace 类 m_vlVotes 的对应元素，由每一个 HoughVote 中的 m_nOccMapId，索引到训练样本中的图片块分割掩模，再累加 m_dValue 到前景和背景概率，即可计算每个像素的前景似然度。算法详见表 2-5。

4 基于隐形状模型的手掌检测器的优化与扩展

在完成基本 ISM 手掌检测器的实现之后，本章根据手掌的特点对 ISM 模型的应用作了两点扩展，并对扩展后的检测流程和相关参数进行了详细的阐述。

4.1 肤色模型与膨胀 mask

4.1.1 肤色模型介绍

肤色区域有稳定、均匀的特点，在颜色空间上往往表现出聚类特性。根据这种聚类特性，对肤色在颜色空间上的分布建立模型，即为肤色模型，也可以称之为肤色分类器[37]。理论已经证明，每一种颜色空间都存在相应的肤色分类器，能够得到基本一致的检测结果。

1）YUV 空间上的肤色模型

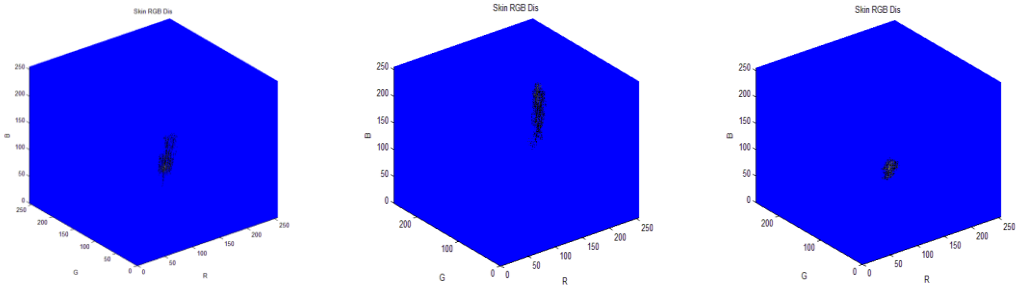
YUV 色彩空间可由 RGB 空间根据公式转换而来。它的特点是将亮度分离到 Y 通道，而 U 和 V 通道分别为色调和饱和度，UV 通道共同描述了像素的色度。由于手掌检测的应用场景的光照条件不稳定，肤色分类器需要有光照鲁棒性。由于 UV 二维不受亮度的影响，YUV 空间上的肤色点在 UV 平面上就表现出很好的聚类特性。可以根据肤色在 UV 平面上的聚类范围，来判断一个像素点是否属于肤色区域。前人已经通过不断地调试与验证，总结出 YUV 空间上的肤色范围为$133 \le U \le 173, 77 \le V \le 127$。

2）RGB 空间上的肤色模型

前面介绍的 YUV 空间上的肤色模型是一种通用的模型，其检测范围往往较大。对一个给定的肤色区域，会发现其在 YUV 空间上只分布在肤色范围的一个子区间内。为了得到更精确的肤色模型，在用 YUV 模型初步确定肤色区域后，可根据手掌的分布规律取手心位置的肤色块作为肤色样本，再对样本建模。图 4-1 为三组不同光照条件下的肤色取样示意图，黑色矩形框标注了取样范围。图 4-2 为对应的样本在 RGB 空间上的频数分布，颜色越深表示频数越低，越浅表示频数越高。可以看出，偏白色区域非常小，着色区域亦比较紧凑，表明大量的肤色点落入很小的空间内，肤色在 RGB 空间上有很好的聚类特性。

(a) 肤色样本1　(b) 肤色样本2　(c) 肤色样本3

图 4-1　不同光照条件下的肤色样本



(a)肤色样本1的分布　　(b)肤色样本2的分布　　(c)肤色样本3的分布

图 4-2　肤色样本在 RGB 空间上的分布

由于肤色的均匀性质，肤色样本可以代表该个体的肤色特性。对肤色样本建立一个概率模型，则可以根据像素点的肤色似然概率值确定其是否属于肤色。如果采用高斯模型模拟肤色分布，则 $N_{skin}(\vec{\mu}_{skin}, \vec{\Sigma}_{skin})$ 的参数计算方法如下：

设 R、G、B 分别为 row*col 的样本像素在 RGB 颜色空间三个维度上的分量矩阵。则均值

$$\begin{aligned}\vec{\mu}_{skin} &= [\bar{R}\ \bar{G}\ \bar{B}] \\ &= \frac{1}{row * col}[\sum\sum R(\mathrm{row},\mathrm{col}), \sum\sum \mathrm{G}(\mathrm{row},\mathrm{col}), \sum\sum \mathrm{B}(\mathrm{row},\mathrm{col})]\end{aligned} \tag{4-1}$$

设方差矩阵为

$$\vec{\Sigma}_{skin} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \tag{4-2}$$

则 $\vec{\Sigma}_{skin}$ 各元素计算公式如式（4-3）至式（4-5）。

$$\begin{aligned} a_{11} &= \frac{1}{row * col}\sum_{row}\sum_{col}(\mathrm{R}(\mathrm{row},\mathrm{col}) - \bar{R})^2 \\ a_{12} &= \frac{1}{row * col}\sum_{row}\sum_{col}(\mathrm{R}(\mathrm{row},\mathrm{col}) - \bar{R}) * (\mathrm{G}(\mathrm{row},\mathrm{col}) - \bar{G}) \\ a_{13} &= \frac{1}{row * col}\sum_{row}\sum_{col}(\mathrm{R}(\mathrm{row},\mathrm{col}) - \bar{R}) * (\mathrm{B}(\mathrm{row},\mathrm{col}) - \bar{B}) \end{aligned} \tag{4-3}$$

$$\begin{aligned} a_{21} &= \frac{1}{row * col}\sum_{row}\sum_{col}(\mathrm{G}(\mathrm{row},\mathrm{col}) - \bar{G}) * (\mathrm{R}(\mathrm{row},\mathrm{col}) - \bar{R}) \\ a_{22} &= \frac{1}{row * col}\sum_{row}\sum_{col}(\mathrm{G}(\mathrm{row},\mathrm{col}) - \bar{G})^2 \\ a_{23} &= \frac{1}{row * col}\sum_{row}\sum_{col}(\mathrm{G}(\mathrm{row},\mathrm{col}) - \bar{G}) * (\mathrm{B}(\mathrm{row},\mathrm{col}) - \bar{B}) \end{aligned} \tag{4-4}$$

$$
\begin{aligned}
a_{31} &= \frac{1}{row * col}\sum_{row}\sum_{col}(\mathrm{B}(\mathrm{row},\mathrm{col}) - \bar{B}) * (\mathrm{R}(\mathrm{row},\mathrm{col}) - \bar{R}) \\
a_{32} &= \frac{1}{row * col}\sum_{row}\sum_{col}(\mathrm{B}(\mathrm{row},\mathrm{col}) - \bar{B}) * (\mathrm{G}(\mathrm{row},\mathrm{col}) - \bar{G}) \\
a_{33} &= \frac{1}{row * col}\sum_{row}\sum_{col}(\mathrm{B}(\mathrm{row},\mathrm{col}) - \bar{B}))^2
\end{aligned}
\tag{4-5}
$$

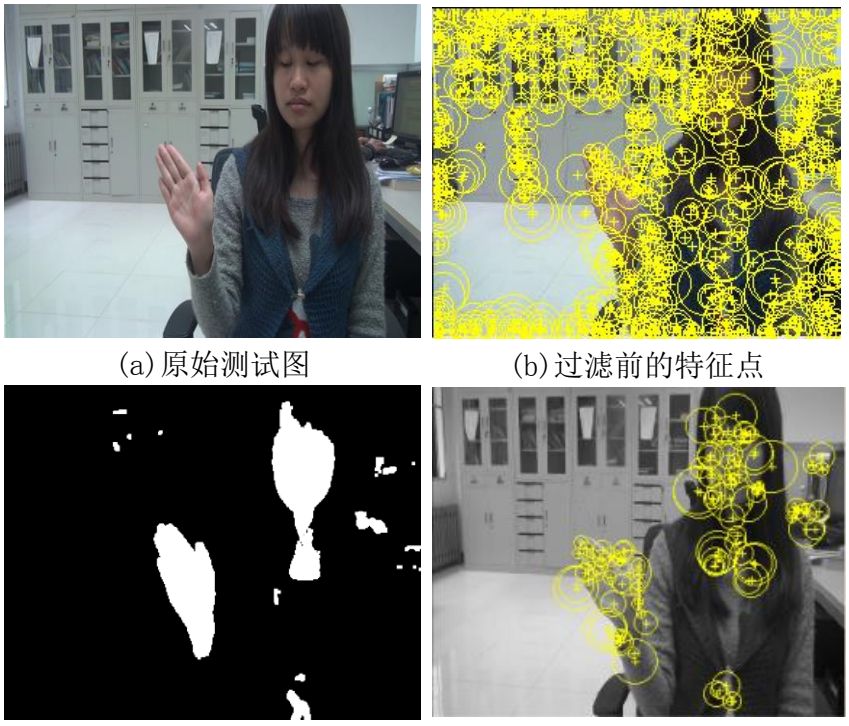
建立好肤色模型后，对于未知像素点，可以先在 YUV 空间上进行初步筛选，再通过计算该像素与模型的距离，确定其属性。计算该像素点的 R、G、B 值和 $N_{skin}(\vec{\mu}_{skin}, \vec{\Sigma}_{skin})$的距离，公式如下：

$$
DisToSkin = ([\mathrm{R},\mathrm{G},\mathrm{B}] - \vec{\mu}_{skin}) * (\vec{\Sigma}_{skin})^{-1} * ([\mathrm{R},\mathrm{G},\mathrm{B}] - \vec{\mu}_{skin})^{T} \tag{4-6}
$$

当 $DisToSkin$ 小于某一经验阈值 $DisThresh$ 时，认为该点为肤色点。

4.1.2 肤色模型在手掌检测中的应用

利用肤色模型筛选肤色区域，可有效过滤大量干扰点，缩小检测范围。为了避免目标区域的特征点被过滤，使用较为宽松的阈值，并对掩模进行膨胀操作，保证目标区域能够完整地被检测到。图 4-3(a)展示了一幅测试样本，图 4-3(c)展示了对应的肤色掩模。



(a) 原始测试图　(b) 过滤前的特征点

(c) 肤色掩模　(d) 过滤后的特征点

图 4-3　使用肤色掩模缩小检测区域

在提取特征点后，首先根据肤色掩模对特征点进行过滤，只有肤色区域的特征点被保留并计算特征值。图 4-3(b)和(d)展示了过滤前和过滤后剩下的特征点。图 4-3(a)为 720*576 的 png 格式图，其中手掌尺寸 108*214。对该测试图进行检测过程中的时间对比如表 4-1。

表 4-1　使用肤色掩模的检测时间对比

时间项（单位：s）	不使用肤色掩模	使用肤色掩模
特征点提取	2.21	2
特征值匹配码本	0. 977	0.442
投票	1.63	0.376
最大假设搜索	5.4	1.9
自顶向下分割	3.27	0.24

由于本文采用的 DoG 角点是整幅图像中提取，因此特征提取的时间并没有被压缩。在对提取的特征点进行过滤后，参与后续特征值计算和匹配等相关步骤的数据量均减少，因此计算时间也减少，从表 4-1 可以对比看出。

综上，膨胀的肤色 mask 可以有效减少计算量。

4.2 码本的组合

4.2.1 多码本组合检测流程

有的目标具有多个子类，如手掌可分为 3 种姿势（见 3.1.1 节）。对每个子类单独训练的码本，只能用于识别该子类的测试样本。为了能够对不同子类的目标进行识别，使用户免于考虑目标所属的子类类型、检测器应用更自然，需要将子类码本组合成父类目标的检测器。

回顾检测流程，主要有四步（参考图 3-11）：提取特征、匹配码本、投票到投票空间、搜索假设。第一步独立于码本；第二步只与码本入口的特征值有关；第三步的输入依赖于第二步的输出，但其结果影响到第四步的搜索范围。

对于不明确子类的测试样本，搜索范围直接影响了检测结果。例如图 4-4 所示的测试样本，其中的手掌既可算作属于张开朝上，也可算作属于张开偏左。用这两种子类的码本检测时，投票空间内均包含与来自该类码本记录的票，可能都有足够大的局部最大值作为假设，也可能会因为票不够密集而不足以形成假设（$score < \theta$）。

由于检测的目标类别是父类，而无需划分子类，因此应该在对父类的投票空间中搜索结果。对于上面这种可能与多个码本匹配的测试样本，如果把各个投票结果统一到一个投票空间中，则票的密集度能反映测试样本与父类的匹配程度，有利于提高检测率。综上所述，投票阶段应把所有的票置于一个投票空间。

图 4-4　目标有子类歧义的测试样本

多码本组合检测器的检测流程如下：

表 4-2　多码本组合检测流程

初始化：码本集 $\mathbb{C}=\{C_1,C_2,...,C_p\}$ ，

$V=createVotingSpace(img.w,stepW,img.h,stepH,s_{\min},s_{\max},stepS)$ 算法表 2-3 为 $vote(img,C,V)$ ，算法表 2-4 为 $doMSME(V)$ 。

1　对于 $\mathbb{C}$ 中的所有码本 C_i

2　　$vote(img,C_i,V)$

3　假设集 $H=doMSME(V)$

4　$removeDupicateH(H)$

$removeDupicateH(H)$ 为假设去重，即对于足够接近的假设，认为是重复的假设，只需保留一个。$removeDupicateH(H)$ 算法流程分为两步，第一步在投票空间去除相近假设，第二步在二维图像上去除重叠率高的假设，详见 4.2.2 节。

4.2.2 假设去重

检测结果为向量 H ， H 的每个元素为一个_Hypothesis 类对象（表 3-10），表示测试图片上目标所在的矩形包围框。对非常靠近或重叠率高的假设，认为指向的是同一目标，只需要保留一个。实验中，对同一个目标产生多个相近的假设的情况是非常多的。这是因为投票空间的票集常表面出密集性。消除重复假设在投票空间和二维图像上分两个阶段进行。

第一阶段根据假设在投票空间的距离去重，算法如图 4-5。

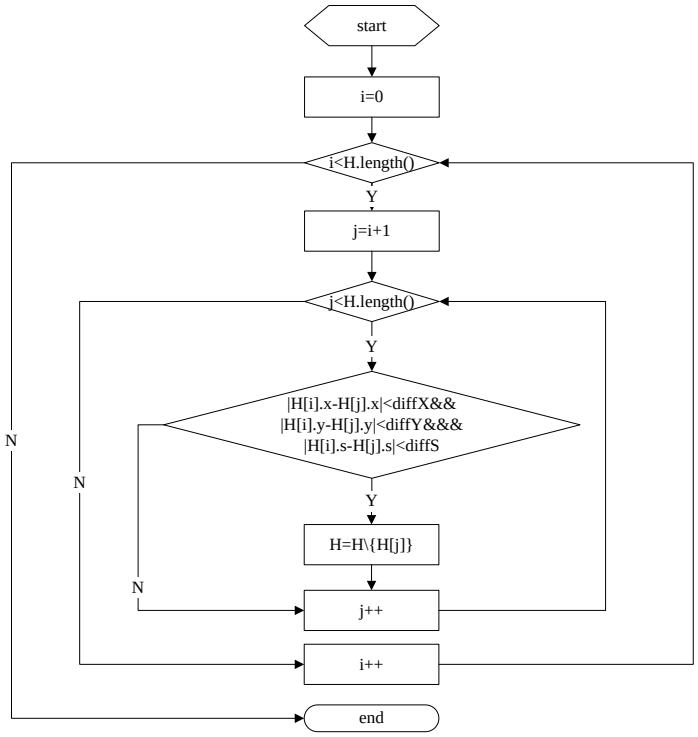


图 4-5　假设去重算法 1

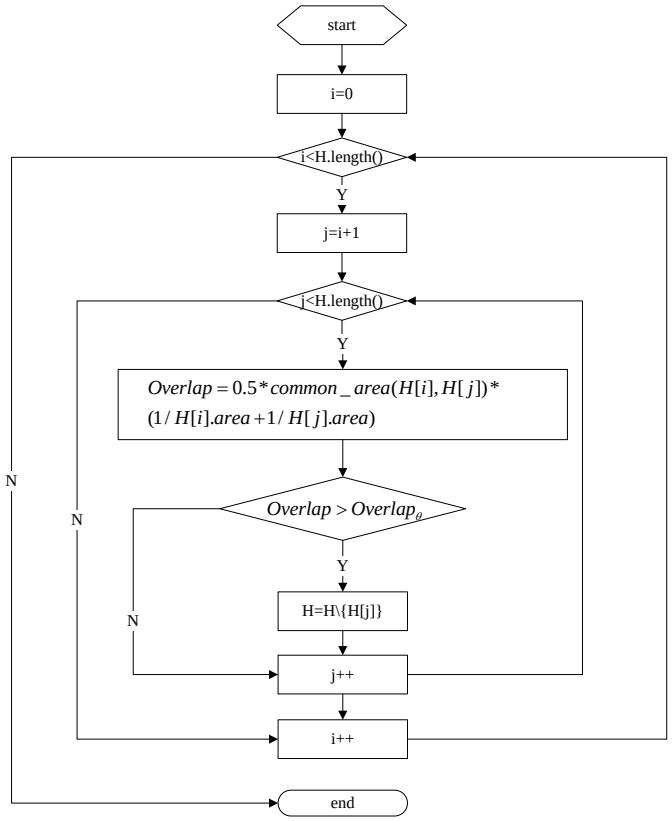


图 4-6　假设去重算法 2

其中 diffX、diffY、diffS 为假设空间三维网格划分的单元 bin 的尺寸。该步骤保证输出的假设在投票空间V 不在一个 bin 内，即保证了一个 bin 的范围内投票结果的唯一性。

第二阶段根据假设在二维图像上的重叠率去重，算法如图 4-6。

算法 2 与算法 1 的区别仅在于消除假设的条件。该步骤确保假设的重叠面积不超过比例阈值$Overlap_\theta$。实验中，$Overlap_\theta$取 0.2。

5　实验过程及结果分析

本章在不改变码本的前提下，通过设定假设阈值θ_h（即表 2-4 的 score 阈值），控制假设生成的条件，得到 ROC 曲线。采用控制变量法，对多个因素对检测结果的影响分别进行了对比分析。

5.1 使用肤色模型前后结果对比

使用肤色掩模对特征点进行过滤时，由于被过滤的都是非肤色区域的噪点，与码本错误匹配的噪点特征将减少，最终在保证检测率不受影响的情况下，误检率将降低。对朝右手掌训练得到的码本，在使用肤色掩模和不使用肤色掩模的情况下，分别用朝右手掌测试集测试，得到了不同的检测结果。

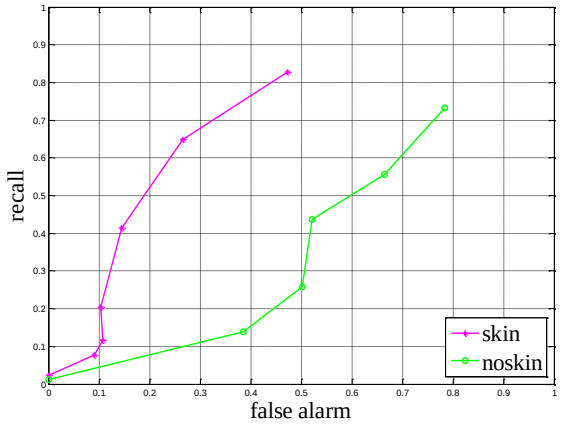


图 5-1　使用肤色掩模前后的 ROC 曲线

如图 5-1，“skin”为使用肤色掩模的 ROC 曲线，“noskin”为不使用肤色掩模的 ROC 曲线。图中(0,1)点是理想情况，达到该点时，误检为零且所有的目标都被正确检测到。ROC 曲线越靠近(0.1)点，则检测器的性能越好。图中“skin”曲线显然更靠近(0,1)点，且在达到同样的 recall 值时，其误检率总是相对较小。可见肤色掩模可以有效排除误检。

表 5-1　不同的假设阈值下使用肤色 mask 前后的检测结果

θ_h	Skin		noskin	
	false alarm	recall	false alarm	recall
1.0	0.103448	0.203125	0.384960	0.138000
0.6	0.145161	0.414062	0.522006	0.437500
0.4	0.265487	0.648438	0.664230	0.556452
0.2	0.472637	0.828125	0.784114	0.733564

上表为图 5-1 中的部分采样点，显示θ_h相同时，使用肤色掩模对误检率的降低效果。可以看出肤色的过滤效果较明显，但另一方面，由于是否使用掩模对投票空间的票布局产生较随机的影响，检测率会发生少量波动。

5.2 码本降噪效果分析

在聚类阶段形成的类中心将作为码本的入口，并在检测阶段与测试样本的特征值匹配。由于本文将与特征 f 匹配成功的各码本入口视作相同，即 $p(c_i \mid f)=1/|c^*|$（见 2.5.1 节），故码本中的非目标特征将会是噪声特征，且有与非噪声特征同样的投票值。当噪声特征与测试样本匹配时，由于其码本入口后的记录往往较少，根据式（2-2）可知记录投票权值 $p(x|c_i,l)$ 较大，因此将会在投票空间产生票权重较大的错误投票。

为了减少码本中的噪声，在聚类完成时，将规模小的类，视作是偶然出现的噪声特征，可以直接舍弃，如此将提高码本中特征的纯度。为了展示降噪效果，本文在训练阶段将类规模为 1 的类舍弃，并分别保存舍弃前后的聚类结果作为码本入口。在空间分布学习阶段，训练样本集分别与两个码本匹配，并得到相应的码本记录。如此产生了两个不同的码本。

用训练集 2 朝上的样本训练，得到降噪前后的两个不同的码本。将测试集 2 的朝上样本分别用两个码本进行测试。

对假设阈值θ_h取各个值时，检测的结果进行了统计，如表 5-2。

表 5-2　不同的假设阈值下码本降噪前后的检测结果

θ_h	cut		nocut	
	false alarm	recall	false alarm	recall
0.5	0.694915	0.337500	0.717514	0.312500
0.4	0.715736	0.350000	0.748744	0.312500
0.3	0.860000	0.568750	0.866460	0.537500
0.2	0.907348	0.725000	0.911129	0.693750
0.1	0.943674	0.812121	0.948405	0.736840

从上表可以看出，假设概率整体偏低。除去噪声后的码本，检测率略高于未除去噪声的码本，而误检率相反。

为更好的对比结果，将两个码本的检测结果以 ROC 曲线呈现。其 ROC 曲线如图 5-2。

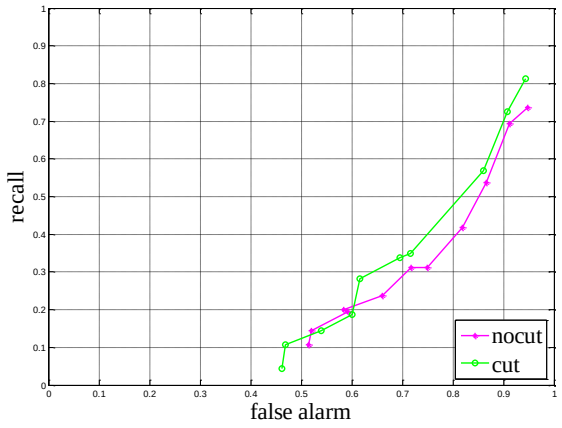


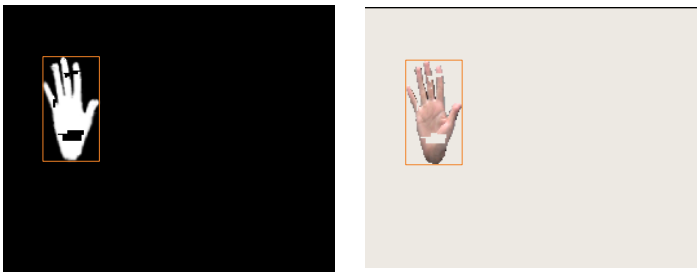
图 5-2　码本降噪前后的 ROC 曲线

如图 5-2 所示，“nocut”为未降噪的码本的 ROC 曲线，“cut”为降噪后的码本的 ROC 曲线。总体上看，对码本降噪有一定的优化效果。

5.3 使用分割使结果更精确

投票结果创建的假设 $h=(x,y,s)$，是对目标所在区域的假设。其中 (x,y) 为假设的目标中心坐标，s 为目标相对于样本的规模系数。若样本中的目标尺寸为 (w_0,h_0)，则测试图的假设目标尺寸为 (sw_0,sh_0)。以 (x,y) 为中心、(sw_0,sh_0) 为尺寸可以创建假设检测框，并根据式（3-1）计算检测结果的正确性。以假设 $h=(x,y,s)$ 作为结果，可以得到假设的 ROC 曲线。

另一方面，由 2.6 节可知，以投票空间局部最大票为中心的核空间内，为参与创建假设的票子集。利用该核空间内的票 $(x,w,occ,l)_i$，对每个特征区域 l，参考 occ 逐像素进行前景与背景投票，最终得到前景与背景的分割结果。对于同一个假设得到的分割结果，用最小的矩形区域包含之，即可得到该假设的分割包围框。图 3-12 的分割包围框如图 5-2。



（a）自顶向下分割　　(b)在原图的分割效果

图 5-3　分割包围框

对图 3-11 的投票结果和分割结果分别标出，如图 5-4，略大的黄色框为分割包围框，相对较小的绿色框为结果假设框。直观地看，分割结果比假设结果更加准确。

图 5-4　假设与分割结果

对朝上的手掌进行训练和测试，测试阶段采用降噪的码本和肤色过滤。将 θ_h 相同时假设和分割结果分别作出 ROC 曲线，见图 5-5。

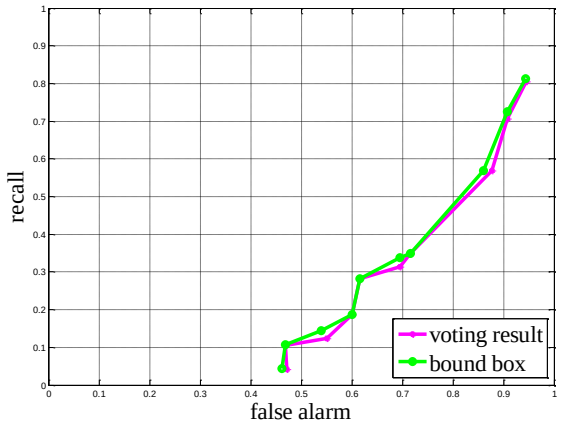


图 5-5　投票结果与分割结果

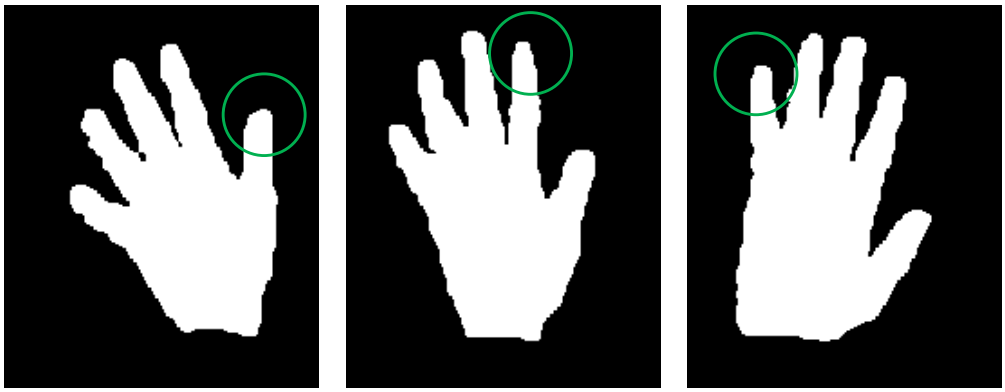
图 5-5 展示了对相同的检测结果，分别用假设框和分割框的形式表示，并与标注的目标区域比较，得到的两条 ROC 曲线。其中“voting result”为假设框的 ROC 曲线，“bound box”为分割包围框的 ROC 曲线。可以看出，分割结果比假设搜索结果更精确。

5.4 组合码本对整体性能的提升

对样本子类单独训练和检测，其码本规模较小，匹配的特征较少。对于手掌检测而言，由于子类之间存在较多的共有特征，一种子类测试样本与另一子类的码本也能匹配。

如图 5-6 所示，(a)、(b)、(c)分别来自三个子类，其圆框内的特征相似，均为竖直朝上的手指头。那么在三个子类的码本中，都会出现这种特征及对应的码本记录。例如对一个偏左的测试样本，如果对该样本提取的特征中包含竖直朝上的手指这一特征，

该特征不仅与偏左的码本匹配，朝上和偏右这两个码本也有与之匹配的入口。



(a) 偏左样本掩模　　(b) 朝上样本掩模　　(c) 偏右样本掩模

图 5-6　子类共有特征示例（绿色圆框内为共有特征）

将码本进行组合（组合码本的检测流程详见 4.2.1 节），可以有效提高检测率。对于具备共有特征的测试样本，组合码本提供了匹配多个码本的机会。由于组合码本的检测流程是依次与各个码本匹配，再将所有的票集中到一个投票空间，这样对于共有特征，在投票空间的票会增加，且单个票的权重并没有降低。如此以来，该特征在假设搜索阶段将对产生相对于单码本更大的影响，该样本被检测的可能性增加。

对于并不具备共有特征的测试样本，从这种样本提取的特征与组合码本匹配和与单个码本匹配的结果基本相同，因此组合码本的方法对于该类样本也不会降低检测率。

对于朝上的测试样本，单码本与组合码本检测的 ROC 曲线见图 5-7。

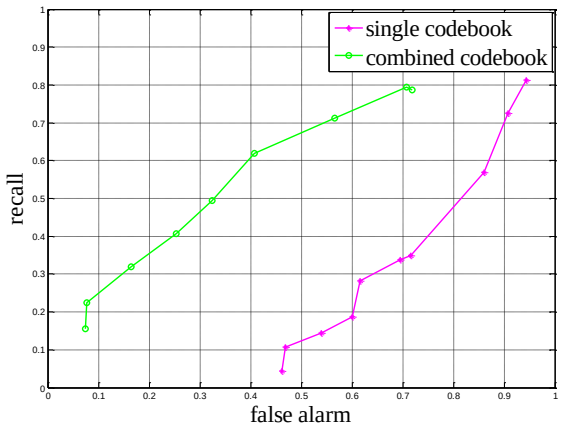


图 5-7　单码本与组合码本对单类的检测结果对比

图 5-7 展示了使用单类码本和组合码本分别对朝上类样本进行测试的 ROC 曲线。其中“single codebook”为单码本的检测结果，“combined codebook”为组合码本的检测结果。可以看出，对手掌这种多姿态目标而言，由于共有特征的大量存在，组合码本对检测率提升较明显。

为更好的对比结果，对假设阈值 θ_h 取各个值时，检测的结果进行了统计，如表 5-3。

表 5-3　不同的假设阈值下单码本与组合码本的检测结果

θ_h	Single Codebook		Combined Codebook	
	false alarm	recall	false alarm	recall
1.0	0.468750	0.106250	0.252874	0.406250
0.8	0.600000	0.187500	0.324786	0.493750
0.6	0.615385	0.281250	0.407186	0.618750
0.4	0.715736	0.350000	0.566540	0.712500
0.2	0.907348	0.725000	0.708046	0.793750
0.1	0.943674	0.812121	0.719376	0.787500

从上表可以看出，单码本检测整体上检测率偏低而误检率偏高。组合码本在整体上提升了检测率，但偶尔亦有小幅波动，如 θ_h=0.1 时，组合码本的检测率略低于单码本。

5.5 整体检测效果分析

使用肤色模型和组合码本优化后，ISM 分类器对所有测试样本的检测性能如图 5-8。

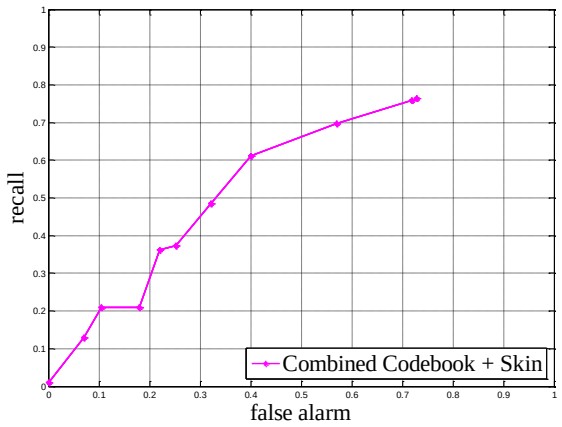


图 5-8　优化后的检测器对所有类型测试集的 ROC 曲线

可以看出，检测器整体上误检率较高。当检测率达到 76.46%时，此时的误检率为 72.91%。分析其原因主要在于两点：

一、选取的特征分辨性能不高，检测阶段背景中的特征可与码本中的特征匹配。背景特征与目标码本特征发生匹配，有特征自身性质的原因，更重要的是特征的挑选方法原因。与分类器不同，隐形状模型中的特征全部由正样本聚类而得，对负样本是否具备分辨能力，则无法得知。

二、手部整体较光滑，特征点较少，可表现手掌的特征也较少。从图 2-1 可见，相对于背景区域，手部的特征点密度较低。

针对当前模型存在的问题，以降低误检率为重点，有若干改进思路。

本文对于均值漂移结果的概率达到假设阈值都予以通过。由于手掌整体形态光滑，特征点较少，与码本匹配偏少，导致投票空间中的有效票数较少，造成假设概率不大。而背景中许多特征与手部特征相似，可以匹配到码本，导致投票空间的噪声票较多，造成误检较多。本文后续没有对假设作进一步检验，而假设阈值对假设的正确性并没有分辨能力。

Leibe[9]在其车辆检测实验中，采用了最小描述长度准则（Minimal Description Length，简称 MDL），根据信息量最小的原理进行了假设筛选，取得了较好的效果。图 1-1 中的红线为采用了 MDL 的检测性能，可以看出，相比基本的 ISM 算法（黑线），MDL 对误检的排除效果较明显。

Seemann[25]在其行人检测实验中，采用 Chamfer Matching，和目标分割结果结合起来，也起到了消除误检的效果。图 5-9 对比了标准 ISM 算法和 Chamfer Matching 扩展后的 ISM 算法，数据来源于对同规模行人检测的结果。此外，Breitenstein[38]采用了置信度检验的方法。假设检验是手掌识别后期工作的方向之一。

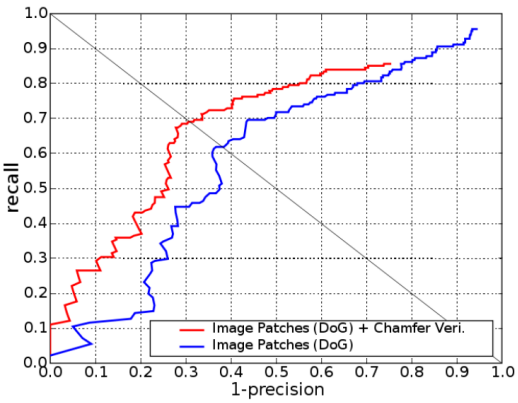


图 5-9　标准 ISM 与添加了 Chamfer 检验模块的 ISM 对行人检测的性能对比

除了对检测结果进行二次过滤，也可以参考分类器的训练思想，制作正样本和负样本，挑选对于正负样本有分辨性能的特征。如 Adaboost 进行弱分类器训练时，选择对正负样本有弱分类性能的特征作为弱分类器，再将弱分类器经过概率组合构成强分类器。借鉴这种思想，可以对随机背景构成的负样本构建负码本，用于剔除正码本中的与负码本共有的特征，或者在投票阶段建立负票机制。构建具有分类性质的隐形状模型，也是值得尝试的方向。

6 结论与展望

6.1 结论

本文以西安交大与长虹研究院合作的软件与智能交互方向的新型手势识别项目为背景，研究和实现了一种基于隐形状模型的手掌检测器用于手势检测系统。本文通过分析人体手势的构成特点，考虑到多姿态特征，在进行文献调研分析之后，采用隐形状模型进行实验。并根据手掌的肤色和多姿态特点，采用肤色过滤、分姿态训练并组合码本的方式，对原始的检测器进行了优化。

本文通过对隐形状模型的研究和运用，对手掌识别方法进行了有益的探索。本文的主要工作如下：

（1）研究了隐形状模型的原理。从训练与检测两个阶段分别阐述，为后文的结果分析与优化引入了理论基础。

（2）介绍了将隐形状模型应用于手掌检测的过程。经过形态上的分析比较，定义了三种手势，并介绍了样本的制作方法。根据距离可变的应用需求采用具有规模不变性的 DoG 特征点，根据手掌灰度稳定的特点采用灰度特征。并分别介绍了 DoG 特征点检测算子和特征值计算方法。

（3）针对手掌的特点对检测流程进行了优化。首先根据肤色特征，用肤色过滤减少了检测阶段输入的特征数量，从而压缩了运算时间，并降低了误检率。其次根据多姿态手掌共享特征的特点，将多个姿态的手掌训练得到的码本加以组合，该方法有效提高了检测率。

（4）从多个角度对比了检测结果。不仅对肤色和组合码本两个优化策略的效果进行了比较分析，还从码本降噪和投票结果与像素分割结果两个方面进行了分析。

本文的测试结果表明，优化后的隐形状模型检测器，对于大小为 720*576、其中目标手掌宽度范围为 100~130、高度范围为 180~250 的样本集，其最高检测率可以达到 76.46%，此时的误检率为 72.91%。

6.2 展望

由于利用摄像头进行手掌检测的应用目前不如人脸识别、车辆识别等的应用广泛，对此展开的研究相对而言偏少，因此目前没有通用的手掌样本库。本文由于实验条件的限制，收集的样本来源较少。建立通用的手掌样本库，不仅有益于训练得到通用性更强的检测器，也有益于学术交流。

特征方面，本文是采用单个特征进行训练和检测，未来可以考虑其它特征点检测子和特征类型，也可以将多种特征融合使用。本文将聚类结果中类规模为 1 的类进行了裁剪，对于特征池过大的情况，可以采取更加严厉的小类消除策略。

速度方面，虽然本文通过肤色缩小了计算量，但检测耗时较长，达不到实时。对一幅典型测试图检测，各阶段耗时见表 6-1。

表 6-1 整体检测时间

时间项	耗时（单位：s）
特征值提取	1.94
特征值匹配码本	0.352
投票	0.349
最大假设搜索	1.84
自顶向下分割	0.026
总计	4.507

对表 6-1 分析可知，除了特征点提取耗时较大外，在投票空间进行假设搜索也比较耗时，这与投票空间票数过多和搜索算法都有关系。后期工作可以设法消除码本中的噪点，或控制匹配码本入口的数量。可以将码本中的入口按聚类规模或记录个数排列，使典型特征优先得到匹配。此外，可以尝试更优的假设搜索策略。

精度方面，本文主要对基本 ISM 算法在手掌检测的应用作了研究和分析，后期需要结合手掌检测的特点对算法进行扩充。如 Leibe[9]针对汽车等目标使用的 MDL 检验方法， Seemann[25]针对行人检测的 Chamfer Matching 检验法等，都是后期工作可借鉴的方向。

致　谢

这次毕业论文能够得以顺利完成，是所有曾经指导过我的老师，帮助过我的同学，一直支持着我的家人对我的教诲、支持和鼓励的结果。我要在这里对他们表示深深的谢意！

首先，我要感谢我的导师梅魁志老师。他不仅对我的论文进行了直接的指导，在整个研究生阶段的学习生涯中，也时时从生活和工作上提出有益的建议。他的严谨勤奋的工作态度，是为人师表的最直接的诠释。

感谢人工智能与机器人研究所的各位老师。这两年的学业生活都是在人机所度过的，感谢老师们对我的关心和严格要求，这里浓厚的学术氛围让我受益良多。

感谢课题组的李博良老师。李老师刻苦钻研的精神影响我很深，而他的专业知识和方法教会我怎样去提高效率。在实验过程中，李老师从工程技巧和算法评价方法上给予了精心的指导。

感谢教研室的伙伴们。王芳、彭静帆、席宝、高增辉协助了样本的制作，刘冬冬在我对 Linux 平台不熟悉时提供了许多帮助，解筱娜、林斌、崔继岳、董佩在论文的写作方面传授了许多经验，丑文龙曾亲自调试我的代码，张冀在测试的方法上提出了建议……诸如此类充满关怀的帮助，使我在科研阶段沉浸在友爱的氛围中。在平时的生活中，我们互相帮助，共同进步，建立了深厚的友谊。

感谢两位舍友，她们单纯可爱的个性，让我的宿舍总是充满了欢乐和正能量。她们让我及时发现并弥补了自身的缺点和不足，使我更加成熟。两年同吃同住的岁月，是我人生中一段美丽的回忆。

最后，我要感谢家人。他们在我成长的路上付出了巨大的心血，是我最坚强的后盾。他们的支持与理解、无私的爱，是我源源不断的动力。

参考文献

[1] 王晓琳．基于计算机视觉的手势识别人机交互技术[J]．硅谷，2009，(19)．

[2] 武霞，张崎，许艳旭．手势识别研究发展现状综述[J]．电子科技，2013，26(6)：171-174．

[3] 胡友树．手势识别技术综述[J]．中国科技信息，2005，(2)：42-42．

[4] 翁汉良．基于单目视觉的手势识别算法的研究与实现[D]：广东工业大学，2011．

[5] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features[C]: in Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001,1: 511-518.

[6] Lienhart R, Maydt J. An extended set of haar-like features for rapid object detection[C]: in Proceedings of the IEEE International Conference on Image Processing, 2002,1: 900-903.

[7] Benenson R, Mathias M, Timofte R, et al. Pedestrian detection at 100 frames per second[C]: in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2012: 2903-2910.

[8] Zhu Q, Yeh MC, Cheng KT, et al. Fast human detection using a cascade of histograms of oriented gradients[C]: in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2006, 2: 1491-1498.

[9] Leibe B, Leonardis A, Schiele B. Robust object detection with interleaved categorization and segmentation[J]. International Journal of Computer Vision, 2008, 77 (1): 259-289.

[10] Sabzmeydani P, Mori G. Detecting pedestrians by learning shapelet features[C]: in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2007: 1-8.

[11] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]: in Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005, 1: 886-893.

[12] Sun Z, Bebis G, Miller R. On-road vehicle detection: A review[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28 (5): 694-711.

[13] Leung B. Component-based car detection in street scene images[D]: Massachusetts Institute of Technology, 2004.

[14] Francke H, Ruizdel SJ, Verschae R. Real-time hand gesture detection and recognition using boosted classifiers and active learning. Advances in Image and Video Technology[M]. Springer, 2007: 533-547.

[15] Zhang Q, Chen F, Liu X. Hand gesture detection and segmentation based on difference background image with complex background[C]: in Proceedings of IEEE International Conference on Embedded Software and Systems, 2008: 338-343.

[16] Angelopoulou A, Rodr ́guez JG, Psarrou A. Learning 2d hand shapes using the topology preservation model gng. European Conference on Computer Vision[M]. Springer, 2006: 313-324.

[17] Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting[J]. Journal of Computer and System Sciences, 1997, 55 (1): 119-139.

[18] Wu B, Nevatia R. Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors[C]: in Proceedings of IEEE 10th International Conference on Computer Vision, 2005, 1: 90-97.

[19] Leibe B, Ettlin A, Schiele B. Learning semantic object parts for object categorization[J]. Image and Vision Computing, 2008, 26 (1): 15-26.

[20] Leibe B, Leonardis A, Schiele B. Combined object categorization and segmentation with an implicit

shape model[C]: in Proceedings of Workshop on Statistical Learning in Computer Vision, 2004, 2: 7-14.
[21] Leibe B, Schiele B. Interleaving object categorization and segmentation[M]: Springer, 2006.
[22] Jones MJ, Poggio T. Model-based matching by linear combinations of prototypes[J]. MIT AI Memo 1583, MIT, 1996.
[23] Ullman S. Three-dimensional object recognition based on the combination of views[J]. Cognition, 1998, 67 (1): 21-44.
[24] 张路平，李飚，王鲁平等．基于两级隐式形状模型的抗遮挡目标跟踪[J]．国防科技大学学报，2013，35：6-6．
[25] Seemann E. Pedestrian Detection in Crowded Street Scenes[M]: Hartung-Gorre, 2007.
[26] Shi J, Tomasi C. Good features to track[C]: in Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1994: 593-600.
[27] Leibe B, Schiele B. Scale-invariant object categorization using a scale-adaptive mean-shift search[M]: Springer, 2004.
[28] Hough PV. Method and means for recognizing complex patterns[M]. Google Patents, 1962.
[29] Ballard DH. Generalizing the Hough transform to detect arbitrary shapes[J]. Pattern Recognition, 1981, 13 (2): 111-122.
[30] Lowe DG. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60 (2): 91-110.
[31] Fukunaga K, Hostetler L. The estimation of the gradient of a density function, with applications in pattern recognition[J]. IEEE Transactions on Information Theory, 1975, 21 (1): 32-40.
[32] Cheng Y. Mean shift, mode seeking, and clustering[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1995, 17 (8): 790-799.
[33] Collins RT. Mean-shift blob tracking through scale space[C]: in Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003: II-234-240 vol. 232.
[34] Comaniciu D, Meer P. Mean shift: A robust approach toward feature space analysis[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24 (5): 603-619.
[35] Comaniciu D. Image segmentation using clustering with saddle point detection[C]: in Proceedings of IEEE International Conference on Image Processing, 2002, 3: 297-300.
[36] Stauffer C, Grimson WEL. Adaptive background mixture models for real-time tracking[C]: in Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1999, 2:18-20.
[37] 陈锻生，刘政凯．肤色检测技术综述[J]．计算机学报，2006，29(2)：194-207．
[38] Breitenstein MD, Reichlin F, Leibe B, et al. Robust tracking-by-detection using a detector confidence particle filter[C]: in Proceedings of IEEE 12th International Conference on Computer Vision, 2009: 1515-1522.

学位论文独创性声明（1）

本人声明：所呈交的学位论文系在导师指导下本人独立完成的研究成果。文中依法引用他人的成果，均已做出明确标注或得到许可。论文内容未包含法律意义上已属于他人的任何形式的研究成果，也不包含本人已用于其他学位申请的论文或成果。

本人如违反上述声明，愿意承担以下责任和后果：

1．交回学校授予的学位证书；

2．学校可在相关媒体上对作者本人的行为进行通报；

3．本人按照学校规定的方式，对因不当取得学位给学校造成的名誉损害，进行公开道歉。

4．本人负责因论文成果不实产生的法律纠纷。

论文作者（签名）：　　日期：　　年　　月　　日

学位论文独创性声明（2）

本人声明：研究生　　所提交的本篇学位论文已经本人审阅，确系在本人指导下由该生独立完成的研究成果。

本人如违反上述声明，愿意承担以下责任和后果：

1．学校可在相关媒体上对本人的失察行为进行通报；

2．本人按照学校规定的方式，对因失察给学校造成的名誉损害，进行公开道歉。

3．本人接受学校按照有关规定做出的任何处理。

指导教师（签名）：　　日期：　　年　　月　　日

学位论文知识产权权属声明

我们声明，我们提交的学位论文及相关的职务作品，知识产权归属学校。学校享有以任何方式发表、复制、公开阅览、借阅以及申请专利等权利。学位论文作者离校后，或学位论文导师因故离校后，发表或使用学位论文或与该论文直接相关的学术论文或成果时，署名单位仍然为西安交通大学。

论文作者（签名）：　　日期：　　年　　月　　日

指导教师（签名）：　　日期：　　年　　月　　日