

工学博士学位论文

基于局部纹理特征的物体检测方法

OBJECT DETECTION BASED ON LOCAL TEXTURE  
FEATURES

张洪明

哈尔滨工业大学

2006 年 2 月

Classified Index: TP393.41

U.D.C.: 681.39

Dissertation for the Doctoral Degree in Engineering

# OBJECT DETECTION BASED ON LOCAL TEXTURE FEATURES

<b>Candidate:</b>	Zhang Hongming
<b>Supervisor:</b>	Prof. Gao Wen
<b>Academic Degree Applied for:</b>	Doctor of Engineering
<b>Specialty:</b>	Computer Application Technology
<b>Date of Oral Examination:</b>	February 25, 2006
<b>University:</b>	Harbin Institute of Technology

## 摘要

从计算机诞生以来，让计算机或机器人像人类一样具有视觉能力，是科研工作者一直不懈追求的目标。物体检测是人类视觉中的基本步骤和基本功能，为人类了解周围的环境和景物提供了至关重要的前提。物体检测是计算机视觉学科中极具挑战性的理论研究课题，而且在很多领域具有广阔的应用前景，例如多媒体信息分析与搜索、场景分析与理解、视频编码、视频监控、模式识别等。作为一个研究热点，物体检测越来越受到研究者们的重视，研究深度和广度在不断地增强。

本文围绕物体检测展开研究，研究目的是探索一种有效的物体特征表示方法并且应用于物体检测，从而提供物体检测的新方法。针对上述研究目标，本文在基于局部纹理分析的物体表示特征、有效特征的选择和基于局部纹理特征的物体检测算法等方面进行了深入的研究和探讨。论文具体的研究内容如下：

1. 本文基于物体的局部纹理分析，提出了改进的空间直方图特征，用于物体检测任务中的特征表示。空间直方图特征是图像局部纹理的信息分布，它同时反映了物体在不同尺度上的纹理和结构信息。改进的空间直方图特征与目标物体类别相关联，具有对目标物体的判别能力，而且能够适用于各种典型物体的表示。本文以刻画物体的纹理信息和结构信息为着眼点，建立了空间直方图特征的提取方法，并且对空间直方图特征的物体判别能力进行了研究。

2. 提出了基于空间直方图特征的层次化物体检测算法。本文采用空间直方图特征作为物体表示，根据由粗到精的策略，综合利用两种物体检测技术：直方图匹配方法和支持向量机分类器，构建了物体检测的一种通用算法。该算法首先通过直方图匹配方法进行粗检测，达到排除图像中大量非目标物体的目的；其次在精确检测过程中，采用支持向量机分类器提高检测精度。该算法有效的解决了物体检测问题，具有快速、鲁棒的特点。该算法可以处理不同类型的物体模式，包括刚体物体模式、可变形的柔性非刚体物体、以及纹理模式物体。

3. 提出了一种结合类别可分性和特征相关性的特征选择方法。传统的特征选择技术在处理高维数据时，存在训练时间长或者选择所得到的特征子集分类性能不高的局限。本文方法使用 Fisher 准则度量特征的类别可分

性，采用互信息计算特征之间的相关性，按照顺序增加的方式产生候选特征，以分类器错误率最小为目标，选择分类性能高而互相之间相关性弱的特征，构成紧致而有效的特征子集。该方法不但可以选择出有效的分类特征，而且在保证分类性能的条件下降低了特征维数，提高了分类效率。本文将所提出的特征选择方法用于构造空间直方图特征子集，作为物体检测中支持向量机分类器算法的输入特征。实验结果表明该方法可以自动获取与目标物体类别相关联的分类特征。

4. 提出了基于空间直方图特征和纠错码分类器的多姿态人脸检测算法。首先，本文结合纠错码多类分类器算法，研究了如何将基于空间直方图特征的物体检测算法扩展到多姿态人脸检测。其次，针对基于纠错码多类分类算法中单分类器训练困难的问题，本文提出了一种在给定纠错码码本的前提下如何训练有效纠错码多类分类器的方法。该方法以整体纠错码多类分类器的错误率最小为目标，依次选择有效特征和训练单分类器，从而训练得到整体分类性能更高的纠错码多类分类器。最后，本文对基于纠错码的多姿态人脸检测方法进行了实验验证，实验结果表明该方法可以有效的解决多姿态人脸检测问题，取得了良好的检测性能。

**关键词** 物体检测；多姿态人脸检测；空间直方图特征；特征选择；纠错码多类分类器

## Abstract

It was about the initial stage of the computer coming into the world that scientists and engineers began to pursue creating a machine that has the vision ability similar to human beings. Object detection is the fundamental function and process of human vision. It provides essential preconditions for human beings to understand scenes in the real world. As an active research area, object detection has been attracted more and more attention. In computer vision community, object detection has been a very challenging research topic. Moreover, the object detection technology is significant in many real applications.

This dissertation aims to provide a generic object detection framework by exploring novel features for object pattern representation and applying them in automatic object detection. Some key issues and techniques of object detection are discussed and lucubrated, such as object feature extraction methods based on local texture analysis, efficient feature selection methods, object detection algorithms based on local texture features, and multi-view face detection algorithm. The detailed descriptions of the above techniques are as follows.

1. An object representation method based on local texture analysis (i.e. a kind of improved spatial histogram features) is proposed in this dissertation. Motivated by the observation that objects have texture distribution and shape configuration, this dissertation proposes spatial histogram based features (termed as spatial histogram features) to represent objects. As spatial histograms consist of marginal distributions of an image over local patches, the information about texture and shape at different scales of the object can be encoded simultaneously. This dissertation presents the extraction method of spatial histogram features and gives quantitative analysis of the discriminative ability of spatial histogram features.

2. A generic object detection approach based on spatial histogram features is presented. This approach uses spatial histogram features as object representation, employs histogram matching and support vector machine to construct an object detector based on a coarse-to-fine strategy. The object detector receives image samples of a fixed size as inputs and initially produces spatial histogram features

from the image samples. In the coarse detection stage, the cascade histogram matching method rejects most non-object instances. In the fine detection stage, a support vector machine is used to verify whether or not the remained samples are object instances. The proposed method is efficient and robust to object detection, and it can be extended to generic object detection for different object types including rigid objects, non-rigid objects and texture rich objects.

3. A feature selection approach based on class separability and feature correlation is presented in this dissertation. Traditional feature selection methods have limitations, such as expensive training time of selection procedure and low classification accuracy of the selected features. The proposed method uses Fisher criterion to measure class separability of each feature and employs mutual information to calculate features correlation. Optimal feature subset is constructed by selecting uncorrelated and discriminative features through minimizing the classification error rate. The proposed approach is applied to the spatial histogram features based object detection tasks to learn spatial histogram features for support vector machine. The experiment results show that the proposed feature selection method is efficient to extract informative and class-specific features for object detection.

4. A multi-view face detection method based on spatial histogram features and ECOC (error correcting output codes) multi-class classification is presented. First, we use the ECOC multi-class classification framework to extend the spatial histogram features based object detection approach to solve the multi-view face detection problem. Second, we present an approach for designing efficient binary classifiers by learning informative features through minimizing the error rate of the ensemble ECOC multi-class classifier. Finally, we conduct experiments to evaluate the effectiveness and robustness of the proposed multi-view face detection approach.

**Keywords** Object detection; multi-view face detection; spatial histogram features; feature selection; ECOC (error correcting output codes) based multi-class classifier

# 目录

摘要.....	I
Abstract.....	III
<b>第 1 章 绪论.....</b>	<b>1</b>
1.1 课题背景及研究意义 .....	1
1.2 物体检测领域的研究现状.....	3
1.2.1 物体特征表示 .....	3
1.2.2 物体检测方法 .....	7
1.2.3 物体检测中的典型案例：人脸检测 .....	9
1.2.4 物体检测中存在的问题.....	12
1.3 本文主要工作及结构安排.....	12
1.3.1 论文的主要工作 .....	14
1.3.2 论文的结构安排 .....	14
1.3.3 论文的主要贡献 .....	15
<b>第 2 章 空间直方图特征及其判别分析 .....</b>	<b>17</b>
2.1 引言.....	17
2.2 空间直方图特征的提取 .....	17
2.2.1 纹理直方图 .....	18
2.2.2 空间直方图特征 .....	23
2.3 空间直方图特征的判别分析.....	24
2.3.1 类别可分性的实例说明 .....	24
2.3.2 类别可分性的度量 .....	28
2.4 空间直方图特征在彩色图像人脸检测中的应用 .....	29
2.4.1 基于彩色信息的空间直方图特征 .....	29
2.4.2 检测方法 .....	30
2.4.3 实验评估 .....	32
2.5 本章小结 .....	34
<b>第 3 章 基于空间直方图特征的物体检测.....</b>	<b>36</b>
3.1 引言.....	36
3.2 物体检测的系统框架 .....	37

3.3 联合直方图匹配 .....	38
3.3.1 联合直方图匹配的推导 .....	38
3.3.2 联合直方图匹配的训练方法 .....	41
3.4 支持向量机分类器及其特征选择 .....	44
3.4.1 空间直方图特征之间的相关性 .....	45
3.4.2 空间直方图特征的有效选择 .....	45
3.5 实验评估 .....	50
3.5.1 正面人脸检测 .....	51
3.5.2 侧面汽车检测 .....	57
3.5.3 视频文字检测 .....	63
3.5.4 关于系统检测速度的实验 .....	66
3.6 本章小结 .....	68
<b>第 4 章 基于纠错码多类分类器的多姿态人脸检测 .....</b>	<b>69</b>
4.1 引言 .....	69
4.2 相关研究工作介绍 .....	69
4.3 基于纠错码的多类分类方法 .....	71
4.3.1 信息传输系统模型与纠错码 .....	72
4.3.2 将纠错码应用于多类分类方法 .....	74
4.4 基于纠错码的多姿态人脸检测 .....	75
4.5 构建基于空间直方图特征的ECOC分类器 .....	77
4.6 实验评估 .....	80
4.6.1 纠错码性能分析 .....	82
4.6.2 在标准测试集合上的实验结果 .....	83
4.6.3 纠错码与一对多编码的比较 .....	86
4.6.4 实验结论 .....	88
4.7 本章小结 .....	88
<b>结论 .....</b>	<b>90</b>
<b>参考文献 .....</b>	<b>92</b>
<b>攻读学位期间发表的学术论文 .....</b>	<b>101</b>
<b>哈尔滨工业大学博士学位论文原创性声明 .....</b>	<b>103</b>
<b>哈尔滨工业大学博士学位论文使用授权书 .....</b>	<b>103</b>
<b>致谢 .....</b>	<b>104</b>
<b>个人简历 .....</b>	<b>105</b>



## Contents

<b>Chinese Abstract.....</b>	<b>I</b>
<b>English Abstract .....</b>	<b>III</b>
<b>Chapter 1 Introduction .....</b>	<b>1</b>
1.1 Background .....	1
1.2 State of the art of object detection.....	3
1.2.1 Features for object representation.....	3
1.2.2 Object detection methods .....	7
1.2.3 A typical case of object detection:face detection .....	9
1.2.4 Existing problems of object detection .....	12
1.3 Contributions and organization of the thesis .....	12
1.3.1 Main research work of the thesis .....	14
1.3.2 Structure of the thesis .....	14
1.3.3 Summary of the contributions .....	15
<b>Chapter 2 Spatial histogram features and their discriminating analysis..</b>	<b>17</b>
2.1 Introduction .....	17
2.2 Extraction method of spatial histogram features .....	17
2.2.1 Texture histogram .....	18
2.2.2 Spatial histogram features .....	23
2.3 Discriminating analysis of spatial histogram features .....	24
2.3.1 Illustrations of class separability for spatial histogram featuers .....	24
2.3.2 Measurement of class separability .....	28
2.4 Using spatial histogram features for face detection in color images .....	29
2.4.1 Spatial histogram features using color information .....	29
2.4.2 Face detection method .....	30
2.4.3 Experiments .....	32
2.5 Summary .....	34
<b>Chapter 3 Object detection using spatial histogram features .....</b>	<b>36</b>
3.1 Introduction .....	36
3.2 Overview of the proposed object detection method.....	37

3.3 Cascade histogram matching .....	38
3.3.1 Definition of cascade histogram matching .....	38
3.3.2 Training method of cascade histogram matching.....	41
3.4 Feature selection method for support vector machine .....	44
3.4.1 Feature correlation measurement of spatial histogram features.....	45
3.4.2 Efficient selection of spatial histogram features .....	45
3.5 Experiments.....	50
3.5.1 Frontal face detection .....	51
3.5.2 Side-view car detection .....	57
3.5.3 Video text detection.....	63
3.5.4 Performance time .....	66
3.6 Summary .....	68
<b>Chapter 4 Multi-view face detection using ECOC.....</b>	<b>69</b>
4.1 Introduction .....	69
4.2 Related work .....	69
4.3 Multi-class classification method based on ECOC.....	71
4.3.1 Information transimitting model and ECOC.....	72
4.3.2 Using ECOC to solve multi-class problems .....	74
4.4 Multi-view face detection based on ECOC codes .....	75
4.5 Construct ECOC classifier based on spatial histogram features.....	77
4.6 Experiments.....	80
4.6.1 Performance analysis of ECOC codes for face detection .....	82
4.6.2 Testing results on one standard test data set .....	83
4.6.3 Comparison between ECOC codes and one-against-others .....	86
4.6.4 Summary of experiments.....	88
4.7 Summary .....	88
<b>Conclusions.....</b>	<b>90</b>
<b>References.....</b>	<b>92</b>
<b>Papers.....</b>	<b>101</b>
<b>Statements of Copyright .....</b>	<b>103</b>
<b>Letter of Authorization .....</b>	<b>103</b>
<b>Acknowledgement .....</b>	<b>104</b>
<b>Resume .....</b>	<b>105</b>

## 图表目录

图 1-1 物体检测的示例 .....	2
图 1-2 基于局部纹理特征的物体检测方法的研究框架图 .....	13
图 2-1 LBP计算中的邻域像素 .....	18
图 2-2 计算LBP的实例.....	19
图 2-3 人脸(a)和非人脸(b)的图像、LBP图像和LBP直方图 .....	19
图 2-4 侧面汽车（前 3 行）和非汽车（后 3 行）的图像、LBP图像和 LBP直方图 .....	20
图 2-5 通过空间直方图在多尺度上衡量物体的局部纹理信息 .....	22
图 2-6 空间模板的示意图 .....	22
图 2-7 空间直方图特征的提取框架 .....	23
图 2-8 特征分布：侧面汽车和非汽车 .....	25
图 2-9 特征分布：人脸和非人脸 .....	26
图 2-10 特征分布：文字和非文字 .....	27
图 2-11 人脸模式的 23 个空间模板 .....	30
图 2-12 直方图匹配方法 .....	31
图 2-13 直方图匹配特征区域：(a)人脸整体区域，(b)双眼区域，(c) 中 间竖直区域 .....	31
图 2-14 两级直方图匹配的人脸检测流程图.....	32
图 2-15 彩色图像中人脸检测的示例 .....	33
图 2-16 光照变化剧烈的彩色图像中人脸检测示例 .....	34
图 3-1 基于空间直方图特征的单类目标物体检测 .....	38
图 3-2 极大似然判决规则简化形式的示意图：形式(a)和形式(b) .....	40
图 3-3 联合直方图匹配训练算法的流程图.....	43
图 3-4 结合可分性和相关性的特征选择的系统结构 .....	47
图 3-5 正面人脸训练样本的示例 .....	51
图 3-6 正面人脸检测为联合直方图匹配选取的 11 个空间模板 .....	51
图 3-7 正面人脸检测算法在CMU+MIT测试集上的ROC曲线.....	52
图 3-8 正面人脸检测的示例 .....	54
图 3-9 CAS_PEAL_R1 子集上的正面人脸检测实例：(a)标准集合， (b) 表情集合， (c)背景集合， (d) 光照集合 .....	56

图 3-10 侧面汽车检测的训练样本示例 .....	57
图 3-11 侧面汽车检测中为联合直方图匹配选取的 15 个空间模板.....	57
图 3-12 UIUC 测试集A上的ROC曲线.....	58
图 3-13 UIUC 测试集B上的ROC曲线.....	59
图 3-14 UIUC 测试集A上的RPC曲线 .....	59
图 3-15 UIUC 测试集B上的RPC曲线.....	60
图 3-16 侧面汽车检测实例: (a)测试集A, (b) 测试集B, (c)一些其它数字图像 .....	62
图 3-17 侧面汽车检测结果实例: 漏检的情况(a,b)和误检的情况(c,d) .	63
图 3-18 文字检测训练样本示例 .....	63
图 3-19 文字检测中为联合直方图匹配选取的 17 个空间模板 .....	64
图 3-20 文字检测算法在MSRA测试集合上的ROC曲线.....	65
图 3-21 文字检测算法在MSRA测试集合上的RPC曲线 .....	65
图 3-22 文字检测的实验结果: (a)MSRA测试集和(b)一些其它图像 ....	67
图 4-1 人脸的多姿态: (a)正面 (b)上下旋转 (c)左右旋转 (d)平面旋转 .....	70
图 4-2 信息传输系统的简化模型 .....	72
图 4-3 多姿态人脸检测的系统框架 .....	76
图 4-4 基于纠错码分类器的多姿态人脸检测.....	77
图 4-5 构建基于空间直方图特征的ECOC分类器的流程 .....	80
图 4-6 人脸图像训练样本示例: (a)正面人脸, (b)左侧人脸, (c)右侧人脸 .....	81
图 4-7 多姿态人脸检测实验的ROC曲线: CMU_PROFILE测试集 .....	83
图 4-8 多姿态人脸检测的实例: CMU_PROFILE测试集.....	84
图 4-9 多姿态人脸检测的实例: 其它图像 .....	85
图 4-10 不同编码方式的多姿态人脸检测结果示例: (a) one-against-others, (b) ECOC .....	87
图 4-11 多姿态人脸检测算法在CMU_PROFILE测试集上的ROC曲线比较: ECOC vs. One-against-others .....	88
表 1-1 不同特征描述方法比较 .....	7
表 2-1 彩色图像中人脸检测的实验结果 .....	33
表 3-1 正面人脸检测算法不同阶段在CMU+MIT集合上的测试结果 ....	53

表 3-2 不同方法在CMU+MIT测试集的人脸检测率的对比.....	53
表 3-3 用于正面人脸检测测试的CAS_PEAL_R1 子集的图像情况 .....	55
表 3-4 CAS_PEAL_R1 子集上的正面人脸检测测试结果.....	55
表 3-5 侧面汽车检测算法不同阶段在UIUC 测试集A上的测试结果 ....	60
表 3-6 汽车检测方法的性能比较 .....	61
表 3-7 不同方法在UIUC测试集A上等错误率的对比情况.....	61
表 3-8 文字检测算法不同阶段在MSRA测试集上的测试结果 .....	64
表 3-9 视频文字检测算法在MSRA集合上的测试结果对比 .....	66
表 3-10 物体检测系统的执行速度 .....	66
表 4-1 纠错码的一个例子 .....	73
表 4-2 多姿态人脸检测ECOC编码的码本 .....	81
表 4-3 单分类器的分类错误率 .....	82
表 4-4 基于纠错码的多姿态人脸检测器的分类性能 .....	83
表 4-5 不同方法在CMU_PROFILE测试集的人脸检测率 .....	86
表 4-6 多姿态人脸检测的一对多编码 .....	86

## List of Figures and Tables

Figure 1-1 One example of object detection.....	2
Figure 1-2 Framework of object detection based on local texture features	13
Figure 2-1 Neighborhood for LBP computation .....	18
Figure 2-2 An example of LBP computation .....	19
Figure 2-3 Examples of face and non-face image samples :(a) face images, their LBP images and histograms, (b) non-face images, their LBP images and histograms .....	19
Figure 2-4 Examples of side-view car and non-car image samples .....	20
Figure 2-5 Texture and spatial distribution is encoded by spatial histograms in image scale space .....	22
Figure 2-6 Illustration of spatial template .....	22
Figure 2-7 Framework for extraction of spatial histogram features .....	23
Figure 2-8 Feature distribution: side-view car and non-car .....	25
Figure 2-9 Feature distribution: face and non-face.....	26
Figure 2-10 Feature distribution: text and non-text .....	27
Figure 2-11 The 23 spatial templates of face pattern.....	30
Figure 2-12 Histogram matching method for face detection .....	31
Figure 2-13 Spatial templates for histogram matching in face detection :(a) entire face region, (b) eye region, and (c) middle vertical region.....	31
Figure 2-14 Flowchart of two-step histogram matching for face detection	32
Figure 2-15 Examples of face detection in color images.....	33
Figure 2-16 Examples face detection under various lighting conditions ....	34
Figure 3-1 Process of object detection using spatial histogram features ....	38
Figure 3-2 Simplification of maximum likelihood decision rule: form (a), and form (b).....	40
Figure3-3 Training method of cascade histogram matching.....	43
Figure 3-4 Feature selection based on class separability and feature correlation.....	47
Figure 3-5 Some training examples of frontal face .....	51
Figure 3-6 11 Selected spatial templates for cascade histogram matching in	

frontal face detection .....	51
Figure 3-7 ROC obtained on CMU+MIT face detection test set .....	52
Figure 3-8 Some examples of frontal face detection .....	54
Figure 3-9 Some examples of frontal face detection in CAS_PEAL_R1 subsets: (a) Normal subset, (b) Expression subset, (c) Background subset, (d) Lighting subset.....	56
Figure 3-10 Some training examples for side-view car detection.....	57
Figure 3-11 15 Selected spatial templates for histogram matching in side- view car detection .....	57
Figure 3-12 ROC obtained on UIUC car detection test set A .....	58
Figure 3-13 ROC obtained on UIUC car detection test set B .....	59
Figure 3-14 RPC obtained on UIUC car detection test set A.....	59
Figure 3-15 RPC obtained on UIUC car detection test set B.....	60
Figure 3-16 Car detection examples on (a) UIUC test set A, (b) UIUC test set B, and (c) some other digital photos.....	62
Figure 3-17 Car detection results: missing detections (a, b) and false detections(c, d) .....	63
Figure 3-18 Some training examples for text detection.....	63
Figure 3-19 17 selected spatial templates for cascade histogram matching in text detection .....	64
Figure 3-20 ROC obtained on MSRA text detection test set .....	65
Figure 3-21 RPC obtained on MSRA text detection test set.....	65
Figure 3-22 Some examples of text detection: (a) MSRA testing set, and (b) some other video frames .....	67
Figure 4-1 Multi-view faces: (a) front view (b) pitch view (c) profile view (d) in-plane rotation .....	70
Figure 4-2 Model of information transmission.....	72
Figure 4-3 System overview of multi-view face detection .....	76
Figure 4-4 Architecture of the ECOC-based multi-view face detection .....	77
Figure 4-5 Flowchart of constructing spatial histogram features based ECOC multi-class object detector.....	80
Figure 4-6 Some examples of face samples: (a) front view, (b) left profile, and (c) right profile.....	81

Figure 4-7 ROC of multi-view face detection on CMU_PROFILE set .....	83
Figure 4-8 Examples of multi-view face detection on CMU_PROFILE set	84
Figure 4-9 Examples of multi-view face detection on other images.....	85
Figure 4-10 Examples of results with different codes for face detection: (a) one-against-others, and (b) ECOC .....	87
Figure 4-11 Comparison of RCO curve of multi-view face detection on CMU_PROFILE set: ECOC vs. One-against-others .....	88
Table 1-1 Comparison of different methods for representation features.....	7
Table 2-1 Testing results of face detection in color images .....	33
Table 3-1 Results on CMU+MIT set by different phases of frontal face detection .....	53
Table 3-2 Comparison of different frontal face detection methods: detection rates on CMU+MIT set .....	53
Table 3-3 CAS_PEAL_R1 subsets used for frontal face detection testing .	55
Table 3-4 Face detection testing results in CAS_PEAL_R1 subsets .....	55
Table 3-5 Results in UIUC set A by different phases of side-view car detection .....	60
Table 3-6 Car detection results comparison.....	61
Table 3-7 Comparison of RPC equal error rates on UIUC set A .....	61
Table 3-8 Results in MSRA text set by different phases of text detection..	64
Table 3-9 Comparison of text detection results on MSRA test set.....	66
Table 3-10 Performance time of our object detection systems .....	66
Table 4-1 An example of error correct code .....	73
Table 4-2 ECOC codebook of multi-view face detection .....	81
Table 4-3 Classification error rates of the basic classifiers .....	82
Table 4-4 Classification performance of the ECOC_based multi-view face detector .....	83
Table 4-5 Comparison of multi-view face detection rates of different methods.....	86
Table 4-6 One-against-others codebook of multi-view face detection.....	86



## 第1章 绪论

### 1.1 课题背景及研究意义

俗话说：“百闻不如一见”，人类所感知的外界信息 80% 以上来自于视觉。人类似乎天生具有这样一种能力，就是可以从纷繁复杂的客观世界中，迅速而准确地找到他(她)所感兴趣的物体，例如街道、人脸、行人、汽车、文字、建筑物等，以及其它日常生活中所见的各种物体。在视觉感知领域，这种能力被概括成为物体检测。物体检测是人类视觉中的基本步骤和基本功能，为人类了解周围的环境和景物提供了至关重要的前提。

让计算机或者机器人具有包括物体检测在内的视觉感知能力，是人类多年以来的一个科学梦想。为了达到这个目标，有两种途径。第一种途径是仿生学方法，科学家们从生理学、心理学、神经学等各方面对人类视觉系统进行了观察和探索，并且对视觉的物理和化学性质以及人类视觉系统的组织已经积累了相当多颇有价值的知识，但是视觉过程的机理至今还是一个谜。第二种途径是工程方法，主要思路就是利用计算机来模拟人类的视觉功能。科研工作者用各种计算理论和工程方法在该领域内进行了大量的研究，取得了很大进展，并一直在不懈的努力。沿着这个方向进行研究和开发所取得的进展(包括各种计算机视觉技术的原理、算法和系统等)，促进了物体检测领域和计算机视觉学科的发展。

本文的研究内容属于工程方法的研究范畴。围绕物体检测展开研究，本论文的主要研究目的是探索一种有效的物体特征表示方法并且应用于物体检测，从而提供物体检测的新方法。

计算机视觉中，物体检测是一个充满挑战的研究热点。给定一幅输入图像以及感兴趣的目标物体，物体检测的过程就是在图像中检测目标物体的存在，并且定位出其位置。

从认知角度看，物体检测属于语义层次上的分割问题。图 1-1 是物体检测的示例。一幅图像中含有多种类别的物体，例如行人、汽车、文字等。物体检测的任务是在图像中定位目标物体，如果图像中含有目标物体，则获取物体的类别、位置、大小、姿态等信息。图像中物体的各种信息，可以为场景分析和图像内容描述等任务提供有价值的线索。在物体检测的基础上，图

像内容可以被组织成由图像中所含有的物体构成，通过分析各个物体之间的关系，能够对图像内容产生基于物体的高层语义解释。

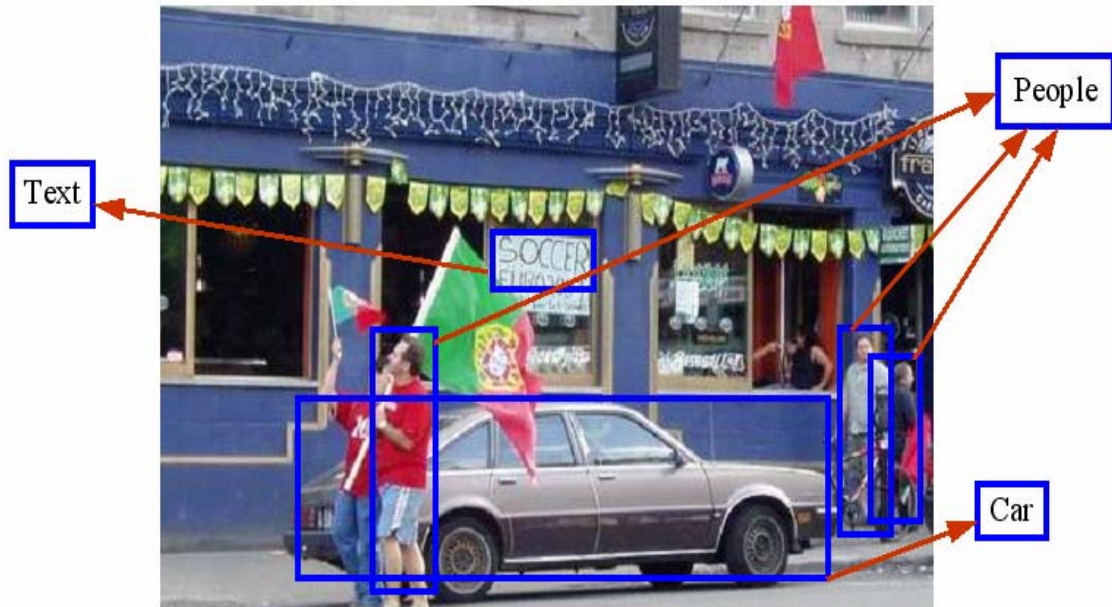


图 1-1 物体检测的示例

Figure 1-1 One example of object detection

目前，物体检测技术受到来自学术界和工业界越来越多的关注，究其原因，是这项技术在理论和应用层面上都获得了多方面的发展动力。

在理论上讲，物体检测是一个极具挑战性的人工智能研究课题，尚未得到彻底而理想的解决方案。物体检测技术用计算机模拟人类的视觉感知过程，是对计算机计算能力和知识表示能力的检验。物体检测是一项跨领域的研究工作，它涉及到图像处理、模式识别和机器学习等多个学科分支，其研究成果影响着这些领域的发展。作为计算机视觉领域中的一个典型问题，物体检测的研究成果对于进一步认识和探索人类视觉系统的工作原理，具有相当大的参考价值。

从应用角度看，物体检测技术具有很高的现实价值和潜在价值。将图像内容解析成为基于物体的表示，在众多的的实际应用中有着广泛的用途，例如多媒体信息分析与搜索、场景分析与理解、视频监控、模式识别等。

在多媒体内容分析领域，物体检测技术可以增强对多媒体数据进行描述和检索的功能。其基本途径是应用物体检测技术获得图像和视频中的语义层

次上物体的信息，形成有效的内容描述，将这些描述和文本信息、音频信息相结合，用来有效的组织、表达、存储、管理、查询和检索多媒体数据。

在图像理解领域，物体检测技术为场景分析与理解的任务提供了前提条件，即快速、有效的物体定位。根据物体的属性和物体之间的空间关系所构成的场景描述可以应用到机器人的视觉系统中，使得机器人系统能够适应周围的环境。

在视频监控领域，物体检测技术可以发现被监控的对象，同时自动监控对象的行为。利用物体检测技术，还可以在大规模存储的监控录像数据中建立基于内容的视频检索机制，能够提供快速的检索和浏览功能。

在特定任务的模式识别领域，物体检测技术为识别算法提供了输入，形成了识别对象。例如，人脸检测模块是自动人脸识别系统中的不可或缺的重要组成部分，是人脸识别系统中的最前端处理过程。

此外，物体检测还是其它领域中的一项关键技术，例如视频会议、视频编码、多媒体通信和三维游戏等。

综上所述，物体检测技术的研究具有重要的理论意义，而且具有很高的应用价值。

## **1.2 物体检测领域的研究现状**

物体检测的困难主要是同类物体之间具有很大的灰度差异，其原因是物体具有不同的姿态，而且成像光照和成像背景复杂多变。因此，物体检测的关键就是如何有效的提取共性特征来描述物体类别的模式，然后利用这些特征进行目标分类。

本文将物体检测领域中的研究重点划分为两个方面。首先是物体特征表示方法，所要解决的问题是如何在图像中提取特征，进而有效的表示物体。其次是物体检测方法，所要解决的问题是如何利用特征表示来判断物体的存在。按照这种划分，本文分别总结物体特征表示方法和物体检测方法的研究现状。人脸检测作为一个典型问题，在物体检测领域有着重要的地位，因此本文还单独介绍人脸检测方法的研究现状。

### **1.2.1 物体特征表示**

人们提出了很多形式的特征来表示物体，并且广泛应用于物体检测及识别领域中。归纳起来，大致有如下几种类型的特征：图像像素特征、边缘特

征、频域特征、局部区域描述特征和直方图特征等。

#### 1.2.1.1 图像像素特征

物体检测中常用的是空域图像上的基本特征，即像素值特征，包括灰度值和彩色信息值。

灰度值对应客观景物被观察到的亮度，只包含灰度信息的图像称为灰度图像。利用灰度值作特征的最简单方法是把图像像素值按行展开构成一个高维向量。采用这种方式，Rowley 等[1]和 Garcia 等 [2]进行了人脸检测，Kim 等[3]进行了文字检测。更复杂的一种处理方法是把高维向量进行主成分分析(PCA: Principle Component Analysis)，保留能量集中的部分，去除特征之间的相关性，得到低维特征在最大程度上来描述原始特征。这种方法称为像素 PCA 方法，首先在人脸识别领域由 Turk 等[4]提出，并且极大的推动了物体检测与识别领域的发展。另一种降维方式是线性判别分析(LDA: Linear Discriminative Analysis)。和主成分分析方法不同，这种方式企图在高维的灰度值向量空间中刻画类别可区分性方向，从而获得低维特征向量来最大程度上来区分物体类别。这种方法称为像素 LDA 方法，它是目前人脸识别领域中的流行方法，取得了比 PCA 方法更好的识别性能[5]。

客观世界中，彩色信息是物体的一个本质属性，因此采用彩色信息作为物体的特征表示方法，在很多领域中得到广泛的应用。本文简单介绍这方面与物体检测有关的工作。Swain 等[6]首先引入彩色直方图作为物体描述特征，随后 Huang 等人[7]把空间相关性概念嵌入彩色直方图特征，提出了彩色相关图特征(Color correlogram)。这两种方法在图像检索、物体跟踪[8]等领域取得了明显的效果。在物体识别领域中，Schiele[9]等介绍了一个彩色不变性模型，并用于描述物体的局部信息，通过高维直方图匹配的方法区分物体。有很多人脸检测方法也使用了彩色信息特征，例如，Hsu 等[10]在  $YC_bC_r$  彩色空间刻画人脸模式的特征，在解决彩色图像中的正面人脸检测问题的时候取得了较好效果。

图像像素特征具有计算简单的特点，可以直接从原始数据中获取，但是容易受光照、背景变化和噪声的影响。

#### 1.2.1.2 边缘特征

物体的几何或物理性质的突变，例如深度、反射或表面方向的不连续性等，总是以图像中灰度突变的形式出现，这些灰度突变构成了图像的边缘和

边线。不同类别的物体所具有的边缘分布是不同的，因此很多研究者使用边缘特征来描述物体。

在图像中提取边缘特征的一般过程是首先通过卷积方法把原始灰度图像转换成为边缘图像，然后利用边缘信息构成特征。边缘梯度是边缘信息的一种度量。Felzenszwalb [11]在行人检测算法中使用 Canny 边缘图对行人的形状进行建模，在输入灰度图像中提取边缘后，通过形状匹配方法检测行人。边缘梯度是文字检测方法中常用的特征。Smith 等[12,13]在他们提出的算法中，把文字块定义为包含大量剧烈边缘的图像区域。Cai 等[14,15]提出了在 YUV 彩色空间通过边缘检测方法定位文字区域的算法。边缘方向是边缘信息的另一种度量。Bernhard 等[16]采用基于边缘方向作为特征，实现了一个实时的人脸检测系统。

### 1.2.1.3 频域特征

频域分析是图像处理的一种有效工具。比空域特征和边缘特征更为复杂的特征提取方法是把图像变换到频率域中，采用变换系数作为特征。在物体检测领域，通过将图像变换到频域上，可以根据频率特性，或者提取物体的特征点区域，或者获取整个物体的频域描述特征量。M.Zobel 等人在 JPEG 图像的离散余弦变换（DCT）系数中提取出额头、眼睛、鼻子、嘴唇、下巴等特征区域，通过人脸结构模型检测人脸[17]。

小波变换是频域分析中常用的办法，它可以同时在空域和频域两个尺度上对图像进行分解，得到一系列变换系数[18]。Papageorgiou 等使用图像的 Harr 小波系数作为特征矢量，搜集了大量正反例样本，通过 SVM 机制实现人脸检测[19]。Christophe Garcia 和 Georgios Tziritias 对人脸图像小波变换系数做了统计分析，刻画人脸图像纹理的特征，然后通过基于概率分布的 Bhattacharyya 距离进行分类得到人脸和非人脸的判别规则[20]。Waring 等[21]把 Gabor 直方图作为特征，通过 SVM 机制实现人脸检测。

频域特征可以有效的表示物体，但是其计算量大。虽然 Harr 小波系数的简化形式矩形特征[21]具有快速计算的特点，但是往往需要很长的训练时间才能够选择出几千个用于物体检测的特征。

### 1.2.1.4 局部区域描述特征

局部区域描述特征的基本思想是首先从物体模式上找到对于各种变化(包括尺度和旋转等方面的变化)相对稳定的局部区域，然后在这些局部区域

上提取特征来描述物体。

Lowe[22]提出了 SIFT (Scale invariant feature transform)描述特征。一个局部图像被均匀划分成  $4 \times 4$  个子区域, 边缘方向被量化成为 8 份。根据这种划分形成一个 128 维的直方图。图像上的每个点都计算出相应的边缘方向, 累加到直方图对应的项上。局部二值模式(LBP: Local binary pattern)是一种可以刻画物体纹理的表示特征。Hadid 等[23]采用 LBP 对灰度图像进行预处理, 得到 LBP 图像, 然后使用高维直方图提取出图像的纹理分布作为分类特征。这种方法在纹理分类、图像分析等方面得到了应用。

局部区域描述特征方法综合考虑了物体的纹理信息和结构信息, 但是最后得到的分类特征维数过高。

#### 1.2.1.5 直方图特征

对于任意一个统计量, 直方图提供了它的分布情况, 也可以说给出了统计量的整体描述。一般而言, 图像直方图可以对任意特征进行统计, 包括上述的图像像素值特征、边缘算子特征、频域特征等类型的特征。除此之外, 还有扩展类型的直方图, 例如多分辨率直方图(Multiresolution Histogram)和空间直方图(Spatial Histogram)。

多分辨率图像是指在原始图像上采用高斯滤波函数进行卷积而得到的图像。在多分辨率图像上统计获得的直方图, 称为多分辨率直方图。Hadjidemetriou 等[24,25]运用多分辨率直方图进行纹理分类, 取得了很好的结果。空间直方图是在把图像划分成子区域的基础上获得的子图像的直方图。利用空间直方图把图像纹理和空间结构结合起来构成的特征, 对基于内容的图像检索具有明显的效果[26,27]。

直方图特征的优势在于可以综合描述物体在纹理、尺度和空间结构等方面的信息, 但是最后所得到的特征向量维数过高。

#### 1.2.1.6 物体特征的性能比较

这里, 本文总结上述特征描述方法, 并在表 1-1 给出了一个简单的性能比较。本文比较不涉及详细的技术细节, 只是给出它们的优缺点。这几种方法从不同的角度提取物体的特征。值得指出的是, 上述几种方法之间并非完全彼此独立, 它们之间具有相关联的部分。针对不同应用的需求, 可以选用不同特征。

表 1-1 不同特征描述方法比较

Table 1-1 Comparison of different methods for representation features

特征类型	技术特点	优点	缺点
图像像素值特征	包括灰度值和彩色信息值	可以直接从原始数据中获取	受光照、背景变化和噪声的影响
边缘特征	包括边缘梯度和边缘方向	刻画了物体区域的突变，计算简单	容易受噪声的影响
频域特征	选取频域中的变换系数作为特征	可以有效的表示物体	计算量大
局部区域描述特征	在相对稳定的局部区域上提取特征	综合考虑了物体的纹理信息和结构信息	特征维数过高
直方图特征	获取图像特征的分布情况	描述物体在纹理、尺度和空间结构的信息	特征向量的维数过高

## 1.2.2 物体检测方法

按照目标物体的类别数目，物体检测可以被划分为两种任务。第一种是单类目标物体检测任务，即检测目标为单一类型物体。第二种是多类目标物体检测任务，需要在图像中同时把多种类型的目标物体提取出来。

物体检测问题可以用机器学习的方法解决。图像样本由一个特征集合表示，采用机器学习的方法对样本进行分类，判断其是否属于目标类。一般来说，物体检测方法可以划分为两大类：基于全局表观特征的方法和基于局部特征的方法。

### 1.2.2.1 基于全局表观特征的方法

基于全局表观特征的方法将物体看作整体来提取特征，采用分类器算法区分物体的类别。Roweley 等[1]和 Carcia 等[2]采用人工神经网络进行人脸检测算法研究。基于小波系数特征，Osuna 等[28]和 Papageprgiou 等 [29]训练支持向量机算法来检测人脸和汽车。Schneiderman 等[30]用弱贝叶斯规则检测多姿态的人脸和汽车。最近，Boosting 算法广泛用于物体检测。Viola 等[21,31,32]首先提出应用 Adaboost 算法检测人脸，Li 等[33]将这个方法推广到多姿态人脸检测中去。Chen 等[34]在文字检测中使用了 Adaboost 算法提高系统性能。其他研究者采用了另外的学习算法描述和检测物体，包括概

率分布 [35,36], 主成分分析 [37]和混合线性子空间方法[38]等。

#### 1.2.2.2 基于局部特征的方法

基于局部特征的方法将物体看作是很多部件的有机组合, 根据部件之间的关系来判断物体的类别。

Mohan 等[39]构建了行人的部件模型, 通过支持向量机检测部件和行人, 他们的方法把行人模型分解为头部、上肢、下肢和躯干等部件。Naquest 等[40]以图像分块的子区域作为特征, 使用线性分类器判断物体类别。Agarwal 等[41]使用特征点提取方法构建了侧面汽车的部件集合, 把部件之间的位置关系构成特征, 采用神经网络来检测侧面汽车。Fergus 等[42]和 Leibe 等[43,44]在使用特征点提取方法获得物体的组成部件后, 用概率分布表示部件之间的关系, 并通过概率规则进行多种物体的识别, 包括摩托车、人脸、飞机和汽车等。

#### 1.2.2.3 多类目标物体检测

上述方法大多用于单类目标物体检测任务, 多类目标物体检测的研究近年来越来越引起人们的关注。Schneiderman 等[30]把物体的三维姿态分解成为互相不重叠的姿态的集合, 针对每个姿态训练单独的检测器, 通过对单检测器结果进行融合的方式最终确定物体是否存在, 以及属于哪种姿态。这个方法应用于多姿态的人脸检测和汽车检测, 取得了很好的效果。Lin 等[45]提出一种多类分类器和 Adaboost 相结合的办法, 直接用于多个类别物体的检测。他们以不同姿态、不同光照、以及有部分遮挡的人脸检测为例, 介绍了算法的实现过程。Amit 等[46]提出了一种由粗到精的多类别形状检测的算法, 并且应用于汽车牌照中的文字检测问题。要区分的物体类别包含 26 个英文字母、10 个阿拉伯数字和 1 个特殊符号, 总共有 37 个类别。Fei-fei Li 等[47,48,49]使用贝叶斯框架, 提出了一种可增量学习的物体建模和检测方法。这个方法强调在每类物体只有不超过 10 个样本的情况下, 如何不断学习新增加的物体类别的几何模型。目前, 此方法扩展到了 101 种物体的检测。为了提高物体建模和物体检测器的效率, Torralba 等[50]提出了一种特征共享算法, 在不同物体类别之间选择共同的特征, 用于不同类别的物体检测器。

将图像分割和物体检测与识别结合起来, 是物体检测发展的新方向。Tu 等 [51]提出了一个图像内容分解的框架进行场景理解, 这个框架把图像



分割和物体检测与识别通过马尔可夫链蒙特卡洛(MCMC:Markove Chain Monte Carlo)算法统一处理, 最终得到图像中的不同物体和背景部分。

### 1.2.3 物体检测中的典型案例: 人脸检测

许多国家展开了人脸检测的研究, 主要是美国、欧洲国家、日本、新加坡、韩国等, 著名的研究机构有美国 MIT 的 Media lab 和 AI lab、CMU 的 Robotics Institute 和 Human-Computer Interface Institute、Microsoft Research、英国的 Department of Engineering in University of Cambridge 等。文献[52]对人脸检测研究工作进行了总结。

#### 1.2.3.1 基于器官特征的方法

这种方法首先提取人脸器官图像特征, 然后根据人脸中各器官的几何关系来确认人脸的存在。文献[53]利用了大量的特征, 包括几何信息、空间结构、灰度等各种度量, 对人脸进行粗定位。人脸的几何关系由一个树型结构表示, 并计算出概率属性, 作为判断是否为人脸的条件。这是一种自底向上的方法。另外还有自顶向下的方法, 即根据一个人脸模型(一般是正面人脸模型)先在一个比较大的范围内寻找人脸候选区, 由粗到精地在一个最佳范围内定位人脸候选区, 然后检测各种人脸器官特征。Morimoto 等在一个鲁棒性强的瞳孔定位技术的基础上, 结合人脸结构信息, 实现了图像中多人脸的检测[54]。

#### 1.2.3.2 示例学习

示例学习的基本思想是从某一概念的已给正例和反例的集合中归纳产生出接受所有正例, 同时排斥所有反例的该概念的一般规则。将人脸检测视为从模式样本中区分非人脸样本和人脸样本的模式识别问题, 通过对人脸样本集和非人脸样本集进行学习以得到分类算法。为了获得较高的精度, 学习过程需要大量的样本。另外, 样本数据本身是高维矢量。因此, 研究通用而有效的学习算法的关键是精确的可区分度和数据维数的降低。

很有意思的是, 几乎所有的统计模式识别方法都应用到了人脸检测这个问题上。这里有两个方面, 一方面是在特征提取方法上, 使用了统计的方法, 比如主成分分析。另一方面是在判别分类的算法上使用了统计的方法, 如支持向量机、隐马尔科夫模型、贝叶斯决策等等。最近 Viola 等[32]使用

了一种同时提取特征并且训练分类器的方法，所采用的是 AdaBoost 算法。

主成分分析是最小方差准则下最优的描述方法，它将输入数据的维数从图像空间的像素数减少为特征空间的维数。其中，特征脸方法得到了广泛应用[4]，就是用一组特征向量的加权线性组合来代表人脸。主成分分析考虑的是样本模式的整体非几何的特征表示，而局部特征分析（LFA, Local Feature Analysis）则从局部的几何特征出发进行数据空间的降维。将它用于人脸检测，分析和实验证明 LFA 是一种有效的目标检测方法[55]。

支持向量机（SVM, Support Vector Machines）是一种二次规划（Quadratic Programming）方法，它在处理大批量的人脸正例和反例时，只选取那些位于分界面附近的数据来完成学习过程[28]。

Schneiderman 等用小波描述人脸模式的局部特征，根据局部特征之间的条件独立性，采用 Naive Bayes 分类器判断人脸[30]。Liu 等通过构造人脸和非人脸的概率高斯分布，采用 Bayes 分类器判断人脸[36]。用 LBP 直方图特征通过 SVM 分类器检测人脸，也是一种新方法[23]。

将多个表示人脸模式的线性空间进行组合，是示例学习的另一条途径。文献[56]采用了 Kohonen 自组织映射网络对人脸样本和非人脸样本聚类，对每一类样本进行 Fisher 线性判别，得到每一类的判别平面，从而构成图像子空间，并运用高斯模型描述每个子空间，估计出类条件概率密度函数；这样，对于测试图像，计算其属于各个子空间的概率，分类决策为概率最大的类是它所属的类，从而判断测试图像是否为人脸。这些方法一般用于正面人脸检测。

### 1.2.3.3 神经网络

从本质上讲，神经网络也是一种基于样本的学习方法。将神经网络用于人脸检测，取得了很大的进展。

MIT 的学者首先对人脸样本集和非人脸样本集聚类，以测试样本与人脸样本集和非人脸样本集的子类之间的距离作为识别特征向量，利用多层感知器(MLP)网络作为分类器[35]。CMU 的研究人员直接以图像作为神经网络的输入，设计了一个具有独特结构的适用于人脸特征的神经网络分类器，并且通过前馈神经网络对检测结果优化[1]。他们工作的共同之处是都采用了自调整（bootstrap）的学习原理，对分类器一边训练，一边测试，并把在测试过程中的错误分类结果作为反例样本加入学习过程，从而减少了样本集的规模，并逐步的提高了神经网络的分类性能。

Feraud 等[57]利用多个神经网络：多层感知器(MLP)和约束产生式模型(CGM ,Constrained Generative Model)，实现了一个可应用于 Web 中人脸图像检索的快速而准确的人脸检测方法。Lin 等[58]训练了三个基于概率决策的神经网络(PDBNN ,Probabilistic Decision Based Neural Network),用于人脸检测、眼睛定位和人脸识别，实现了一个完整的人脸识别系统。

#### 1.2.3.4 基于隐马尔可夫模型的方法

马尔可夫模型的概念是一个离散时域有限状态自动机，隐马尔可夫模型 HMM 是指这一马尔可夫模型的内部状态外界不可见，外界只能看到各个时刻的输出值。HMM 的打分、解码和训练相应的算法是前向算法、Viterbi 算法和前向后向算法[59]。

对于人脸模式来说，我们可以把它分成前额、眼睛、鼻子、嘴巴和下巴这样一个序列。那么人脸模式就可以通过对这些区域的有序识别来检测。这正好是隐马尔可夫模型容易做到的。Samaria[60]等人提出了使用 HMM 模型进行人脸检测的算法。他们使用人脸区域的结构信息来作为隐马尔可夫模型的状态迁移的条件。

#### 1.2.3.5 基于 AdaBoost 的方法

Boosting 是一种分类器融合算法。Adaboost 学习算法原本是用来提高某种简单分类算法的性能的，例如，可以用来提高简单的感知器的性能。它通过对一些弱的分类器的组合来形成一个强的分类器。在 Adaboost 算法中，简单的分类算法被称为弱学习算法。Adaboost 算法通过一个迭代的训练过程来得到一个强的分类器。在第一次训练出一个弱分类器后，训练样本的权重得到调整，从而使没有被第一次训练出的弱分类器正确分类的样本的权重增加。如此迭代下去，最终得到的分类器是对每次训练得到的弱分类器的一个线性组合。Viola 和 Jones[32]提出了一种基于 AdaBoost 和快速特征提取的正面实时人脸检测算法。他们的算法在个人计算机上达到了 15 帧/秒的速度，在 MIT 和 CMU 的测试集上有很好的表现。MSR 的研究者们通过利用 3 个视角的这样的检测器，实现了一个多视角的人脸检测算法[33]。他们还改进了 Adaboost 人脸检测方法[61,62,63]。

#### 1.2.3.6 基于彩色信息的方法

人脸的肤色在颜色空间中的分布相对比较集中，利用这个特点可以检测

人脸。这种方法的最大优点是对姿态变化不敏感。用彩色信息检测人脸的关键是合理选择色度坐标。常用的方案是将彩色的  $R$ 、 $G$ 、 $B$  分量归一化。目前人们研究更多的是如何提取彩色的色度信息，即将  $RGB$  彩色空间转化为其它彩色空间，以突出色度信息。

H.Martin 设计了肤色模型表征人脸颜色，利用一个感光模型对输入图像修正和补偿，并建立人脸颜色分类器，其输出作为神经网络的输入进行检测[64]。Tony 等采用高斯混合模型(GMM, Gaussian Mixture Model)表示人脸肤色  $R$ 、 $G$ 、 $B$  各分量的统计分布，通过阈值比较判断像素是否为人脸像素[65]。Ying Dai 等利用彩色信息，并结合人脸的纹理特征，设计了 SGLD 共发矩阵方法进行检测[66]。Wu 等通过在  $XYZ$  彩色空间中，对人脸肤色和头发颜色的分布建立基于模糊逻辑理论的描述模型，通过计算隶属度来确定人脸肤色区域[67]。Hsu 等[10,68]在  $YC_bC_r$  彩色空间刻画人脸模式的特征，在解决彩色图像上的正面人脸检测问题的时候取得了较好的结果。

#### 1.2.4 物体检测中存在的问题

通过对研究现状的回顾，可以发现，人们在物体检测领域中已经获得了很大成就，但是仍然面临着许多实际问题。存在的主要问题总结为如下几个方面。

1. 首先在物体表示方面，采用何种方式抽取稳定、有效而具有灵活性的特征来表示物体，并且满足物体检测任务的要求，仍然是一个尚未解决的开放问题。
2. 其次在检测方法方面，如何通过选择有效的特征、采用何种检测策略、以及如何解决多姿态或多类目标的物体检测等，都需要在理论上和实践中进行更深入的探索。
3. 其它方面，例如：如何进行更加客观的检测方法性能评价、如何建立大规模的物体数据库等，还需要研究者们长期工作，以期获得进一步的提高。

### 1.3 本文主要工作及结构安排

本文针对物体检测中的主要问题，在物体表示方法和物体检测算法两个层面展开研究，主要研究目的是探索一种有效的物体特征表示方法并且应用于物体检测，从而提供物体检测的新方法。本文的主要研究内容可以概括为

以下几个方面。首先在物体表示方法方面，研究基于局部纹理分析的物体特征提取方法，提出了改进的空间直方图特征；其次通过研究有效的特征选择方法，提出基于局部纹理特征的物体检测方法，并且将该方法应用于三种典型类别的物体检测任务(分别是侧面汽车检测、视频文字检测和正面人脸检测)；最后将这种新的物体检测方法推广至多姿态人脸，提出了一种多姿态人脸检测算法。

本论文从整体考虑，将思路模块化，表示成如图 1-2 所示的研究框架图。论文工作建立在物体的局部纹理分析的基础上，按照特征提取、物体检测方法、多类物体检测的顺序层层推进，最后形成了以空间直方图特征为中心的整体框架。

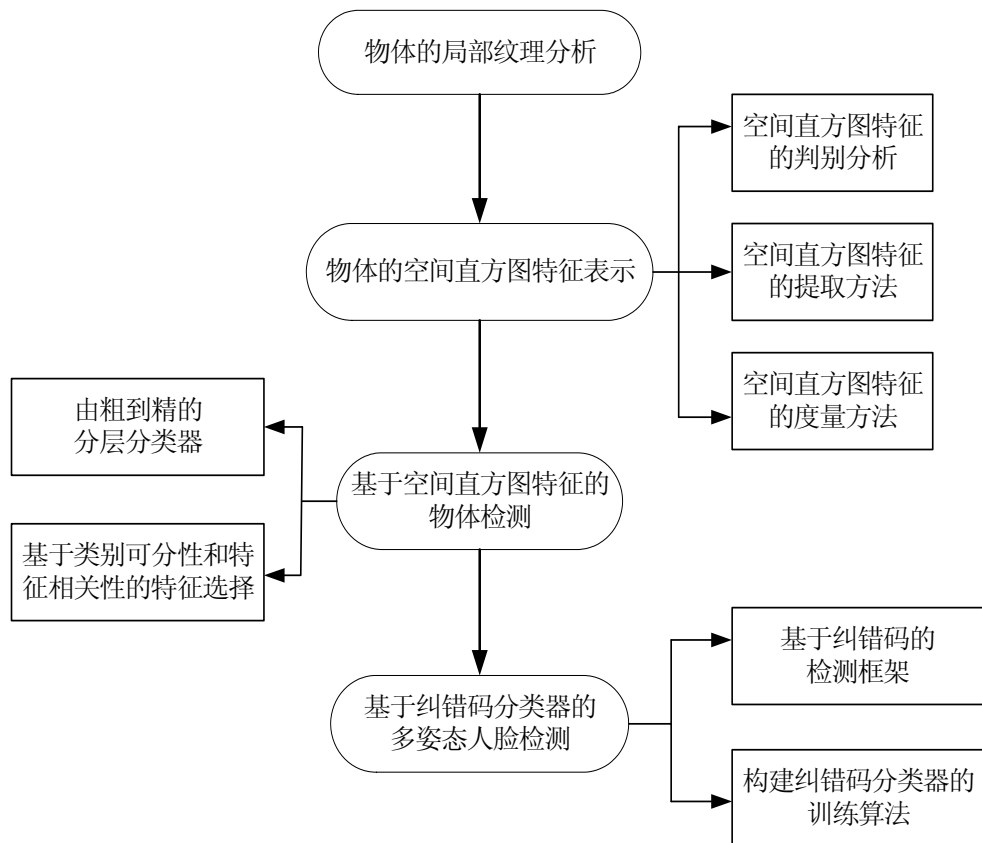


图 1-2 基于局部纹理特征的物体检测方法的研究框架图

Figure 1-2 Framework of object detection based on local texture features

### 1.3.1 论文的主要工作

论文的主要工作概括如下。

1. 在物体表示方法方面，提出了改进的空间直方图特征，用于物体检测任务中的特征表示。空间直方图特征是一个结合纹理分布和空间结构的物体表示方法。以刻画物体的纹理信息和结构信息为着眼点，建立了空间直方图特征的提取方法。该方法采用图像局部区域的空间直方图来表示物体，并且提取与目标物体类相适应的空间直方图特征。本文对空间直方图特征的物体判别能力进行了分析，并且提出了度量空间直方图特征有效性的方法。
2. 在物体检测方法方面，提出了基于空间直方图特征的层次化物体检测算法。该算法采用空间直方图特征作为物体表示，通过由粗到精的策略，将直方图匹配和支持向量机分类器构成分层结构的分类器进行物体检测。
3. 在特征选择方面，提出了基于类别可分性和特征相关性的特征选择方法。为了构建有效的分类特征，本文采用 Fisher 准则函数和互信息熵分别度量空间直方图特征的类别可分性和特征之间的统计相关性，并且提出新的特征选择方法来构造紧凑的空间直方图特征子集，作为分层分类器算法的输入特征。
4. 在多姿态人脸检测方法方面，提出了基于纠错码多类分类器的多姿态人脸检测方法。该方法以纠错码多类分类器为框架，将基于空间直方图特征的物体检测算法推广至多姿态人脸检测。其基本思路是通过纠错码编码方式把多姿态人脸检测问题分解成多个单独的两类分类问题，每个两类分类问题采用基于空间直方图特征的物体检测方法求解。在构建多类分类器的过程中，本文提出了以最小化分类错误率为目标的联合训练算法，逐步选择特征子集和训练分类器。

### 1.3.2 论文的结构安排

本论文的后续章节安排如下。

第 2 章讨论了基于局部纹理分析的物体表示方法。提出了改进的空间直方图特征，并且论述了空间直方图特征的提取方法；然后对空间直方图特征进行了判别分析；最后以彩色图像中的人脸检测为实验示例，验证空间直方

图特征是否具有对物体检测有利的判别能力。

第3章的内容是基于空间直方图特征的物体检测算法。首先提出了物体检测算法的系统框架；然后分别针对直方图匹配和支持向量机分类器，提出了基于类别可分性和特征相关性的特征选择方法；最后给出了所提出的物体检测算法的实验评估结果。

第4章对多姿态人脸检测进行研究，将基于空间直方图特征的物体检测算法扩展到多姿态人脸检测的范围。首先介绍了纠错码多类分类器的基本方法；其次在此基础上，提出了基于纠错码多类分类器的检测方法；然后，提出了以最小化分类错误率为目标的联合训练算法，构建基于空间直方图特征的纠错码多类分类器检测算法；最后对多姿态人脸检测算法进行实验验证，给出了实验结果。

最后是论文的结论部分，对论文工作进行总结，并且探讨了未来的研究方向。

### 1.3.3 论文的主要贡献

论文工作的主要贡献总结如下。

1. 本文基于物体的局部纹理分析，提出了改进的空间直方图特征，用于物体检测任务中的特征表示。空间直方图特征同时刻画了物体的纹理信息和空间结构信息。改进的空间直方图特征与目标物体类别相关联，具有对目标物体的判别能力，而且能够适用于各种典型物体的表示。
2. 提出了基于空间直方图特征的层次化物体检测方法。该方法采用分层分类器对物体类别进行分类，完成图像中物体的检测。该方法是一个通用的物体检测框架，能够运用到各种典型类别的物体检测任务中，例如侧面汽车检测、视频文字检测、正面人脸检测等。
3. 提出了基于类别可分性和特征相关性的特征选择方法。采用 Fisher 准则函数和互信息熵分别度量特征的类别可分性和特征之间的统计相关性，以分类器错误率最小为目标选择特征子集。该方法不但可以选择出有效的分类特征，而且在保证分类性能的条件下降低了特征的维数，提高了分类效率。
4. 提出了基于纠错码多类分类器的多姿态人脸检测方法。该方法结合纠错码多类分类器算法，将基于空间直方图特征的物体检测算法扩

展到多姿态人脸检测的范围。同时，针对基于纠错码多类分类器的训练困难的问题，提出了以最小化分类错误率为目标的联合训练算法。该算法以最小化整体分类错误率为目标，逐步选择特征子集和训练分类器，从而解决了基于纠错码的多类分类器算法中训练困难的问题。



## 第2章 空间直方图特征及其判别分析

### 2.1 引言

物体表示和特征提取是物体检测研究中的基本问题。一般来说，衡量特征的有效性有几个标准。它们分别是：(1)特征是否具有较强的类别可分性；(2)特征之间是否相互独立或者不相关；(3)特征用于分类时的分类错误率是否最小；(4)特征是否具有较低的计算复杂度。概括而言，如果一个特征具有较大的类别可分性，而且同时和其它特征之间具有较小的相关性，那么这个特征对于物体表示和判别是有效的。在一些实时的应用当中，计算复杂度也是一个重要的指标。由于实际的物体检测系统中的物体模式变化往往非常大，不容易找到同时满足这些条件的特征，使得物体表示特征的设计方法成为检测系统中首先面临的难点问题。

本文基于物体的局部纹理分析，提出了改进的空间直方图特征，用于物体检测任务中的特征表示。为了简略起见，本文其余部分直接将改进的空间直方图特征称为空间直方图特征。本章的重点是研究空间直方图特征的提取方法和判别分析，并且通过彩色图像中的人脸检测来考察空间直方图特征的类别可分性。

### 2.2 空间直方图特征的提取

一般而言，每一类物体根据它的特性而具有其特定的纹理模式。例如，人脸包含各种不同的器官，这些器官按照一定方式组合成为人脸。因此人脸图像具有一定的灰度分布、颜色分布、边缘分布以及频率域分布，这些特性构成了人脸图像独特的纹理模式。纹理模式具有两方面的作用。第一是表征功能，可以用来表示和描述物体的组成和分布；第二是判别功能，不同物体具有不同的纹理模式，利用这一点可以区分不同物体。

物体模式不但包含纹理分布，还具有结构信息，这两种信息都能够用于表示和识别物体。为了描述物体模式，本文提出采用空间直方图特征表示物体，基本思想是用直方图来刻画物体的整体和局部的纹理模式。具体而言，主要采用两种策略：第一，采用纹理特征来描述物体，用直方图反映纹理特征的分布；第二，物体具有空间结构关系，在物体图像的各个子区域上提取

直方图特征。

空间直方图特征是一种结合纹理分布和空间结构的物体表示方法。该方法采用图像局部区域的空间直方图来表示物体，并且提取与目标物体类相适应的局部纹理特征。

### 2.2.1 纹理直方图

纹理是图像描述中的一个重要概念。关于图像纹理的精确定义至今尚未作出。一般来说，纹理是指在图像中反复出现的局部模式和它们的排列规则。纹理分析的方法可以归纳成三类：统计模型、结构模型、变换域方法[106]。本文采用的纹理特征是局部二值模式(Local Binary Pattern, 简称 LBP) 特征。LBP 算子是新近提出的一种图像纹理描述方法，它运算简单，已经被成功的应用于图像纹理分类和人脸描述等领域[23,69,70,71]。

LBP 算子有很多类型，本文只使用了基本类型 LBP 算子。如图 2-1 所示，将邻域像素的灰度值表示为  $g_i (1 \leq i \leq 8)$ 。对于图像中的任意像素，根据其八邻域的像素灰度值来计算其 LBP 值。

$g_1$	$g_2$	$g_3$
$g_4$	$g_0$	$g_5$
$g_6$	$g_7$	$g_8$

图 2-1 LBP 计算中的邻域像素

Figure 2-1 Neighborhood for LBP computation

邻域像素根据中心像素灰度值进行比较，获得一个二值数值，计算公式如下：

$$s(g_0, g_i) = \begin{cases} 1, & g_i \geq g_0 \\ 0, & g_i < g_0 \end{cases} \quad (1 \leq i \leq 8) \quad (2-1)$$

将 8 个邻域像素的二值数值，依照顺序编码求出中心像素的 LBP 值，计算方法见式(2-2)：

$$LBP = \sum_{i=1}^8 s(g_0, g_i) 2^{i-1} \quad (2-2)$$

一个计算 LBP 的实例如图 2-2 所示。

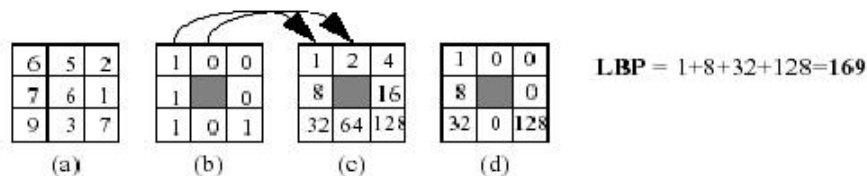


图 2-2 计算 LBP 的实例

Figure 2-2 An example of LBP computation

直方图反映了物体的纹理分布，我们采用直方图来表示物体。对于任意图像，计算基于直方图的模式表示步骤如下：

- 首先采用方差归一化方法处理灰度图像以消除光照影响；
- 然后使用基本类型 LBP 算子把图像转换成 LBP 图像；
- 最后求出 LBP 图像的直方图。

图 2-3 显示了几幅人脸图像和非人脸图像、它们的 LBP 图像以及 LBP 直方图。图 2-4 显示了几幅侧面汽车图像和非汽车图像、它们的 LBP 图像以及 LBP 直方图。

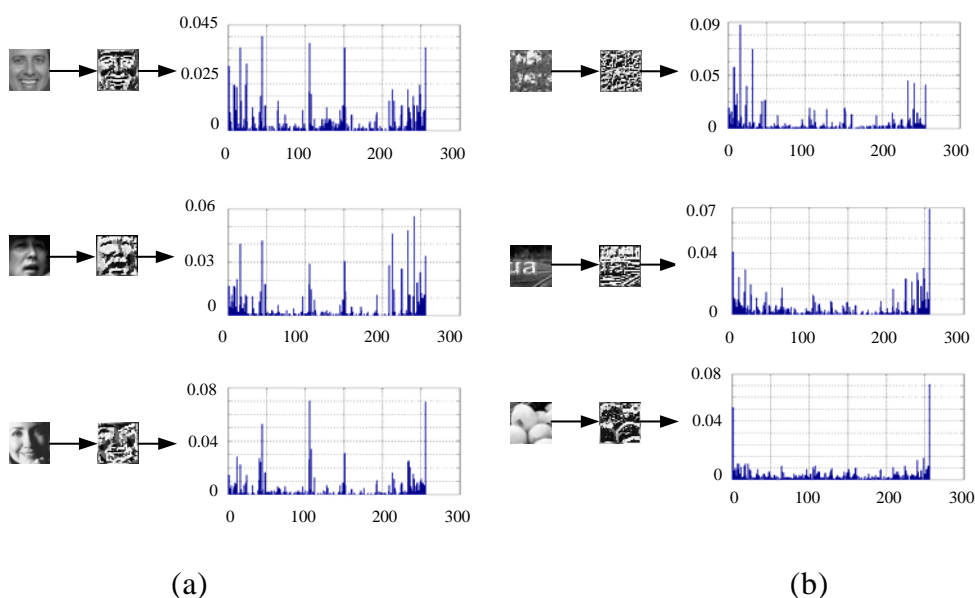


图 2-3 人脸(a)和非人脸(b)的图像、LBP 图像和 LBP 直方图

Figure 2-3 Examples of face and non-face image samples : (a) face images, their LBP images and histograms, (b) non-face images, their LBP images and histograms

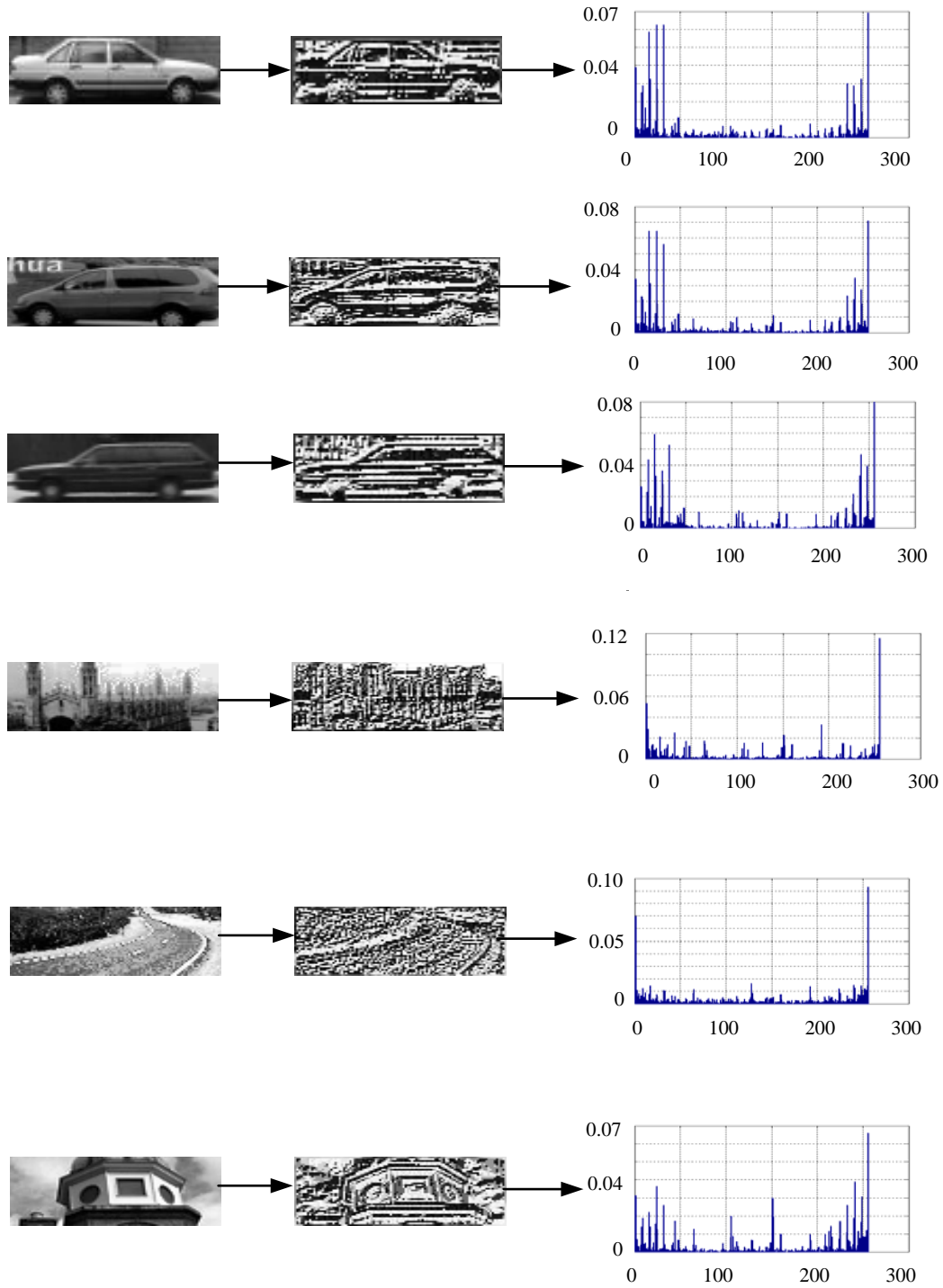


图 2-4 侧面汽车（前 3 行）和非汽车（后 3 行）的图像、LBP 图像和 LBP 直方图

Figure 2-4 Examples of side-view car and non-car image samples

从直观上分析, 基于直方图的纹理特征能够刻画物体类别的共性, 同时也可以用于区分物体类别。例如从图 2-3 和图 2-4 中可以看出, 同一类别的物体具有整体分布相似的纹理直方图, 而且与反例物体样本的纹理直方图之间存在比较明显的差异。因此, 纹理直方图具有一定的物体判别能力。这个性质在图像检索和物体识别领域中得到了验证。

纹理直方图提供了物体图像的纹理分布情况, 给出了一幅图像所有纹理特征的整体描述, 在图像分析中得到了广泛的应用。但是, 纹理直方图的表征能力满足不了物体检测的要求, 其根本原因是直方图忽视了物体的空间结构信息, 没有对物体的空间分布进行编码。对于一些非目标物体图像和物体图像, 它们的纹理直方图非常相似, 甚至具有相同的分布。这些现象说明了纹理直方图对物体类别的区分能力远远不能达到物体检测的精确要求。

传统直方图的缺陷是它没有考虑物体的局部结构关系。物体是由有不同的结构按照一定规则组成的纹理模式, 如果只采用传统直方图的话, 这种结构关系就会丢失。为了克服这个缺陷, 并且增强纹理直方图表示方法对物体的可区分性能, 本文引入空间直方图的概念。

定义一个物体模式为固定尺寸的灰度窗口图像, 空间直方图使用空间模板来衡量物体的局部纹理信息。如图 2-5 所示, 在不同尺度的图像上提取出灰度窗口图像, 经过 LBP 基本算子预处理后形成物体模式图像, 简化表示为  $IW$ 。空间模板是一个矩形区域(如物体模式窗口中的阴影区所示), 表示为  $rt(x, y, w, h)$ , 模板的左上角位置是  $(x, y)$ , 尺寸是  $(w, h)$ 。

空间模板的含义如图 2-6 所示。给定物体模式图像  $IW$  和空间模板  $rt(x, y, w, h)$ , 提取出模板所覆盖子图像的直方图。这样得到的直方图, 称之为空间直方图(spatial histogram), 记作  $SH^{rt(x, y, w, h)}(IW)$ 。

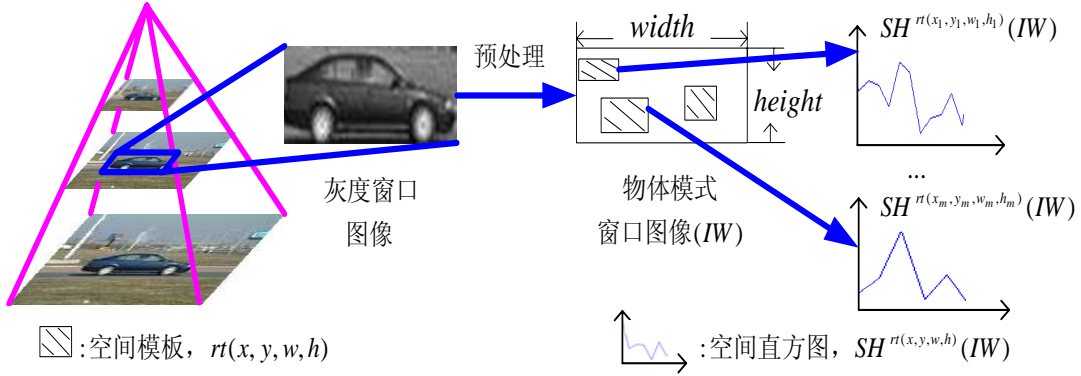


Figure 2-5 Texture and spatial distribution is encoded by spatial histograms in image scale space

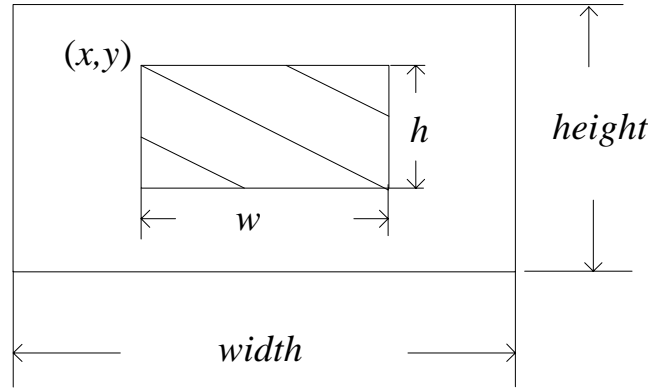


Figure 2-6 Illustration of spatial template

有很多方法来度量两个直方图之间的相似性，例如直方图二次距离 (Histogram quadratic distance)， $\chi^2$  统计量 ( $\chi^2$ -statistics) 和直方图的交 (Histogram intersection)[9]。基于稳定性和计算复杂性考虑，本文采用直方图的交。假设  $H_1$  和  $H_2$  是两个  $k$  维直方图，直方图的交按照公式(2-3)计算：

$$D(H_1, H_2) = \frac{\sum_{i=1}^k \min(H_1^i, H_2^i)}{\sum_{i=1}^k H_1^i} \quad (2-3)$$

### 2.2.2 空间直方图特征

基于上述空间直方图，可以发现直方图的相似性适合用于物体检测。具体做法是把直方图的交作为距离度量，通过判断到样本和物体类别模板的距离来检测物体。以此为起点，本文提取空间直方图特征来表示物体和判别物体类别。

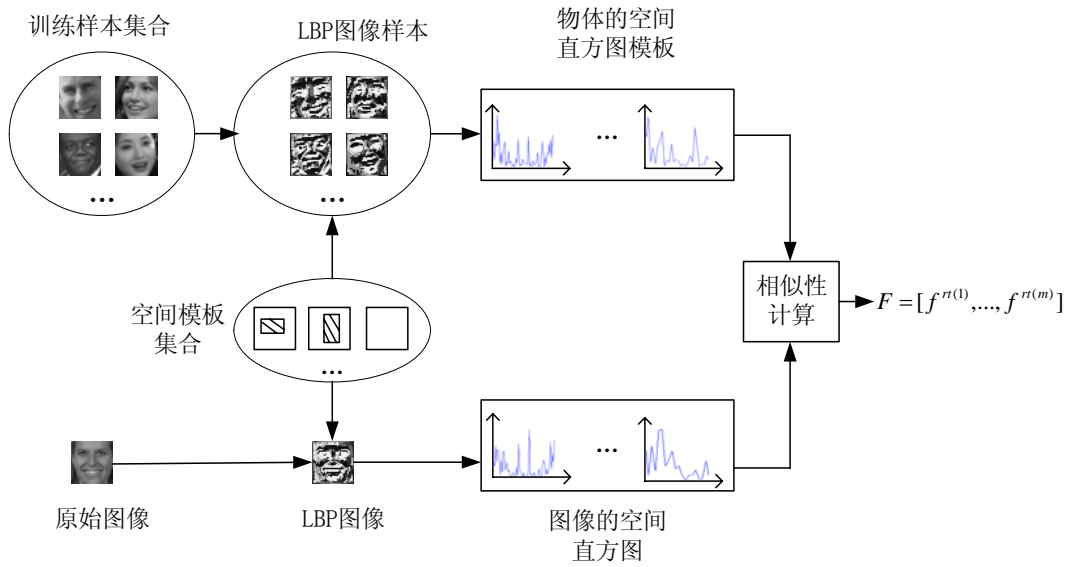


图 2-7 空间直方图特征的提取框架

Figure 2-7 Framework for extraction of spatial histogram features

空间直方图特征本质上与目标物体类别相关联，可以在目标物体大量数据样本的基础上获取。空间直方图特征的提取方法可以概括为如图 2-7 所示的框架。它包括两个部分：物体空间直方图模板的提取和空间直方图特征的计算。简述如下：

1. **物体空间直方图模板：**给定一个物体类别的样本集合（其中包含  $n$  个物体样本）和一个空间模板，那么基于此空间模板的物体直方图模型是样本集合的平均直方图，定义如下：

$$SH^{rt(x,y,w,h)} = \frac{1}{n} \sum_{j=1}^n SH^{rt(x,y,w,h)}(P_j) \quad (2-4)$$

其中  $P_j$  是训练样本,  $rt(x, y, w, h)$  是空间模板。

2. **空间直方图特征的计算:** 对于任意一个样本  $P$ , 定义空间直方图特征(spatial histogram feature)为  $P$  的空间直方图和物体空间直方图模型的交, 计算方法由公式(2-5)给出:

$$f^{rt(x, y, w, h)}(P) = D(SH^{rt(x, y, w, h)}(P), SH^{rt(x, y, w, h)}) \quad (2-5)$$

综上所述, 一个物体模式对应一个空间模板集合  $\{rt(1), \dots, rt(m)\}$  ( $m$  是模板数目), 一个物体样本由  $m$  维的直方图特征表示, 其特征向量为:  $F = [f^{rt(1)}, \dots, f^{rt(m)}]$ 。空间模板可以设置为不同位置和不同尺寸, 导致空间直方图特征集合的基数很大。因此, 空间直方图特征集合可以描述物体的纹理分布和结构信息。

## 2.3 空间直方图特征的判别分析

每种类型的空间直方图特征都具有判别目标的能力, 从而为其成为物体检测特征创造了客观条件。根据不同结构类型的物体, 本文依次选择了刚体结构类型、非刚体结构类型和纹理模式类型的三种物体实例, 对空间直方图特征的类别可分性进行定性说明。在此观察基础上, 本文采用 Fisher 准则来度量空间直方图特征的类别可分性。

### 2.3.1 类别可分性的实例说明

类别可分性是模式识别领域中用于分析特征有效性的重要概念。某个特征的类别可分性是指采用此特征集合来描述样本时, 各个类别之间的可区分能力。

我们分别以侧面汽车、人脸和视频文字为例, 选取特定的空间直方图为例说明空间直方图特征的类别可分性。具体做法是选取一批图像样本和一个空间模板, 在对应的空间直方图特征上求取物体正反例样本集合的特征值分布, 然后采取阈值化操作来分析空间直方图特征的类别可分性。

#### 2.3.1.1 刚体结构物体

刚体结构物体是指物体的形变程度小, 它具有各种显著的部件, 是由这些部件按照固定的空间结构排列而构成的。日常生活中常见的许多人工制造



物体，比如汽车、飞机、电脑等物体，都属于刚体结构物体的范畴。我们以侧面汽车作为刚体结构物体的例子，来分析空间直方图特征对此类物体的类别可分性。

定义侧面汽车图像模式的大小为  $100 \times 40$ ，采用的模板  $rt(40,20,20,20)$  是图像上位于  $(40,20)$  的大小为  $20 \times 20$  的空间模板。 $SH_{car}^{rt(40,20,20,20)}$  是根据公式 (2-4) 在一个包含 200 幅侧面汽车样本的集合上，计算得到的侧面汽车的一个直方图模板。测试特征采用样本到  $SH_{car}^{rt(40,20,20,20)}$  的距离  $f^{rt(40,20,20,20)}(P)$ 。

图 2-8 显示了侧面汽车正负样本集的特征值分布图。其中，侧面汽车样本有 2000 个，非汽车样本有 15000 个。横轴表示特征值的范围，纵轴反映了物体样本在特征值上的分布。

如图 2-8 所示，如果采用 0.7 阈值来划分目标，可以得到 99.1% 的正确检测率和 45.1% 的误检率；阈值 0.8 产生 93.8% 的正确检测率和 12.1% 的误检率。

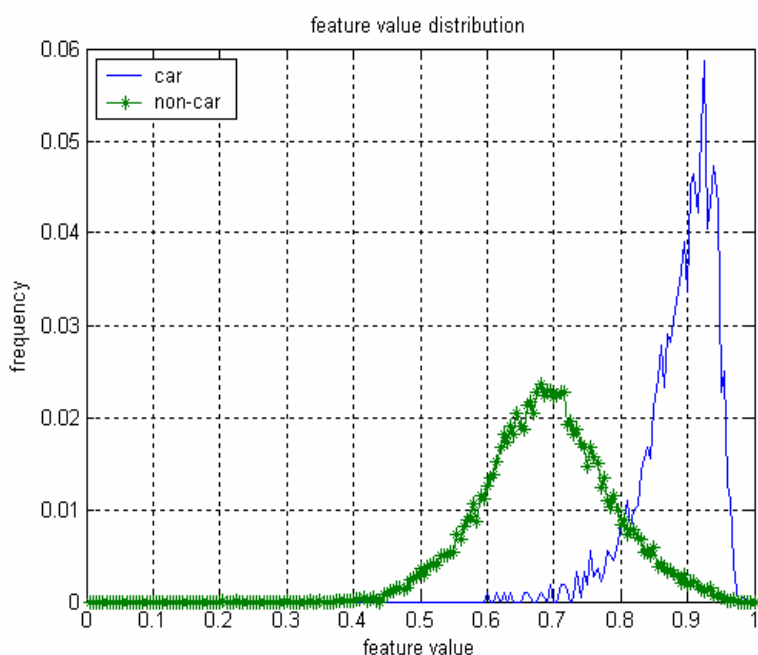


图 2-8 特征分布：侧面汽车和非汽车

Figure 2-8 Feature distribution: side-view car and non-car

### 2.3.1.2 非刚体结构物体

非刚体结构物体是指物体具有柔性的形变程度。与刚体物体类型似，这种类型的物体具有各种显著的部件，是由这些部件按照一定的空间结构排列而构成的。不同之处在于物体本身及其部件都是非刚体的柔性三维结构体，具有可变形的性质。人脸是一种典型的非刚体结构物体，我们以人脸为例，分析空间直方图特征对非刚体结构物体的类别可分性。

定义人脸图像模式的大小为  $32 \times 32$ ，采用的模板  $rt(0,0,32,32)$  是图像上位于  $(0,0)$  的大小为  $32 \times 32$  的模板。 $SH_{face}^{rt(0,0,32,32)}$  是根据公式(2-4)在一个包含 230 幅人脸样本的集合上，计算得到的人脸的一个直方图模板。采用样本到  $SH_{face}^{rt(0,0,32,32)}$  的距离  $f^{rt(0,0,32,32)}(P)$  作为测试特征。

图 2-9 显示了正面人脸正负样本集的特征值分布图。其中，人脸样本有 11200 个，非人脸样本有 12500 个。如图 2-9 所示，如果采用 0.7 阈值来划分目标，可以得到 99.6% 的人脸正确检测率和 30.2% 的误检率。

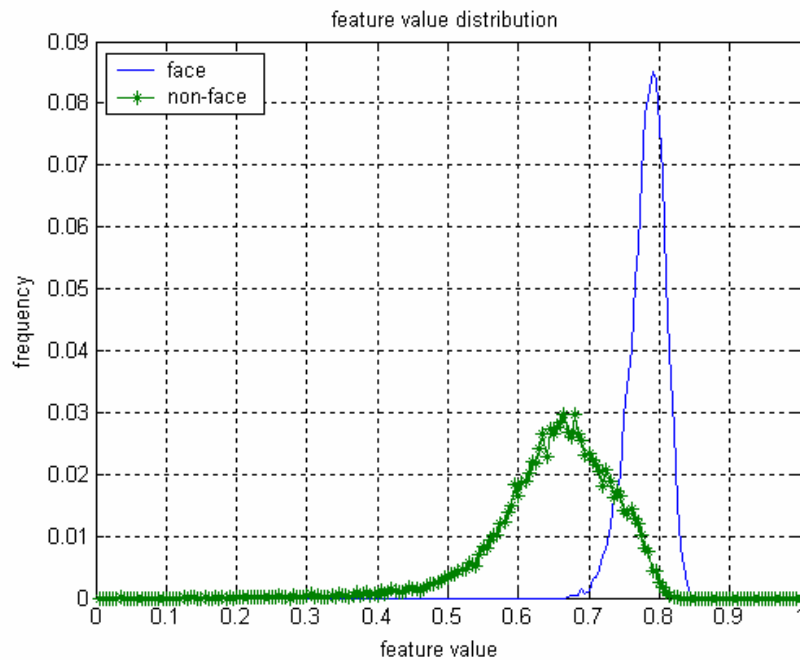


图 2-9 特征分布：人脸和非人脸

Figure 2-9 Feature distribution: face and non-face

### 2.3.1.3 纹理模式物体

客观现实中还有一种类型的物体，它们以纹理特点为主，不但不具备空间结构，而且也没有刚体性质。我们称此类型物体为纹理模式物体。日常生活中很多物体都属于纹理模式物体，例如文字、树叶、窗帘、衣服等。我们以文字图像为例，分析空间直方图特征对纹理模式物体的类别可分性。

定义文字图像模式的大小为  $50 \times 20$ ，采用的模板  $rt(0,5,50,10)$  是图像上位于  $(0,5)$  的大小为  $50 \times 10$  的模板。 $SH_{text}^{rt(0,5,50,10)}$  是根据公式(2-4)在一个包含 180 幅文字样本的集合上，计算得到的一个文字图像直方图模板。采用样本到  $SH_{text}^{rt(0,5,50,10)}$  的距离  $f^{rt(0,5,50,10)}(P)$  作为测试特征。

图 2-10 显示了文字正负样本集的特征值分布图。其中，文字图像样本有 1936 个，非文字图像样本有 12313 个。如图 2-10 所示，如果采用 0.75 阈值来划分目标，可以得到 96.9% 的文字正确检测率和 43.1% 的误检率；如果采用 0.8 阈值来划分目标，可以得到 87.5% 的文字正确检测率和 17.6% 的误检率。

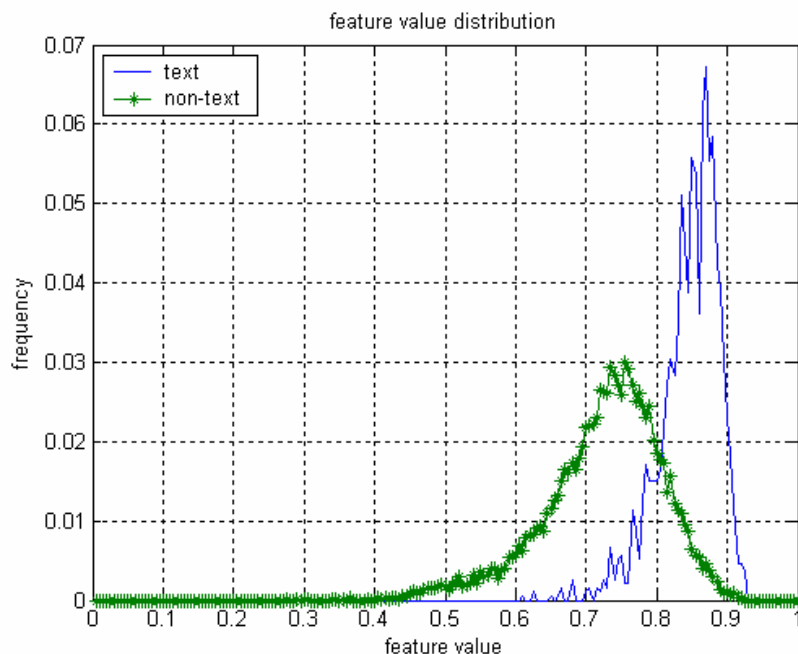


图 2-10 特征分布：文字和非文字

Figure 2-10 Feature distribution: text and non-text

### 2.3.2 类别可分性的度量

本文使用 Fisher 准则函数[72]作为空间直方图特征的类别可分性判据的度量。对于任意一个空间直方图特征  $f_j$ ， $1 \leq j \leq m$ ，假设有  $N$  个样本  $x_1, x_2, \dots, x_N$ ，每个样本  $x_i$  都由  $f_j$  的特征值表示。在这些样本中， $N_1$  个样本属于目标物体类  $\omega(1)$ ， $N_2$  个样本不属于目标物体类，归为  $\omega(2)$ 。样本类间距离  $S_b$  表示两个类别间的距离：

$$S_b = (m_1 - m_2)^2 \quad (2-6)$$

其中， $m_i = \frac{1}{N_i} \sum_{x \in \omega(i)} x, i \in \{1, 2\}$ ， $m_i$  是样本均值向量。各类的样本类内距离  $S_i$  计算公式如下：

$$S_i = \frac{1}{N_i} \sum_{x \in \omega(i)} (x - m_i)^2, i \in \{1, 2\} \quad (2-7)$$

总类内距离  $S_w$  是各类的样本类内距离之和，即

$$S_w = S_1 + S_2 \quad (2-8)$$

从分类性能的角度看，我们希望在特征空间里，各类样本尽可能分得开些，即希望样本类间距离  $S_b$  越大越好；同时希望各类样本内部尽量密集，即类内距离越小越好。因此，作为可分性判据，Fisher 准则函数  $J$  被定义为样本类间距离  $S_b$  和总类内距离  $S_w$  的比值：

$$J(f_j) = \frac{S_b}{S_w} \quad (2-9)$$

一般而言， $J(f_j)$  越大，表示空间直方图特征  $f_j$  的类别可分性越强。

## 2.4 空间直方图特征在彩色图像人脸检测中的应用

从以上分析可以发现，空间直方图特征具有对物体的类别可分性。为了从更深入的角度来验证这一点，本文把人脸检测作为应用实例，引入空间直方图特征来表示人脸。

除了灰度信息之外，物体还具有更丰富的彩色信息。空间直方图特征可以推广至彩色信息空间。对于人脸检测，彩色信息也是一种重要的区分特征。而且，在彩色图像中，人脸肤色的彩色信息特征易于提取。

因此，本节选用彩色图像中人脸检测进行研究。其目的包含两个方面：第一个目的是提取基于彩色信息的空间直方图特征；第二个目的是通过人脸检测来初步验证空间直方图特征的物体表征能力和区分能力。

### 2.4.1 基于彩色信息的空间直方图特征

彩色信息常常作为一种有效的特征，在物体表示和物体识别中得到了很多应用。常见的彩色表示方法是基于三基色理论的 RGB 空间，RGB 空间经过变换，可以得到彩色的色度表示。本文综合颜色空间研究中的理论和技术 [73,74]，把彩色信息从 RGB 空间转换到 YUV 空间，在此基础上提取空间直方图特征。

一般彩色图像中包含了 RGB 空间信息，从 RGB 空间到 YUV 空间的转换可以用矩阵表示为如下公式：

$$\begin{pmatrix} Y \\ U \\ V \end{pmatrix} = \begin{pmatrix} 0.299 & 0.587 & 0.114 \\ -0.147 & -0.289 & 0.436 \\ 0.615 & -0.515 & -0.100 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} \quad (2-10)$$

在 YUV 空间，亮度由 Y 表示，即通常所采用的灰度信息。色调是彩色彼此之间相互区分的特性，如红、橙、黄、绿、蓝等。在 YUV 空间中，色调由 U 和 V 构成的向量的相位角表示，通过公式(2-11)计算：

$$\theta = \tan^{-1}(V/U) \quad (2-11)$$

在这一节中，人脸模式定义为尺寸为 32x32 的彩色图像。综合 RGB 空间和 RGB 空间的彩色信息，可以采用 Y、R、G、B 和  $\theta$  等五个分量来表示

人脸图像，记作  $m(i), i = 1, 2, 3, 4, 5$ ，称为彩色信息向量。针对一幅彩色图像提取空间直方图特征，类似于灰度图像中空间直方图特征提取方法。首先获得图像的彩色信息向量；然后分别对五个分量进行 LBP 预处理，得到五个相应的 LBP 图像；最后根据空间模板集合计算每个分量上的 LBP 直方图。

人脸是由几个显著器官按照一定几何的结构关系构成的物体模式。本文设计了 23 种特殊的空间模板来定义人脸的结构特点。如图 2-11 所示，每个模板用白色矩形区域表示，这些模板涵盖了人脸图像中的眼睛、鼻子、嘴部和腮部等区域。这 23 个人脸空间模板构成了一个模板集合，把它表示为  $\{rect(1), rect(2), \dots, rect(23)\}$ 。

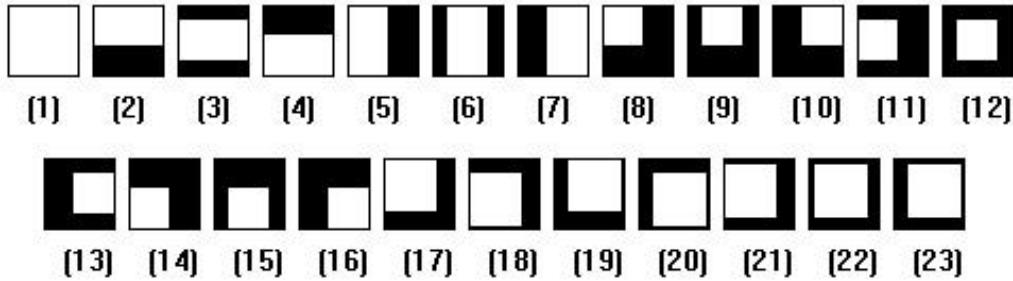


图 2-11 人脸模式的 23 个空间模板

Figure 2-11 The 23 spatial templates of face pattern

基于彩色信息的空间直方图特征是在这个模板集合的基础上进行计算得到的。对于每个分量，可以在彩色图像中获得 23 个空间直方图特征，即  $\{f_{m(i)}^{rect(1)}, \dots, f_{m(i)}^{rect(23)}\}, i = 1, 2, 3, 4, 5$ 。为了表示人脸模式，把所有分量的空间直方图特征组合在一起，构成一个 115 维的基于彩色信息的空间直方图特征向量，即

$$F = [f_{m(1)}^{rect(1)}, \dots, f_{m(1)}^{rect(23)}, \dots, f_{m(5)}^{rect(1)}, \dots, f_{m(5)}^{rect(23)}] \quad (2-12)$$

## 2.4.2 检测方法

人脸检测是一个区分人脸和非人脸的两类分类问题，我们采用简单的直方图匹配的办法来进行目标分类。直方图匹配方法是针对单个空间直方图特征进行的，它的框架如图 2-12 所示。

1. 获取一个人脸空间直方图特征的模板:  $Sh(M)$ 。
2. 决策:  
 图像窗口用  $I$  表示,  $Sh(I)$  是相应的空间直方图。  
  
 如果  $D(Sh(I), Sh(M)) \geq T$ , 则  $I$  属于人脸类;  
  
 否则,  $I$  属于非人脸类。  
 其中,  $T$  是匹配阈值。

图 2-12 直方图匹配方法

Figure 2-12 Histogram matching method for face detection

对于人脸检测, 本文指定了三个人脸区域上的空间直方图特征来进行直方图匹配。它们分别是人脸的全局区域、双眼区域和中间竖直区域, 如图 2-13 所示。

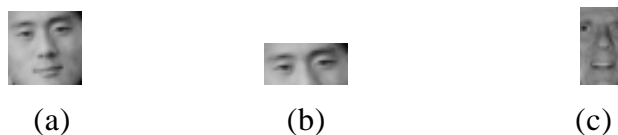


图 2-13 直方图匹配特征区域: (a)人脸整体区域, (b)双眼区域, (c) 中间竖直区域  
 Figure 2-13 Spatial templates for histogram matching in face detection : (a) entire face region, (b) eye region, and (c) middle vertical region

在直方图匹配阶段, 一幅输入图像首先被转换成 LBP 图像, 然后按照 1.2 的比例根据金字塔结构缩小图像, 并且全局搜索得到所有位置上的  $32 \times 32$  子图像。如图 2-14 所示, 这些子图像由两级直方图匹配过程(包括整体匹配和局部匹配)来进行人脸模式的判断。在实验中, 三个直方图匹配过程中的匹配阈值都由人工指定, 取值为 0.7。

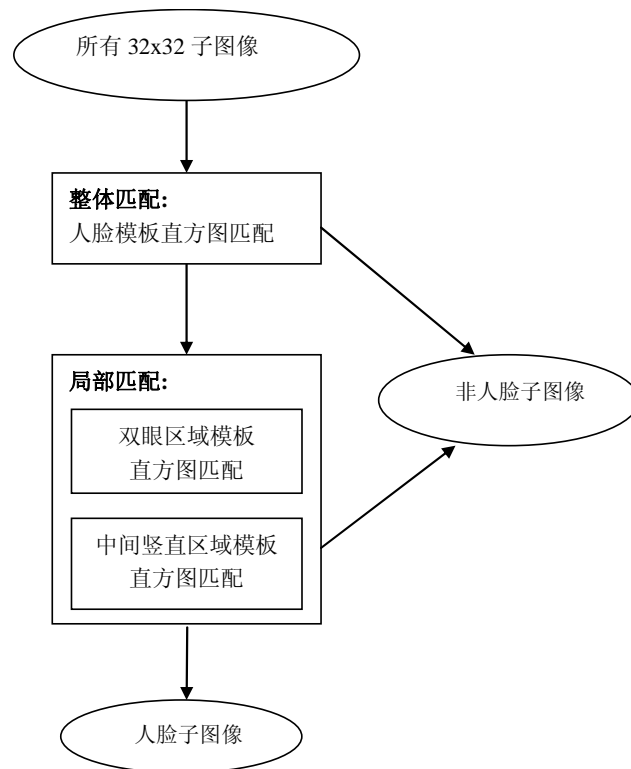


图 2-14 两级直方图匹配的人脸检测流程图

Figure 2-14 Flowchart of two-step histogram matching for face detection

### 2.4.3 实验评估

本文实现了上述检测方法，构建了一个针对彩色图像的正面人脸检测系统，并进行了性能测试。训练数据集是 11400 幅彩色人脸样本和 3000 幅非人脸的样本，尺寸是 32x32。很多研究者使用的人脸测试图像数据都是灰度图像，比如 FERET 人脸库和 CMU\_MIT 人脸库。为了测试本文方法，采集了一个大规模的彩色人脸图像库。采集数据来源于新闻图片、影视图片、个人数码相册和监控视频等多种途径。这些图像中包括了复杂背景下的各种人脸，光照条件变化多端。彩色人脸图像库中总共有 544 张图片，包含了 833 个正面人脸；这些人脸在颜色、位置和尺寸等方面都有很大的变化，而且具有各种不同的表情。

在此人脸图像库上，本文方法在误检为 145 的条件下，正确的检测到了 727 个人脸。测试结果如表 2-1 所示，正确检测率是 87.3%，检测精度是 83.4%。检测精度是指正确检测到的人脸数目与所有检测结果数目(正确检测



到的人脸数目和误检的人脸数目之和)的比值。

表 2-1 彩色图像中人脸检测的实验结果

Table 2-1 Testing results of face detection in color images

人脸数目	正确检测到的人脸数目	误检的人脸数目	正确检测率	检测精度
833	727	145	87.3%	$727/(727+145)=83.4\%$

本文方法可以检测复杂背景下的正面人脸。图 2-15 显示了一些检测实例。本文方法可以同时检测不同颜色、不同大小的人脸；而且还可以处理肤色色调比较暗的人脸和色调比较亮的人脸。同时，本文方法对光照变化具有一定的鲁棒性，图 2-16 显示了光照变化剧烈情况下的一些检测实例。

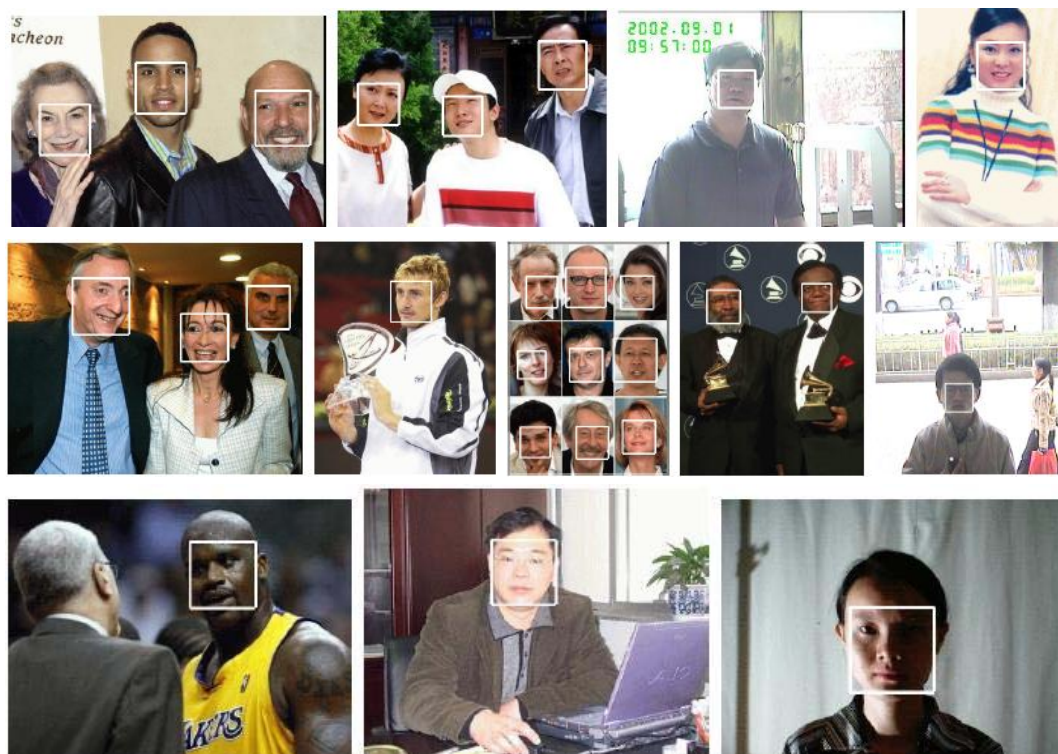


图 2-15 彩色图像中人脸检测的示例

Figure 2-15 Examples of face detection in color images

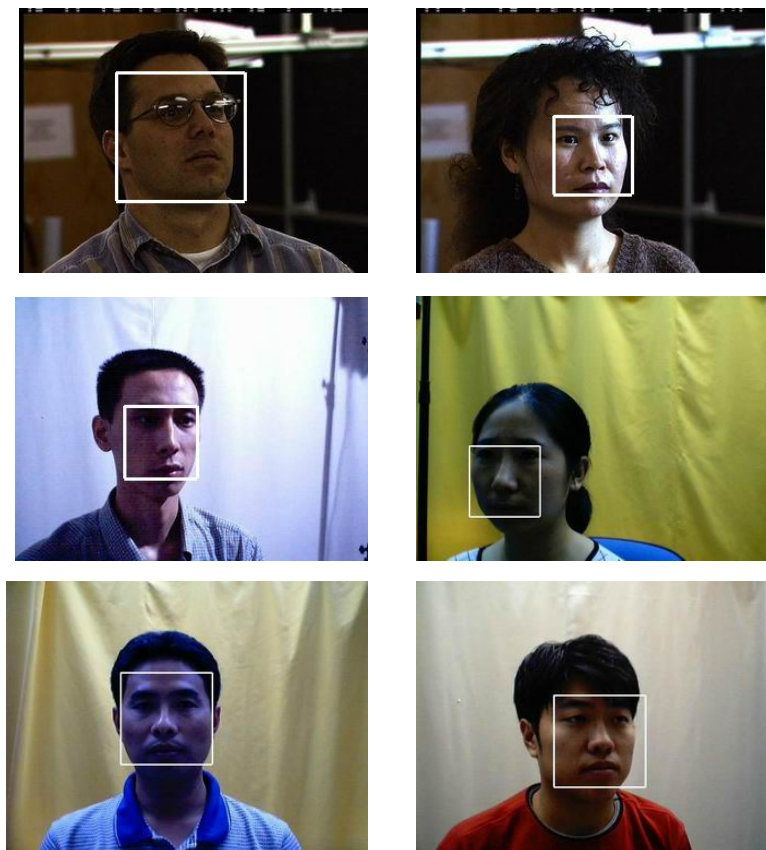


图 2-16 光照变化剧烈的彩色图像中人脸检测示例

Figure 2-16 Examples face detection under various lighting conditions

## 2.5 本章小结

本章围绕物体表示和特征提取问题，提出了改进的空间直方图特征，建立了针对特定目标物体的空间直方图特征提取方法。空间直方图特征是一种局部纹理特征，同时刻画了物体的纹理性质和空间分布。归纳起来，空间直方图特征具有以下性质：

1. 空间直方图特征是局部图像的信息分布，它同时反映了物体的纹理和结构信息；
2. 空间直方图特征可以在不同尺度上表征物体；
3. 空间直方图特征本质上与目标物体类别相关联，可以在目标物体大量数据样本的基础上获取；
4. 除了在灰度信息上可以提取空间直方图特征之外，还可以将空间直

方图特征概念运用到彩色信息分量上。

在实验分析方面，本章以彩色图像中正面人脸检测为例，考察了空间直方图特征的类别可分性。实验结果表明空间直方图特征可以有效的表示人脸模式，所提出的方法能够检测彩色图像中复杂背景下的正面人脸。这充分验证了空间直方图特征有利于物体检测的判别能力。

## 第3章 基于空间直方图特征的物体检测

### 3.1 引言

自上世纪 90 年代以来,物体检测技术得到了蓬勃的发展。例如,人脸检测技术取得了长足的进步,在很多领域中得到了广泛的应用。人们在其它类别物体的检测技术方面也获得了很大的进展,例如汽车检测、行人检测、文字检测等。但是,大多数检测方法是专门针对某一特定类别的物体的,不具备良好的可扩展性。近几年来,人们开始面向非特定类别的物体检测技术展开研究,目前进展不大,主要困难是现有技术在实际应用中达不到检测性能高的要求。

本文针对可扩展性和检测性能两方面的要求,提出了一个可适用于非特定类别物体的检测方法。本文认为首要问题是在特征层面扩大物体类别范围,使得特征表示方法可适用于不同类型的物体。因此,我们采用空间直方图特征来表示物体,并且把所提出的检测方法称为基于空间直方图特征的物体检测方法。

从系统框架角度看,本文方法采用分层分类器的方法进行由粗到精的物体检测。该方法充分利用了空间直方图特征的特点,在粗检测阶段通过联合直方图匹配方法在图像中排除大部分非目标物体区域,在细检测阶段使用支持向量机分类器进行精确检测,最后得到检测结果。

如第 2 章所述,空间直方图特征向量是一个高维向量,不适合直接用作分类器的输入特征;因此进行有效分类器设计的关键是如何构造一个紧致而信息丰富的低维特征向量。目前物体检测领域中常用的特征选择方法大多遵循 Viola 和 Jones 在 2001 年提出的思路[21],即利用 Adaboost 算法学习得到很多由单个特征构成的弱分类器,通过迭代方式改变训练样本和弱分类器的权值,最后把弱分类器组合成为强分类器。虽然很多研究者对这类方法进行了大量的扩展和改进[33,61,63,107],但是这类方法目前依然存在一些不足。首先,Adaboost 算法中只考虑了特征的分类性能,而没有考虑特征之间的相关性;其次,有些训练样本在选择过程的迭代步骤中分类结果前后不一致,造成 Adaboost 训练算法不容易收敛。

与基于 Adaboost 算法的特征选择方法不同,本章提出了一种综合特征

的分类性能和特征相关性度量的特征选择方法，并将该方法融合于物体检测分类器的构造过程中。

本章其它部分组织如下。第2节介绍了基于空间直方图特征的物体检测方法的系统框架。第3节进一步研究了直方图匹配方法，并且提出了联合直方图匹配方法及其训练算法。第4节针对物体检测问题，提出了一种新颖有效的特征选择方法。第5节对所提出的物体检测方法进行了实验评估。最后是本章小结。

## 3.2 物体检测的系统框架

为了检测图像中不同位置 and 不同尺寸的目标物体，本章采用穷举搜索的策略进行物体检测。以侧面汽车检测作为例子，物体检测过程的系统框架概括为如图 3-1 所示，划分为两个阶段：训练阶段和检测阶段。需要指出的是，本章只利用了图像的灰度信息。

在训练阶段，首先采集大量的目标物体和非目标物体的图像样本，并提取出它们的空间直方图特征表示，然后通过特征选择方法获得有效的分类特征，最后训练用于物体检测的分类器。所采用的分类器包括联合直方图匹配方法和支持向量机，构成由粗到精的分层结构分类器。

在输入图像中进行物体检测的阶段包括三个步骤：(1)构建图像金字塔、(2)物体分类、(3)检测结果合并。

在第一步中，输入图像按照固定比例系数（本章选取的比例系数为 1.2）进行若干次缩放，形成图像金字塔。进行缩放操作的目的是在不同尺度上都能够检测到目标物体。

第二步使用固定尺寸的扫描窗口在图像金字塔中各级图像上进行搜索。对于特定位置的扫描窗口，首先在它包含的图像上提取空间直方图特征；然后采用直方图匹配方法和支持向量机分类器，根据所提取的空间直方图特征进行物体分类，从而判断扫描窗口是否为物体目标。

第三步是检测结果融合的过程，将不同位置上检测到的重叠物体目标合并成为单个目标，最后输出物体检测的结果。

第二步中的物体检测过程是一个由粗到精的分类过程。首先，采用联合直方图匹配作为粗检测方法，可以在获得很高的检测率，同时排除了大部分非目标对象，而且它运算简单、运行速度快。其次，使用分类性能高的支持向量机分类器进行精确判断，能够进一步的排除误检对象，从而提高整个系

统的检测精度。

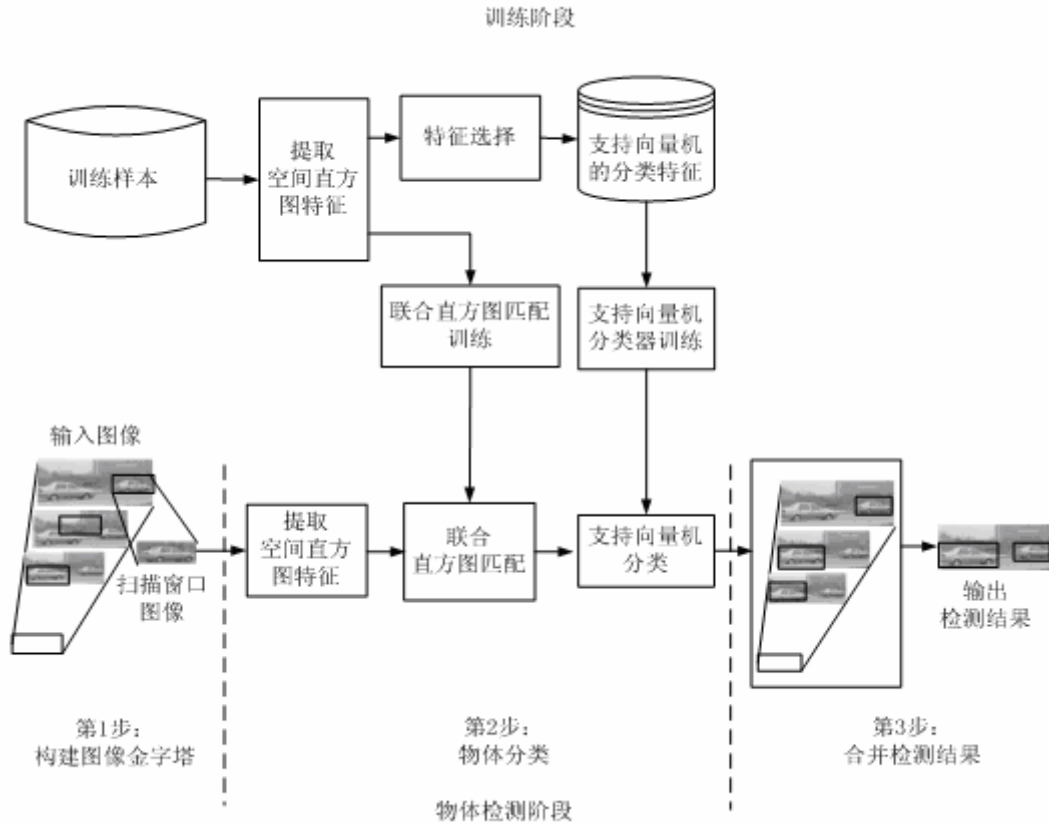


图 3-1 基于空间直方图特征的单类目标物体检测

Figure 3-1 Process of object detection using spatial histogram features

### 3.3 联合直方图匹配

第 2 章采用了直方图匹配的简单方法来区分人脸与非人脸。本章更加深入的研究这种方法，以达到将其应用于一般物体检测的目的。我们首先推导了直方图匹配与基于最小错误率的贝叶斯决策的等价性，然后提出了联合直方图匹配方法，最后给出了联合直方图匹配的训练算法。

#### 3.3.1 联合直方图匹配的推导

对于空间直方图特征，直方图匹配方法可以总结为如下。对于特定的空间模板，首先建立物体类别的空间直方图模型，然后根据样本的空间直方图

到物体的空间直方图模型的距离来判别样本是否属于物体类。假设  $P$  是一个样本， $rt(x, y, w, h)$  是空间模板， $P$  对应于  $rt(x, y, w, h)$  的空间直方图特征是  $f^{rt(x, y, w, h)}(P)$ ，如果  $f^{rt(x, y, w, h)}(P) \geq T$ ，则  $P$  属于目标物体类；否则， $P$  属于非目标物体类。这里， $T$  是分类阈值。

从决策理论角度看，物体检测是一个两种类别的决策问题。在检验一个样本时，其类别  $\omega$  处于两种可能的状态：有可能是目标物体  $\omega_1$ ，也可能是非目标物体  $\omega_2$ 。可以证明：直方图匹配方法在理论上等价于空间直方图特征的极大似然贝叶斯决策。

给定一个空间直方图特征  $f_i (1 \leq i \leq m)$ ，我们用一个变量  $x$  表示它， $(0 \leq x \leq 1)$ 。类别状态的先验概率为  $p(\omega_i)$ ， $i=1,2$ ；类条件概率密度是  $p(x | \omega_i)$ ， $i=1,2$ 。于是，利用贝叶斯法则，根据空间直方图特征  $f_i$  计算类别后验概率的公式如下：

$$p(\omega_i | x) = \frac{p(x | \omega_i) p(\omega_i)}{\sum_{i=1}^2 p(x | \omega_i) p(\omega_i)} \quad (3-1)$$

根据公式(3-1)可以通过观察变量  $x$  (即样本的空间直方图特征值)把类别的先验概率为  $p(\omega_i)$  转化为类别的后验概率  $p(\omega_i | x)$ 。这样，基于最小错误率的贝叶斯决策规则为：

$$\begin{cases} x \in \omega_1 & \text{if } p(\omega_1 | x) > p(\omega_2 | x) \\ x \in \omega_2 & \text{otherwise} \end{cases} \quad (3-2)$$

上述贝叶斯决策规则可以写为：

$$h(x) = \arg \max_{\omega_i, i=1,2} p(\omega_i | x) \quad (3-3)$$

由于无法获知任何先验知识，我们假设目标物体类和非目标物体类具有相同的先验概率，即  $p(\omega_1) = p(\omega_2)$ 。在此基础上，对公式(3-3)进行推导：

$$\begin{aligned}
 h(x) &= \arg \max_{\omega_i, i=1,2} p(\omega_i | x) \\
 &= \arg \max_{\omega_i, i=1,2} \frac{p(x | \omega_i) p(\omega_i)}{\sum_{i=1}^2 p(x | \omega_i) p(\omega_i)} \\
 &= \arg \max_{\omega_i, i=1,2} p(x | \omega_i) p(\omega_i) \\
 &= \arg \max_{\omega_i, i=1,2} p(x | \omega_i)
 \end{aligned} \tag{3-4}$$

因此可以得到与贝叶斯决策规则等价的极大似然规则：

$$\begin{cases} x \in \omega_1 & \text{if } p(x | \omega_1) > p(x | \omega_2) \\ x \in \omega_2 & \text{otherwise} \end{cases} \tag{3-5}$$

空间直方图特征的条件概率密度为单峰分布。在此情形下，极大似然判决规则(3-5)将样本空间分成两个判决区域  $R_1$  和  $R_2$ ，形成了一个分界面  $t$ 。如图 3-2 所示，判决规则可以简化为以下两种形式：

- (a) 如果  $x \geq t$ ，则把  $x$  归类于目标物体  $\omega_1$ ，反之如果  $x < t$ ，则把  $x$  归类于非目标物体  $\omega_2$ ；
- (b) 如果  $x \leq t$ ，则把  $x$  归类于目标物体  $\omega_1$ ，反之如果  $x > t$ ，则把  $x$  归类于非目标物体  $\omega_2$ 。

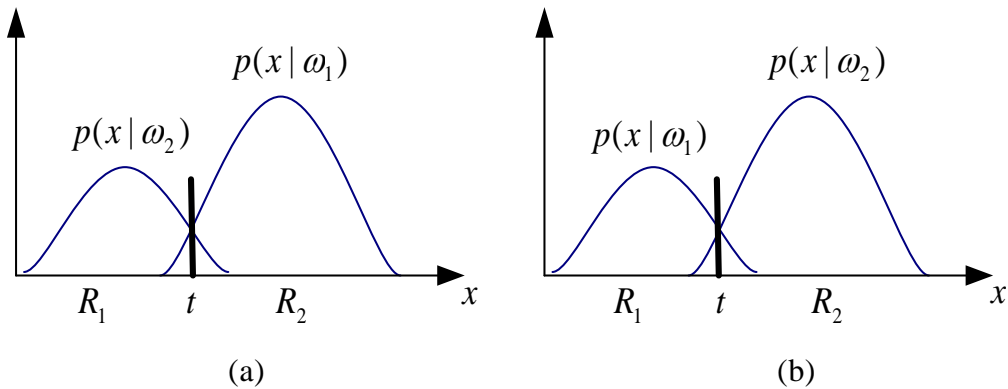


图 3-2 极大似然判决规则简化形式的示意图：形式(a)和形式(b)

Figure 3-2 Simplification of maximum likelihood decision rule: form (a), and form (b)

由于空间直方图特征反映了样本和目标物体模型的相似度，它的特点是：样本的空间直方图特征  $x$  值越大，样本隶属于目标物体类的可能性则越



大。因此，空间直方图特征的贝叶斯决策规则采用的形式是图 3-2 中的形式 (a)，即：

$$\begin{cases} x \in \omega_1 & \text{if } x > t \\ x \in \omega_2 & \text{otherwise} \end{cases} \quad (3-6)$$

显然，直方图匹配方法的原理和公式(3-6)在本质上是相同的。对于任意空间直方图特征  $f_j$ ，采用直方图匹配方法和贝叶斯决策进行物体判别的规则可以统一表示，即：

$$H(P) = \begin{cases} 1 \text{ object} & \text{if } (f_j(P) \geq T_j) \\ 0 \text{ non-object} & \text{otherwise} \end{cases} \quad (3-7)$$

其中  $P$  是待分类样本， $T_j$  是判别阈值。

基于单个特征的直方图匹配显然无法达到物体检测系统所要求的精度。本文选择出一些可分性强的特征，并把它们串联成分级的形式进行直方图匹配，我们称之为联合直方图匹配。如果确定了  $n$  个空间直方图特征  $f_1, \dots, f_n$ ，以及相对应的分类阈值  $T_1, \dots, T_n$ ，那么联合直方图匹配的判决准则如下：

$$H(P) = \begin{cases} 1 \text{ object} & \text{if } (f_1(P) \geq T_1) \wedge \dots \wedge (f_n(P) \geq T_n) \\ 0 \text{ non-object} & \text{otherwise} \end{cases} \quad (3-8)$$

其中  $P$  是待分类样本。

### 3.3.2 联合直方图匹配的训练方法

本文使用可分性判据和检测率衡量特征的分类性能，提出了一个联合直方图匹配的训练方法。检测率是指在一个正例样本集合上的分类正确率。这个方法选择出分类性能强的特征子集  $F_{select}$  和相应的阈值集合  $ThreSet$ ，同时构造联合直方图匹配分类器。

假设联合直方图匹配的训练方法具有以下输入条件:(1)空间直方图特征

集合  $F = \{f_1, \dots, f_m\}$ , (2) 正例训练样本集合  $SP$  和反例训练样本集合  $SN$ , (3) 正例校验样本集合  $VP = \{(x_1, y_1), \dots, (x_n, y_n)\}$  和反例校验样本集合  $VN = \{(x'_1, y'_1), \dots, (x'_k, y'_k)\}$ , 其中  $x_i$  和  $x'_i$  是具有  $m$  维的空间直方图特征的样本,  $y_i = 1$  表示正例样本,  $y'_i = 0$  表示反例样本, (4) 可接受的检测率:  $D$ 。联合直方图匹配训练的具体过程如下所述:

- (1) 初始化:  $F_{select} = \emptyset$ ,  $ThreSet = \emptyset$ ,  $t = 0$ , 将反例校验样本集合  $VN$  上的分类正确率设置为 0, 即  $Acc(pre) = 0$  和  $Acc(cur) = 0$ ;
- (2) 对于每个空间直方图特征  $f \in F$ , 在训练样本集合  $SP$  和  $SN$  上, 计算其 Fisher 准则函数  $J(f)$ ;
- (3) 求出 Fisher 准则函数值最大的空间直方图特征  $f_t$ , 即

$$f_t = \arg \max_{f_j} \{J(f_j) | f_j \in F\};$$

- (4) 在校验样本集合  $V = VP \cup VN$  上, 用空间直方图特征  $f_t$  执行直方图匹配, 找到一个阈值  $\theta_t$ , 使得在正例校验样本集合  $VP$  上的检测率  $d$  大于  $D$ , 即  $d \geq D$ ;
- (5) 计算反例校验样本集合  $VN$  上的分类正确率, 即:

$$Acc(cur) = 1 - \frac{1}{k} \sum_{i=1}^k |H(x'_i) - y'_i|。其中 H(x) 是基于  $f_t$  和  $\theta_t$  的直方图匹配对样本  $x$  的分类结果,  $H(x) \in \{0, 1\}$ ;$$

- (6) 如果分类正确率满足条件:  $Acc(cur) - Acc(pre) \leq \varepsilon$  ( $\varepsilon$  是一个小的正数), 那么训练过程结束, 返回特征子集  $F_{select}$  和阈值集合  $ThreSet$ ; 否则, 执行以下步骤:

$$(a) \quad Acc(pre) = Acc(cur), \quad SN = \emptyset, \quad F_{select} = F_{select} \cup \{f_t\},$$

$$F = F \setminus \{f_t\}, \quad ThreSet = ThreSet \cup \{\theta_t\}, \quad t = t + 1;$$

(b) 在一个不含有物体目标的图像集合上，执行基于  $F_{select}$  和  $ThreSet$

的联合直方图匹配，把误检结果添加到  $SN$  中；

(c) 跳转至(2)，继续下一次选择特征的循环。

联合直方图匹配训练算法的流程图概括为图 3-3 所示。

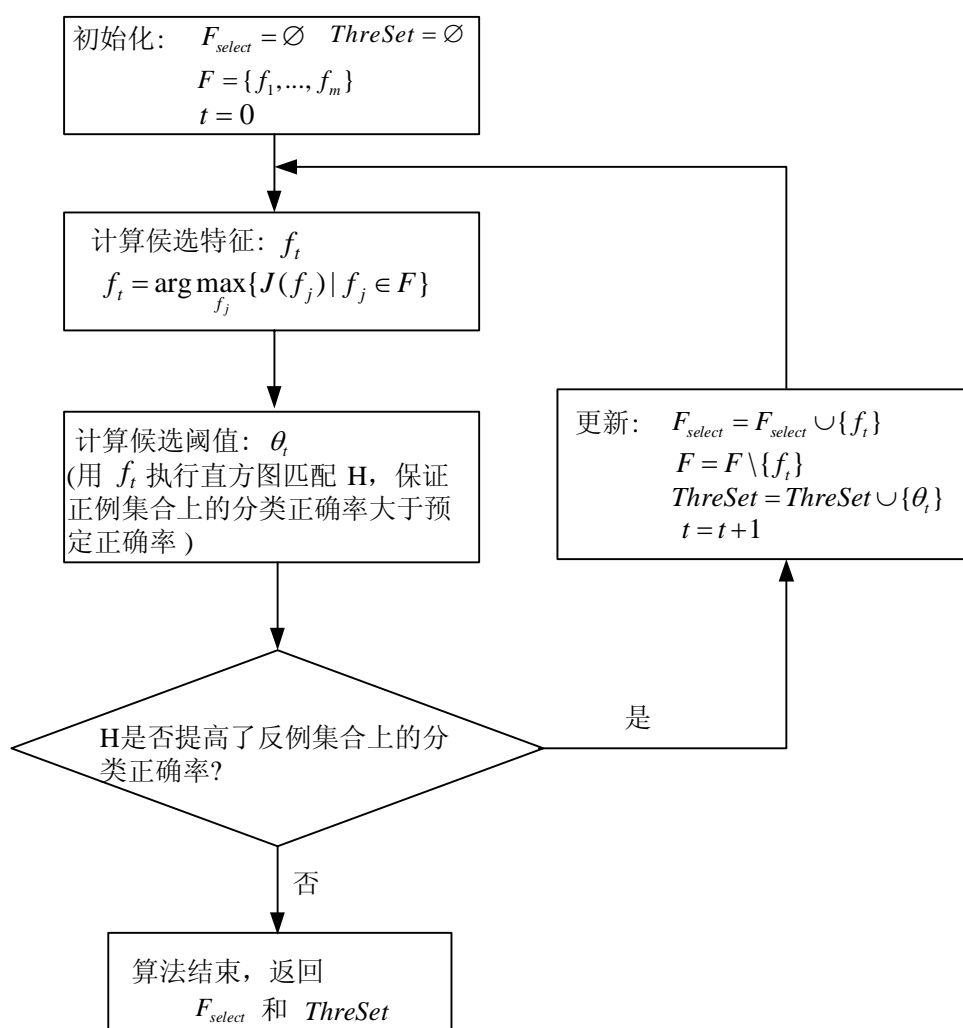


图 3-3 联合直方图匹配训练算法的流程图

Figure3-3 Training method of cascade histogram matching

### 3.4 支持向量机分类器及其特征选择

联合直方图匹配是物体检测的粗检过程，保持了较高的检测率，但是误检率仍然较高。为了提高检测性能，本文采用支持向量机分类器对粗检结果进行精细的检测。支持向量机[75]是解决两类模式分类问题的基本方法之一，在很多领域得到了成功的应用。它的基本思想是通过非线性变换将输入空间变换到一个高维空间，并且在这个新空间中求取最优线性分类面。

支持向量机的训练样本可以表示为  $(x_i, y_i)$ ，其中  $x_i \in R^N$ ， $y_i \in \{-1, 1\}$  是其对应的类别标志， $i=1, 2, \dots, l$ ， $l$  是训练样本的个数。当训练样本线性不可分时，支持向量机的判别函数表示为：

$$f(x) = \text{sgn}(\sum_{i=1}^l y_i \alpha_i K(x, x_i) + b) \quad (3-9)$$

其中， $\alpha_i$  是  $x_i$  对应的系数， $b = y_i - \sum_{j=1}^l y_j \alpha_j K(x_j, x_i)$ ， $K(x, x_i)$  被称为核函数。当  $x_i$  对应的  $\alpha_i$  满足  $0 < \alpha_i < C$  时， $x_i$  称为支持向量，其中  $C$  是预先给定的松弛因子常数。对于任意两个样本  $x$  和  $x'$ ，通常所采用的核函数有以下几种形式：

(1) 多项式(Polynomial)函数：

$$K(x, x') = (x \bullet x' + 1)^p \quad (3-10)$$

(2) 径向基(Radial Basis Function, RBF)函数：

$$K(x, x') = \exp(-\frac{|x - x'|^2}{2\sigma^2}) \quad (3-11)$$

(3) 感知机(Sigmoid Function)函数：

$$K(x, x') = \tanh(v(x, x') + c) \quad (3-12)$$

其中参数  $p, \sigma, v, c$  预先给定。

本文将支持向量机同时应用于特征选择和分类器构造。首先，在特征选择过程中，采用支持向量机分类器在实际的样本集合上来评价所选取特征的分类性能。然后，在选择出的特征子集上训练用于物体检测的支持向量机分类器。

### 3.4.1 空间直方图特征之间的相关性

有效的特征集合不但要求每个特征各自具有较强的分类性能，而且要求特征之间相互独立或者不相关。因此，特征选择过程需要度量特征之间的相关性。有很多方法可以用于度量相关性，例如互信息[76,77]、相关系数[78]等。本文采用互信息来度量特征之间的相关性。

一个空间直方图特征被看作是一个随机变量  $X$ ，它的熵定义为：

$$H(X) = -\int p(x) \log_2 p(x) dx \quad (3-13)$$

其中  $p(x)$  为  $X$  的概率密度函数。

两个空间直方图特征  $f_1$  和  $f_2$  之间的互信息为：

$$I(f_1 | f_2) = H(f_1) + H(f_2) - H(f_1, f_2) \quad (3-14)$$

其中， $H(f_1, f_2)$  是  $f_1$  和  $f_2$  的联合熵。容易证明  $I(f_1 | f_2) = I(f_2 | f_1)$  和  $0 \leq I(f_1 | f_2) \leq H(f_1)$ 。针对特征相关性，我们定义了两个计算指标。

- 定义 1：两个特征之间的相关性的计算指标

两个特征  $f_1$  和  $f_2$  之间相关性计算如下：

$$Corr(f_1, f_2) = \frac{I(f_1 | f_2)}{H(f_1)} \quad (3-15)$$

$Corr(f_1, f_2)$  是介于 0 和 1 之间的实数，其数值反映了两个空间直方图特征  $f_1$  和  $f_2$  之间的相关程度。 $Corr(f_1, f_2)$  值越大，表明  $f_1$  和  $f_2$  越相关。

- 定义 2：单个特征与特征子集之间相关性的计算指标

假设  $F_s$  是一个特征子集，计算特征  $f_m (f_m \notin F_s)$  和  $F_s$  之间的相关性如下：

$$Corr(f_m, F_s) = \text{Max}\{Corr(f_m, f_k) | \forall f_k \in F_s\} \quad (3-16)$$

### 3.4.2 空间直方图特征的有效选择

除了分类器本身的设计和研究之外，模式分类的一个关键问题是特征选择。从数据集合中提取出原始特征之后，特征选择就是按照某种评价方法，从中选出一个最优特征子集的过程，要达到的目标是构造有效的分类器。因此，选出的特征子集不但要减少数据空间的特征维数，而且要能够充当有效的分类特征。

一个典型的特征选择处理过程可以分为两个重要的部分，即候选特征子集的生成和特征子集的评估。候选特征子集的生成是一个最优搜索问题，人们提出了很多启发式的局部最优搜索解法。在特征子集的评估方面，目前的特征选择技术概括起来，可以分为过滤方法(Filter approach)和封装方法(Wrapper approach)两大类[79]。两者的主要区别在于特征子集的评价方法。过滤方法[77,80]按照某种比较准则来评价特征子集，选择出在此准则下最优的特征，作为分类器的输入特征。封装方法[81,82]在产生候选特征子集之后，直接产生分类器进行校验，用分类错误率评价特征子集，最终选择出分类错误率最小的特征子集。这两类方法各有优缺点，过滤方法训练时间短，但是选择得到的特征子集的分类性能不高；相反，封装方法训练复杂，时间较长，但是获得的特征子集的分类性能高。

对于基于空间直方图特征的物体检测而言，如何从高维的特征空间寻找到信息丰富而且判别能力强的低维特征子集，关系到整个算法的检测性能是否足够精确和算法速度是否满足实际应用的要求。

#### 3.4.2.1 基于类别可分性和特征相关性的特征选择算法

为了向支持向量机分类器提供紧致有效的分类特征，本文提出了基于类别可分性和特征相关性的特征选择算法。该算法具有两个特点：它综合考虑了特征的类别可分性和特征相关性，提高了特征选择的效率；另一方面，它采用分类器评价特征子集的性能，保证了分类性能。

假设原始特征集合是  $F = \{f_1, \dots, f_m\}$ ，其中  $m$  是特征数目。本文提出的特征选择方法分别用 Fisher 准则和互信息来度量特征的类别可分性和特征之间的统计相关性。特征选择过程中，通过分类器测试的方法直接评价候选特征子集的分类性能，以分类器错误率最小为目标，最终选择得到特征子集  $F_{select}$ 。图 3-4 描述了特征选择算法的系统结构。

从整体上看，本文方法通过循环的方式不断的将符合条件的候选特征添加到特征子集  $F_{select}$  中。具体步骤简述如下：

1. 首先初始化时将  $F_{select}$  设置为空集；然后进行特征的可分性计算，得到特征的可分性度量。将可分性最强的特征加入  $F_{select}$  中之后，开始选择  $F_{select}$  的循环过程。

2. 每步迭代时，根据可分性和相关性度量产生候选特征，并且训练一个分类器评价加入候选特征后的  $F_{select}$  的分类性能，如果提高了分类正确

率，则把候选特征添加到特征子集  $F_{select}$ 。

3. 算法结束后，返回特征子集  $F_{select}$ 。分类器设计过程使用  $F_{select}$  作为输入特征，进行分类器训练。

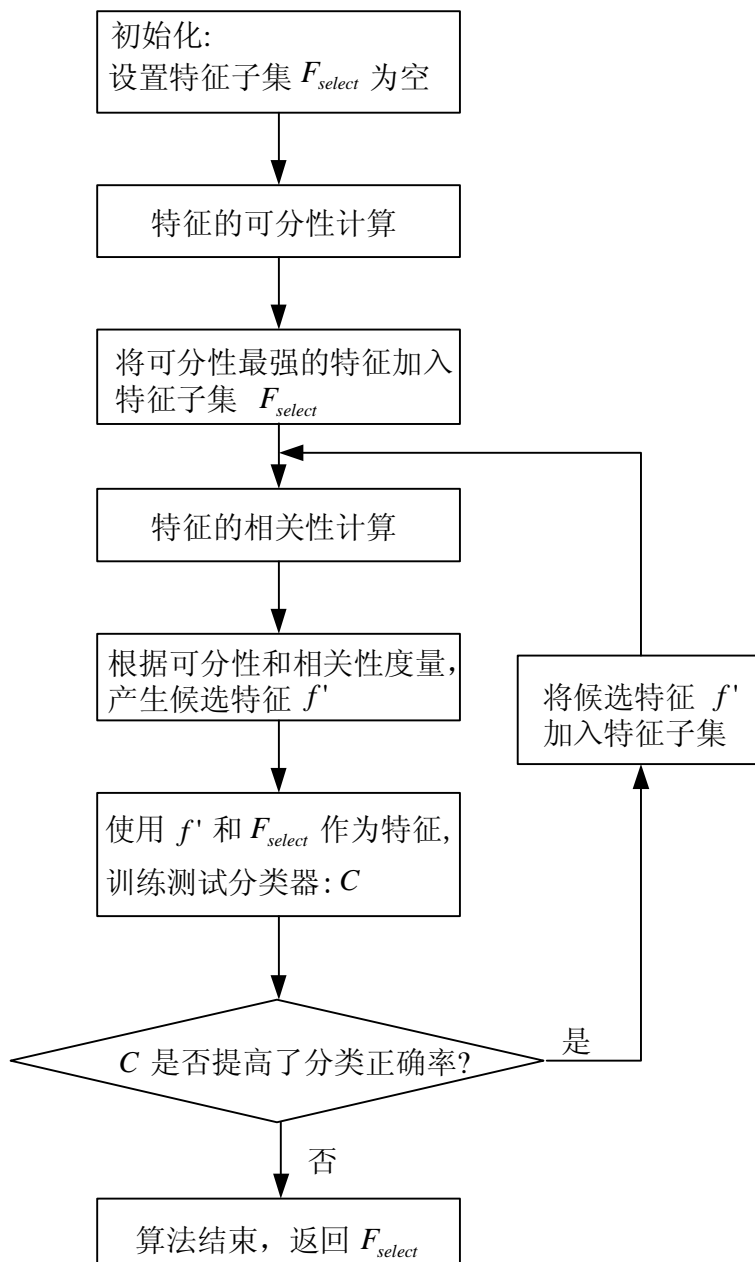


图 3-4 结合可分性和相关性的特征选择的系统结构

Figure 3-4 Feature selection based on class separability and feature correlation

综合特征的类别可分性和特征之间的统计相关性，本文按照一种顺序渐进的方式来选择特征子集  $F_{select}$ ，并且使用测试分类器评价特征子集的分类性能，从而得到有效的分类特征。

假设具有以下输入条件：(1) 空间直方图特征集合  $F = \{f_1, \dots, f_m\}$ ，(2) 训练样本集合  $S = \{(x_1, y_1), \dots, (x_n, y_n)\}$  和测试样本集合  $V = \{(x'_1, y'_1), \dots, (x'_k, y'_k)\}$ ，其中， $x_i$  和  $x'_i$  是具有  $m$  维空间直方图特征的样本， $y_i \in \{0,1\}$  和  $y'_i \in \{0,1\}$  是样本类别标号。特征子集  $F_{select}$  的选择过程如下所述：

(1) 对于  $F$  中的每个特征  $f_j (j=1,2,\dots,m)$ ，在训练样本集合  $S$  上，计算它

的 Fisher 准则函数值  $J(f_j)$ ；

(2) 求出 Fisher 准则函数值最大的特征  $f^*$ ， $F_{select} = \{f^*\}$ ，

$$F_{ori} = F \setminus \{f^*\}；$$

(3) 设置分类正确率初始为 0，即  $Acc(pre) = 0$  和  $Acc(cur) = 0$ ；

(4) 对于每个特征  $f \in F_{ori}$ ，在训练样本集合  $S$  上，计算该特征与特征子

集  $F_{select}$  的相关性度量  $Corr(f, F_{select})$ ；

(5) 计算特征相关性阈值  $Thre$ ：

$$\begin{cases} MinCorr = \min\{Corr(f, F_{select}) \mid f \in F_{ori}\} \\ MaxCorr = \max\{Corr(f, F_{select}) \mid f \in F_{ori}\} \\ Thre = MinCorr * (1 - \alpha) + MaxCorr * \alpha \end{cases} \quad (3-17)$$

这里  $\alpha$  是权值 ( $0 < \alpha < 1$ )，本文在实验中选取  $\alpha = 0.2$ ；



(6) 根据特征相关性阈值  $Thre$ ，求出  $F_{ori}$  中与  $F_{select}$  相关性弱的特征集

$F_{candidate}$ ，即

$$F_{candidate} = \{f_j | Corr(f_j, F_{select}) \leq Thre, f_j \in F_{ori}\} \quad (3-18)$$

(7) 在  $F_{candidate}$  特征集合中求出 Fisher 准则函数值较大的特征  $f' \in F_{ori}$ ，即

$$f' = \arg \max_{f_j} \{J(f_j) | f_j \in F_{candidate}\} \quad (3-19)$$

(8) 使用  $f'$  和  $F_{select}$  作为特征，训练得到一个评价分类器  $C$ ；

(9) 计算校验样本集合  $V$  上的分类正确率：

$$Acc(cur) = 1 - \frac{1}{k} \sum_{i=1}^k |C(x'_i) - y'_i|,$$

其中  $C(x)$  是基于  $f'$  和  $F_{select}$  的分类器  $C$  对样本  $x$  的分类结果，

$C(x) \in \{0,1\}$ ；

(10) 如果分类正确率满足条件： $Acc(cur) - Acc(pre) \geq \varepsilon$  ( $\varepsilon$  是一个小的正数)，那么执行以下更新步骤：

a)  $Acc(pre) = Acc(cur)$ ， $F_{select} = F_{select} \cup \{f'\}$ ， $F_{ori} = F_{ori} \setminus \{f'\}$ ；

b) 跳转至(4)，继续下一次选择特征的循环；

(11) 训练过程结束，返回特征子集  $F_{select}$ 。

上述步骤中的(5)、(6)、(7)步构成了根据类别可分性和特征相关性产生候选特征的过程。首先，计算得到  $F_{ori}$  与  $F_{select}$  的特征相关性阈值  $Thre$ ；然后，求出  $F_{ori}$  中与  $F_{select}$  相关性弱的特征集合  $F_{candidate}$ ；最后，在  $F_{candidate}$  特征集

合中产生可分性较强的候选特征。

实际上，算法中涉及的评价分类器  $C$  可以采用各种形式的分类算法，比如最近邻分类器、贝叶斯法则、决策树、神经网络和支持向量机等。本文采用了支持向量机作为评价分类器  $C$ 。获得特征子集  $F_{select}$  后，利用其作为输入特征训练实际应用于物体检测的支持向量机分类器。

### 3.5 实验评估

在现实世界中，各类物体都具有不同的纹理分布和空间结构，大致可以划分为三种类型。如第 2 章所述，这三种类型分别是刚体结构类型、非刚体结构类型和纹理模式类型。刚体结构类型物体具有显著部件，并且结构关系比较固定，例如汽车。非刚体结构类型物体是可变形的柔性三维结构体，例如人脸。纹理模式类型物体不具有部件结构及空间关系，例如文字。

为了测试基于空间直方图特征的物体检测方法对不同类型物体的有效性和鲁棒性，本文分别对三种典型的物体类别进行了实验，包括正面人脸检测、侧面汽车检测和视频文字检测，以验证空间直方图特征对不同类型物体的表征能力和判别能力。

为了评价物体检测方法的性能，通常采用一些公开通用的标准数据集合作为测试集。主要的衡量指标包括如下几项：

(1)检测率(Detection Rate/ Recall):

$$\text{检测率} = \frac{\text{正确检测到的目标物体的数目}}{\text{测试集中的目标物体的总数}}$$

(2)误检率(False Positive Rate):

$$\text{误检率} = \frac{\text{错误检测到的目标物体的数目}}{\text{测试集中所有扫描窗口的总数}}$$

(3)检测精度(Precision):

$$\text{检测精度} = \frac{\text{正确检测到的目标物体的数目}}{\text{正确检测到的目标物体数目} + \text{错误检测到的目标物体数目}}$$

(4)ROC 曲线(Receiver Operating Curve): ROC 曲线反映了系统的检测率和误检率之间的关系；

(5)RPC 曲线(Recall Precision Curve): RPC 曲线反映了系统的检测率和检测精度之间的关系；

(6)RPC 等错误率 (RPC Equal Error Rate): 和人脸识别不同, 这里 RPC 等错误率的定义是检测率和检测精度相等时的检测率数值。

### 3.5.1 正面人脸检测

我们建立了一个包含 11400 幅正面人脸图像和 17285 幅非人脸图像的训练样本集合, 每个样本的尺寸是  $32 \times 32$ 。这些样本采自于自然图像、网站和视频。另外还收集了 4062 幅人脸图像和 5139 幅非人脸图像作为校验和测试集合。3000 幅不含人脸的自然图像用于产生反例集合。图 3-5 显示了一些人脸训练样本的实例。

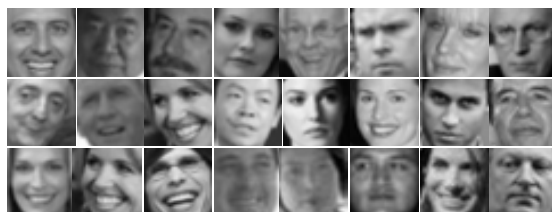


图 3-5 正面人脸训练样本的示例

Figure 3-5 Some training examples of frontal face

在  $32 \times 32$  的图像窗口中定义的空间直方图特征集合的基数很大, 大约是 832351, 此集合对于人脸检测而言过于冗余。实际上, 大部分空间模板都尺寸太小, 或者相互之间位置重叠。为了消除这些冗余的空间模板, 本文限制模板以 5 个像素的步长在水平或者竖直方向上移动, 而且只保留那些尺寸数倍于  $10 \times 10$  的模板用于提取空间直方图特征。

消去冗余模板之后, 我们在联合直方图匹配训练过程中把可接受的检测率  $D$  设为 0.95, 经过训练得到 11 个空间直方图特征(见图 3-6 所示)用于联合直方图匹配。

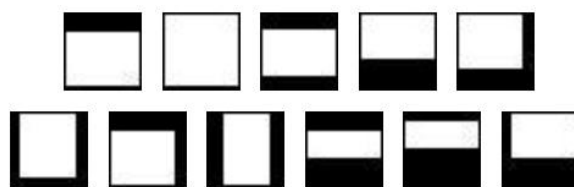


图 3-6 正面人脸检测为联合直方图匹配选取的 11 个空间模板

Figure 3-6 11 Selected spatial templates for cascade histogram matching in frontal face detection

为了提高检测精度，使用基于类别可分性和特征相关性的特征选择算法进行训练，得到 23 个空间直方图特征用于支持向量机分类器。支持向量机分类器采用径向基核 RBF 函数。

### 3.5.1.1 在 CMU+MIT 测试集上的测试结果

本文针对 CMU+MIT 测试集[1]中的图像进行性能比较，这个集合有 130 幅图像，包含 507 个正面人脸。

为了对算法性能做一个整体了解，我们通过逐步改变联合直方图匹配和支持向量机的分类阈值，在测试集上运行正面人脸检测算法，从而获得相对应的检测率和误检率。以下列出了在 CMU+MIT 测试集上的 ROC 曲线，如图 3-7 所示。

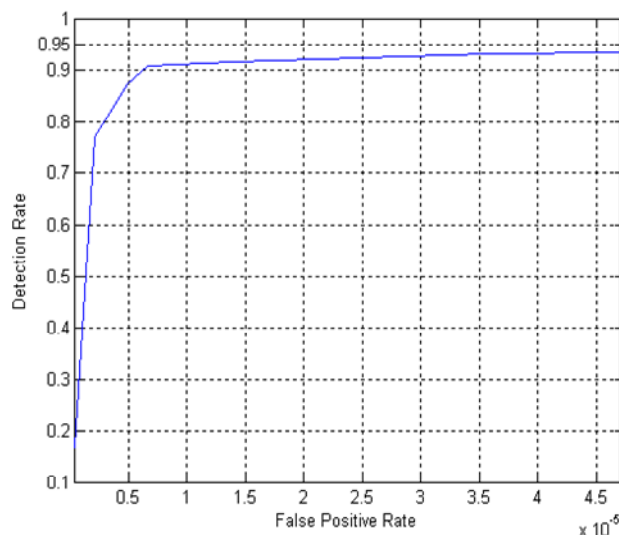


图 3-7 正面人脸检测算法在 CMU+MIT 测试集上的 ROC 曲线

Figure 3-7 ROC obtained on CMU+MIT face detection test set

为了分析层次分类器中不同步骤对系统性能的影响，本文采用了两套实验方案在 CMU+MIT 测试集上进行测试。第一套方案只利用联合直方图匹配方法进行检测；第二套方案利用联合直方图匹配方法和支持向量机分类器组成的分层分类器进行检测。可以通过改变联合直方图匹配的阈值，获得不同参数条件下的测试结果。表 3-1 显示了这两种方案的测试结果。其中参数条件 1 中的联合直方图匹配的阈值比较大，所以获得检测率比较低；参数条件 2 中的联合直方图匹配的阈值比较小，所以获得检测率比较高。

表 3-1 正面人脸检测算法不同阶段在 CMU+MIT 集合上的测试结果

Table 3-1 Results on CMU+MIT set by different phases of frontal face detection

		正确检测 的人脸数目	错误检测 的人脸数目	检测率	检测 精度
参数 条件 1	未使用 SVM	400	230	78.8%	63.4%
	使用 SVM	390	10	76.9%	97.5%
参数 条件 2	未使用 SVM	486	530	95.8%	47.8%
	使用 SVM	478	67	94.2%	87.7%

从表 3-1 的比较结果中可以看出，联合直方图匹配方法能够获得比较高的检测率，但是误检现象比较严重，导致检测精度低。作为精确检测方法，支持向量机分类器大幅度的提高了检测精度，同时依然保持了比较高的检测率。由此可见，本文所采用的分层分类器方法是有效的。

表 3-2 列出了不同方法在 CMU+MIT 测试集合上的实验结果。本文方法取得了比[1,21,33]更好的性能，在相同误检人脸数目的情况下，获得了更高的检测率。本文方法性能接近于[30]中目前最好的性能。与[30]中方法速度慢的特点相比，本文方法具有速度快的特点。

表 3-2 不同方法在 CMU+MIT 测试集的人脸检测率的对比

Table 3-2 Comparison of different frontal face detection methods: detection rates on CMU+MIT set

误检人脸数目 方法	31	65	167
Viola 和 Jones [21]	85.2%	92.0%	93.9%
Rowley et al. [1]	85.0%	N/A	90.1%
Schneiderman 和 Kanade [30]	N/A	94.4%	N/A
Li 和 Zhang [33]	89.2%	N/A	N/A
本文方法	90.7%	92.3%	94.2%

本文方法可以检测复杂背景下的各种正面人脸。图 3-8 显示了一些检测实例。这些图像中含有不同尺寸，不同背景下的正面人脸。



图 3-8 正面人脸检测的示例

Figure 3-8 Some examples of frontal face detection

### 3.5.1.2 在 CAS\_PEAL 人脸库上的测试结果

CAS\_PEAL 库是由中科院计算所的人脸技术研究者专门针对人脸识别而自行创建的一个大规模中国人脸图像数据库[108]。CAS\_PEAL 库的所有图像在专门的采集环境中拍摄，涵盖了姿态、表情、简单饰物和光照 4 种变化，部分人脸图像具有背景、距离和时间跨度的变化。为了方便人脸识别领

域的研究者使用 CAS\_PEAL 人脸库，数据采集单位提供了 CAS\_PEAL 人脸库的共享版本 CAS\_PEAL\_R1，其中包含 1040 人的 30900 幅图像。

在人脸检测领域中，表情变化、光照变化和背景变化这几种情况是影响检测性能的常见而又复杂的因素。为了检验本文方法的鲁棒性，我们选取 CAS\_PEAL\_R1 正面人脸图像子库中的部分子集进行正面人脸检测的测试。具体而言，选取了 4 个子集进行测试，包括标准集合、表情集合、光照集合和背景集合。表 3-3 给出了这几个子集图像情况的详细介绍。

表 3-3 用于正面人脸检测测试的 CAS\_PEAL\_R1 子集的图像情况

Table 3-3 CAS\_PEAL\_R1 subsets used for frontal face detection testing

测试子集	变化种类	人数	图像/人脸数目
标准集合	中性表情、环境光源	1040	1040
表情集合	环境光条件下的 5 种表情：笑、皱眉、惊讶、闭眼、张嘴等	377	1884
光照集合	9 种以上变化的光照模式	233	2450
背景集合	2~4 种不同的采集背景	297	650

图 3-9 给出了这几个集合上的正面人脸检测结果实例图像。表 3-4 显示了本文方法在这几个测试子集上的检测结果。在标准集合的中性表情和环境光源的条件下，正面人脸检测性能达到了 97.0% 的检测率和 94.2% 的检测精度。其它 3 个集合由于表情、光照和背景的变化，检测性能有所下降，但是仍然取得了较好的结果，获得了 93.5% 以上的检测率和 90% 以上的检测精度。这些测试结果表明，本文方法对于表情、光照和背景等复杂变化条件下的人脸检测具有良好的鲁棒性。

表 3-4 CAS\_PEAL\_R1 子集上的正面人脸检测测试结果

Table 3-4 Face detection testing results in CAS\_PEAL\_R1 subsets

测试子集	图像/人脸数目	正确检测到的人脸数目	误检的人脸数目	正确检测率	检测精度
标准集合	1040	1009	62	97.0%	94.2%
表情集合	1884	1804	152	95.7%	92.2%
光照集合	2450	2328	213	95.0%	91.6%
背景集合	650	608	55	93.5%	91.7%

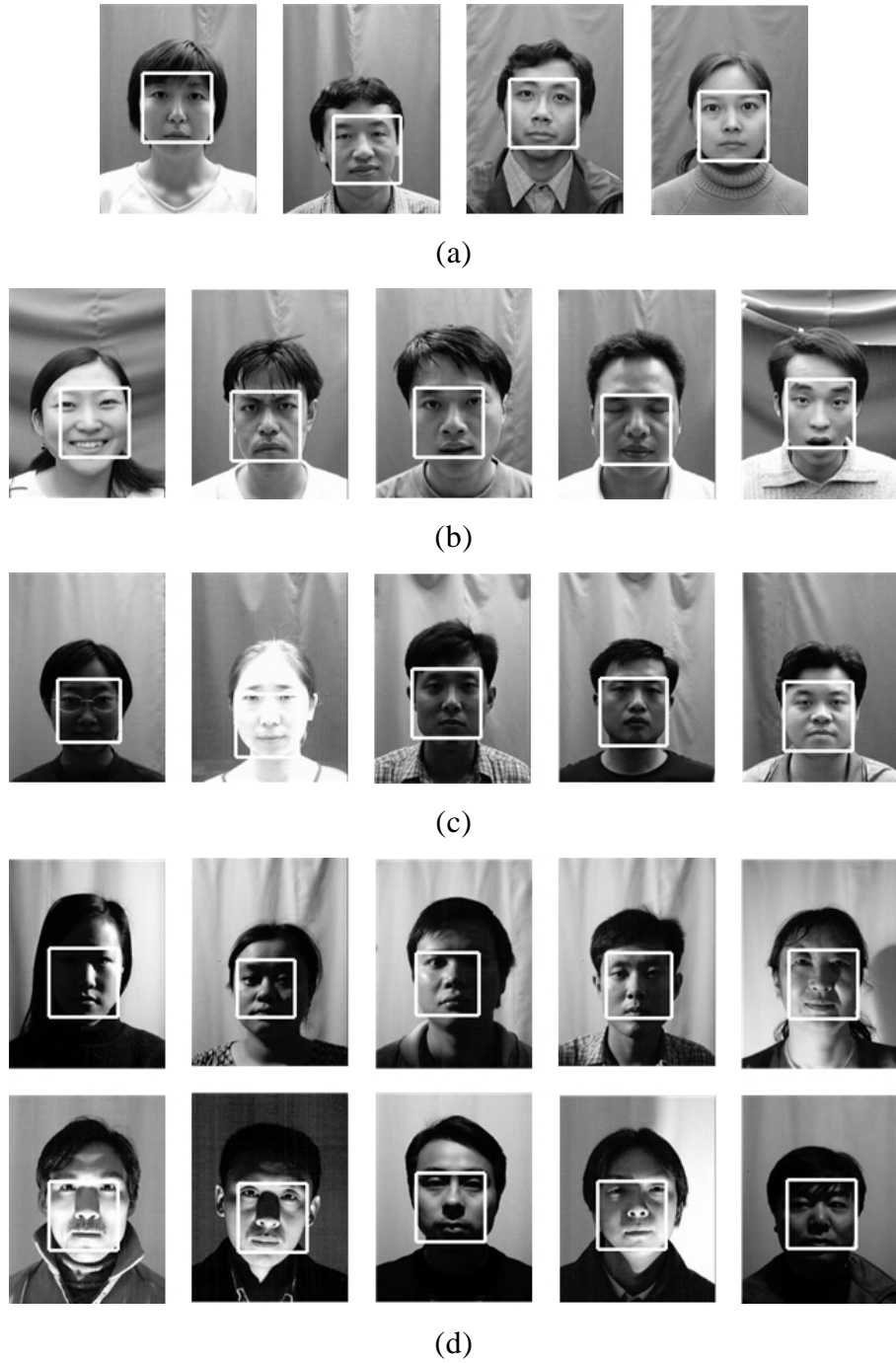


图 3-9 CAS\_PEAL\_R1 子集上的正面人脸检测实例：(a)标准集合， (b)表情集合， (c)背景集合， (d) 光照集合

Figure 3-9 Some examples of frontal face detection in CAS\_PEAL\_R1 subsets: (a) Normal subset, (b) Expression subset, (c) Background subset, (d) Lighting subset



### 3.5.2 侧面汽车检测

从外观上看，侧面汽车是由几个明显的部件构成，包括车身、车轮、车门和车窗等，这些部件按照相对固定的空间结构关系排列。和人脸相比，侧面汽车在空间结构关系上变化更多，其原因在于汽车品牌设计风格迥异。

我们建立了一个包含幅 2725 汽车图像和 14968 幅非汽车图像的训练样本集合，每个样本的尺寸是 100x40。其中 500 个汽车样本来自 UIUC Image Database for Car Detection[83]的训练集合，其它汽车样本采自于自然图像、网页和视频中。另外还收集了 1225 幅汽车图像和 7495 幅非汽车图像作为校验和测试集合。200 幅不含汽车的自然图像用于产生反例集合。图 3-10 中给出了一些侧面汽车的训练样本示例。



图 3-10 侧面汽车检测的训练样本示例

Figure 3-10 Some training examples for side-view car detection

在 100x40 的图像窗口中定义的空间直方图特征集合的基数很大，大约是 3594591，此集合对于物体检测而言过于冗余。消去冗余模板之后，我们在联合直方图匹配训练过程中把可接受的检测率  $D$  设为 0.95，经过训练得到 15 个空间直方图特征（如图 3-11 所示）用于汽车检测的联合直方图匹配。为了提高检测精度，使用基于类别可分性和特征相关性的特征选择算法进行训练，得到 25 个空间直方图特征用于汽车检测的支持向量机分类器。支持向量机分类器采用径向基核 RBF 函数。

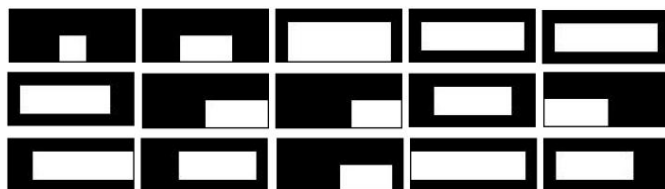


图 3-11 侧面汽车检测中为联合直方图匹配选取的 15 个空间模板

Figure 3-11 15 Selected spatial templates for histogram matching in side-view car detection

本文针对 UIUC Image Database for Car Detection [83]中的测试集合进行性能比较，这个集合分为两个子集。第一个子集 (测试集 A)有 170 幅图像，包含 200 辆侧面汽车，这些汽车尺寸基本上固定在  $100 \times 40$  左右。第二个子集 (测试集 B) 有 108 幅图像，包含 139 辆大小各异的侧面汽车。这两个子集的图像中包含部分被遮挡的汽车，还有一些汽车灰度分布接近于背景，因此对其进行侧面汽车检测，具有很大难度。

为了对算法性能做一个整体了解，我们通过逐步改变联合直方图匹配和支持向量机的分类阈值，在测试集 A 和测试集 B 上分别运行侧面汽车检测算法，从而获得相对应的检测率和误检率。以下列出了在测试集 A 和测试集 B 上的 ROC 曲线，如图 3-12 和 3-13 所示。图 3-14 和 3-15 显示了算法在测试集 A 和测试集 B 上的 RPC 曲线。

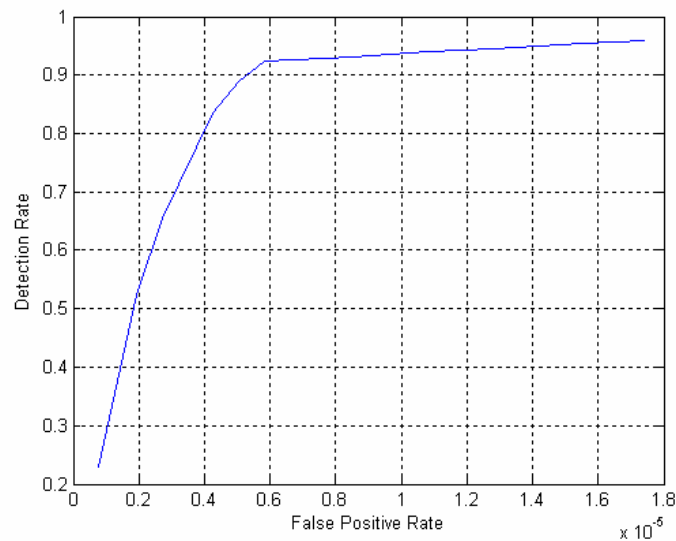


图 3-12 UIUC 测试集 A 上的 ROC 曲线

Figure 3-12 ROC obtained on UIUC car detection test set A

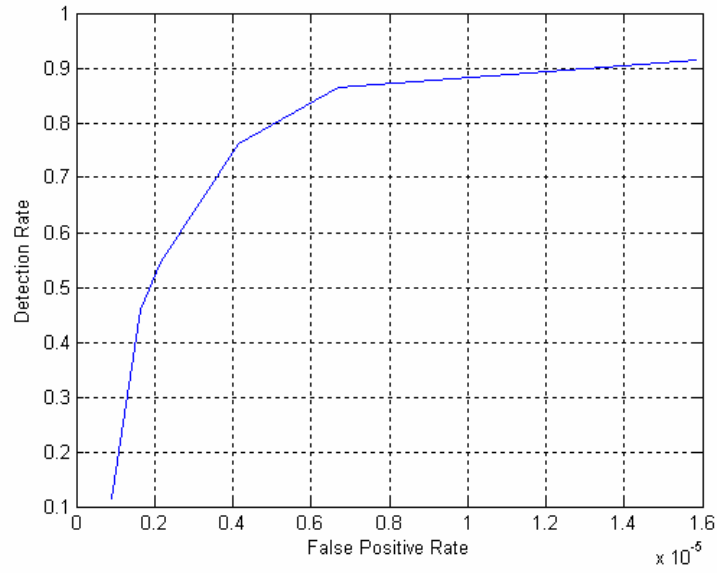


图 3-13 UIUC 测试集 B 上的 ROC 曲线

Figure 3-13 ROC obtained on UIUC car detection test set B

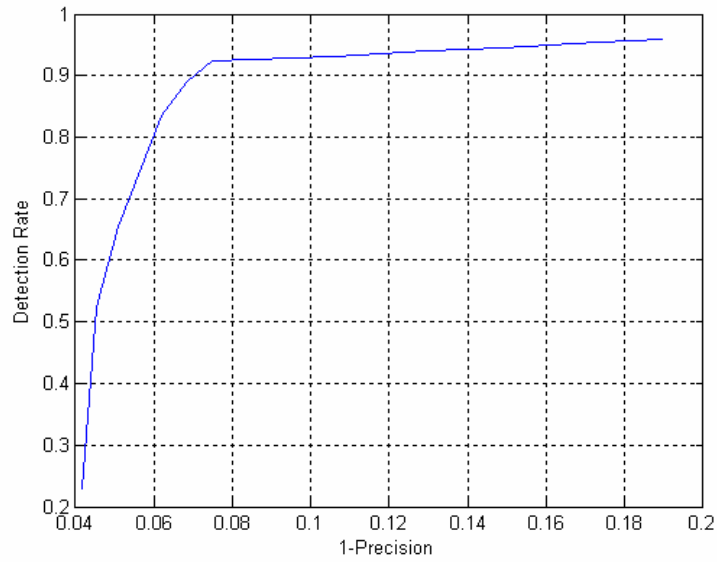


图 3-14 UIUC 测试集 A 上的 RPC 曲线

Figure 3-14 RPC obtained on UIUC car detection test set A

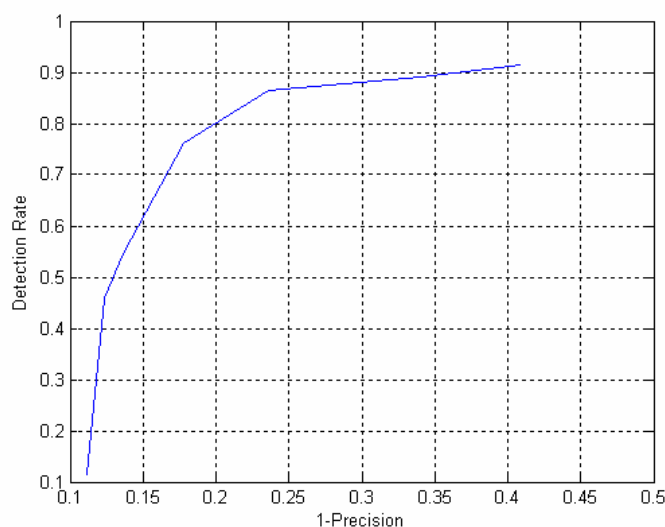


图 3-15 UIUC 测试集 B 上的 RPC 曲线

Figure 3-15 RPC obtained on UIUC car detection test set B

与人脸检测实验类似，本文采用了两套实验方案在 UIUC 测试集 A 上进行测试，从而分析层次分类器中不同步骤对系统性能的影响。第一套方案只利用联合直方图匹配方法进行检测；第二套方案利用联合直方图匹配方法和支持向量机分类器组成的分层分类器进行检测。可以通过改变联合直方图匹配的阈值，获得不同参数条件下的测试结果。表 3-5 显示了这两种方案的测试结果。其中参数条件 1 中的联合直方图匹配的阈值比较大，所以获得检测率比较低；参数条件 2 中的联合直方图匹配的阈值比较小，所以获得检测率比较高。

表 3-5 侧面汽车检测算法不同阶段在 UIUC 测试集 A 上的测试结果

Table 3-5 Results in UIUC set A by different phases of side-view car detection

		正确检测 的数目	错误检测 的数目	检测率	检测 精度
参数 条件 1	未使用 SVM	164	187	82.0%	46.7%
	使用 SVM	158	11	79.0%	93.4%
参数 条件 2	未使用 SVM	196	441	98.0%	30.7%
	使用 SVM	193	45	96.5%	81.1%

表 3-6 列出了本文方法和[41]中的方法的测试结果。从中可以看出，本文方法比[41]中的方法得到了更好的性能，检测率高，而且误检率低。

表 3-6 汽车检测方法的性能比较

Table 3-6 Car detection results comparison

	测试集 A		测试集 B	
	[41]中的方法	本文方法	[41]中的方法	本文方法
侧面汽车数目, $T$	200	200	139	139
正确检测的数目, $TP$	183	193	112	120
错误检测的数目, $FP$	557	45	1216	37
检测率, $TP/T$	91.50%	96.50%	80.58%	86.33%
检测精度, $TP/(TP+FP)$	24.73%	81.10%	8.43%	76.43%

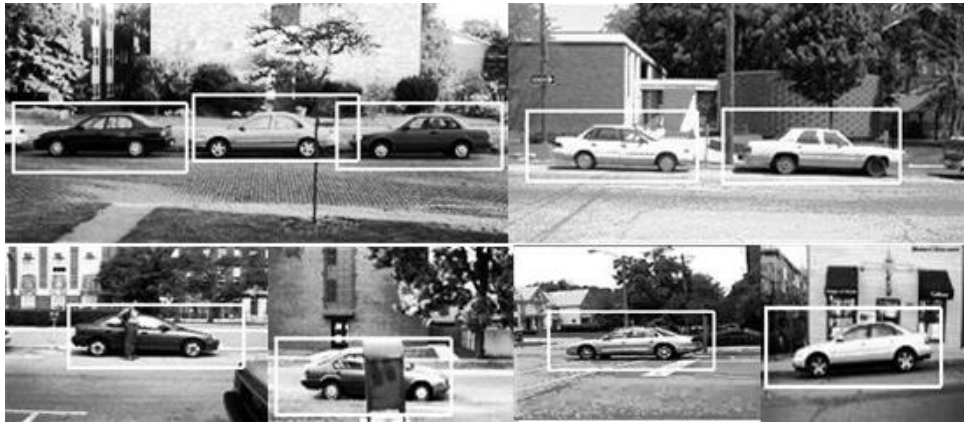
除了[41]中的方法和本文方法，很多算法只在测试集 A 上进行了性能测试。表 3-7 列出了几种方法在 UIUC 测试集 A 上 RPC 等错误率的对比情况。本文方法取得了 92.5%的 RPC 等错误率，高于[41]、[42]和[43]。[44]中方法的设计目标是检测单一尺度侧面汽车，所以它获得了 97.5%的 RPC 等错误率。尽管如此，本文取得了略低于[44]中方法的 RPC 等错误率，而且可以检测多尺度侧面汽车。

表 3-7 不同方法在 UIUC 测试集 A 上等错误率的对比情况

Table 3-7 Comparison of RPC equal error rates on UIUC set A

	[41]中的方法	[42]中的方法	[43]中的方法	[44]中的方法	本文方法
RPC 等错误率	77.5%	88.5%	91.0%	97.5%	92.5%
是否多尺度	是	是	是	否	是

本文方法可以检测复杂背景下的各种侧面汽车。图 3-16 显示了一些检测实例。这些图像中含有不同尺寸，不同背景下的侧面汽车。在图 3-17 的(a)和(b)中，一些汽车远小于 100x40，所以未检测到；(c)和(d)显示了一些误检的图像例子。



(a)



(b)



(c)

图 3-16 侧面汽车检测实例：(a)测试集 A，(b) 测试集 B，(c)一些其它数字图像  
Figure 3-16 Car detection examples on (a) UIUC test set A, (b) UIUC test set B, and (c) some other digital photos



图 3-17 侧面汽车检测结果实例：漏检的情况(a,b)和误检的情况(c,d)

Figure 3-17 Car detection results: missing detections (a, b) and false detections(c, d)

### 3.5.3 视频文字检测

文字检测是在输入图像中判断文字是否存在，并且定位其位置的过程。本文把所提出的方法应用于视频文字检测，以检验空间直方图特征对非固定空间结构类型的物体是否有效。在实验过程中，文字模式定义为尺寸为  $50 \times 20$  的图像块，并且采用空间直方图特征建立文字检测器。文字行检测分为两个步骤：第一步采用文字检测器扫描多尺度的图像金字塔，产生文字区域图；第二步利用文字行分割算法[84]在文字区域图上获取文字行。图 3-18 显示了文字训练样本的一些实例。



图 3-18 文字检测训练样本示例

Figure 3-18 Some training examples for text detection

我们建立了一个包含 2936 幅视频文字图像和 12313 幅非文字图像的训练样本集合(样本尺寸是  $50 \times 20$ )，还收集了 2012 幅视频文字图像和 6865 幅非文字图像作为训练分类器时的校验测试集合。230 幅不含文字的自然图像用于产生反例集合。经过训练，最后得到 17 个空间直方图特征（如图 3-19 所示）用于联合直方图匹配，32 个空间直方图特征用于支持向量机分类器。支持向量机分类器采用径向基 RBF 核函数。

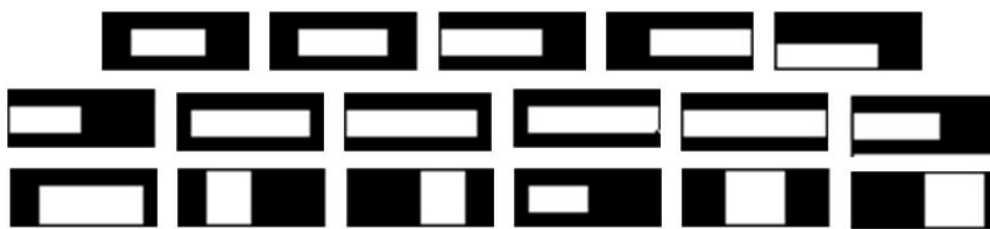


图 3-19 文字检测中为联合直方图匹配选取的 17 个空间模板

Figure 3-19 17 selected spatial templates for cascade histogram matching in text detection

本文方法在公开的Microsoft Research Asia(MSRA)视频文字图像集合上[85,86]进行了测试。在网站([http://www.cs.cityu.edu.hk/~liuwy/PE\\_VTDetect/](http://www.cs.cityu.edu.hk/~liuwy/PE_VTDetect/))上可以获这个测试数据集合。它收集了MPEG-7 数据集中的 45 帧文字图像, 总共包含 128 个人类肉眼可以辨识的文字行。

本文采用了两套实验方案在 MSRA 测试集合上进行测试, 从而分析层次分类器中不同步骤对系统性能的影响。第一套方案只利用联合直方图匹配方法进行检测; 第二套方案利用联合直方图匹配方法和支持向量机分类器组成的分层分类器进行检测。可以通过改变联合直方图匹配的阈值, 获得不同参数条件下的测试结果。表 3-8 显示了这两种方案的测试结果。其中参数条件 1 中的联合直方图匹配的阈值比较大, 所以获得检测率比较低; 参数条件 2 中的联合直方图匹配的阈值比较小, 所以获得检测率比较高。

表 3-8 文字检测算法不同阶段在 MSRA 测试集上的测试结果

Table 3-8 Results in MSRA text set by different phases of text detection

		正确检测 的数目	错误检测 的数目	检测率	检测 精度
参数 条件 1	未使用 SVM	78	112	60.9%	41.0%
	使用 SVM	72	2	56.3%	97.2%
参数 条件 2	未使用 SVM	119	237	93.0%	33.4%
	使用 SVM	114	5	89.0%	95.7%

为了对算法性能做一个整体了解, 我们通过逐步改变联合直方图匹配和支持向量机的分类阈值, 在测试集上运行文字检测算法, 从而获得相对应的检测率和误检率。以下列出了在 MSRA 测试集上的 ROC 曲线和 RPC 曲线, 分别如图 3-20 和图 3-21 所示。



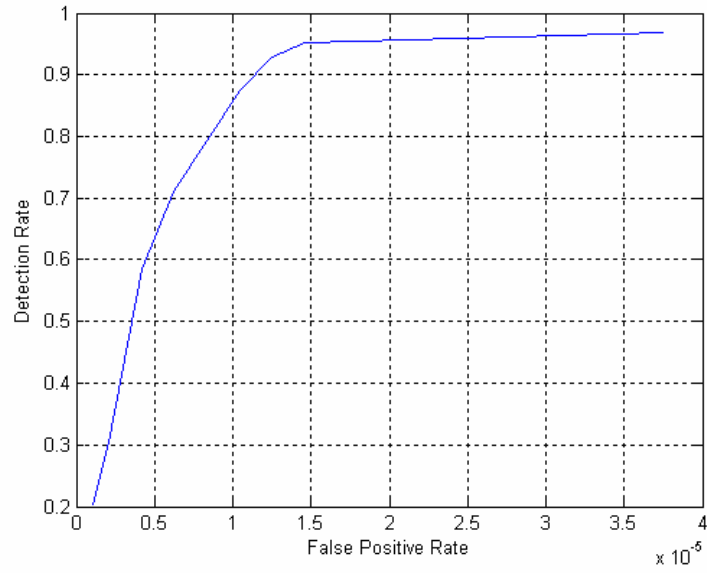


图 3-20 文字检测算法在 MSRA 测试集合上的 ROC 曲线

Figure 3-20 ROC obtained on MSRA text detection test set

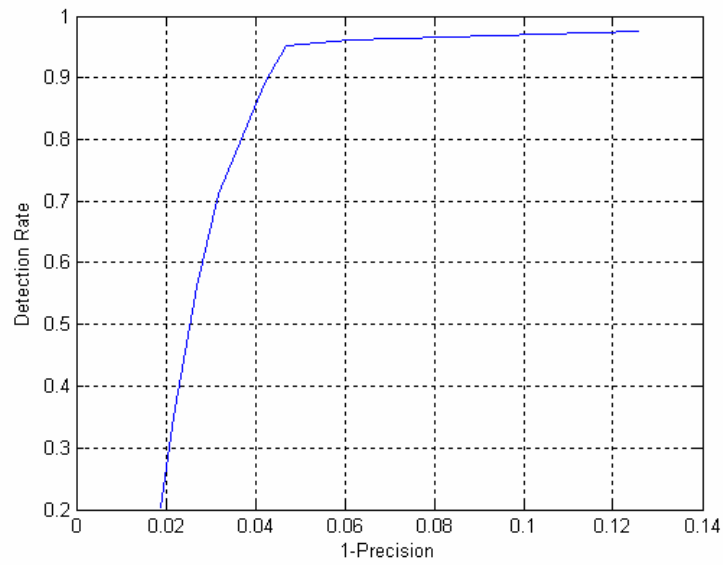


图 3-21 文字检测算法在 MSRA 测试集合上的 RPC 曲线

Figure 3-21 RPC obtained on MSRA text detection test set

表 3-9 列举了在本文方法这个集合上的实验结果，正确检测到了 116 个文字行，误检文字行有 16 个，取得了 95.3% 的检测率和 95.3% 的检测精度。和文献[86]中的方法相比，本文方法在检测率和检测精度上都取得了更优的结果。图 3-22 给出了文字检测的实例。实验结果表明空间直方图特征对于文字检测是有效的，本文方法可以检测复杂背景下的各种文字。

表 3-9 视频文字检测算法在 MSRA 集合上的测试结果对比

Table 3-9 Comparison of text detection results on MSRA test set

	本文方法	Hua et al. [86]
文字行数目, $T$	128	128
正确检测到的文字行数目, $TP$	122	117
错误检测到的文字行数目, $FP$	16	6
检测率, $TP/T$	95.3%	91.4%
检测精度, $TP/(TP+FP)$	95.3%	95.1%

### 3.5.4 关于系统检测速度的实验

除了评价检测率和检测精度等指标之外，系统检测速度对于物体检测算法的研究也极为重要。尤其在一些实际的实时应用系统中，例如视频监控等，系统速度是决定使用何种算法的关键因素。

本文综合上述技术，实现了基于空间直方图特征的物体检测系统。该系统的设计目标是提供物体检测的核心算法和框架技术，不限制目标物体类别。目前，该系统能够处理正面人脸检测、汽车检测和文字检测三种任务。采用的运算平台是 Pentium 4 型号的 3.2GHz CPU，系统的执行速度纪录如表 3-10 所示。从中可以看出，基于空间直方图特征的物体检测系统能够适用于实时应用。

表 3-10 物体检测系统的执行速度

Table 3-10 Performance time of our object detection systems

检测任务	图像大小 (宽 x 高)	单幅图像的平均检测时间(ms)
正面人脸检测	320x240 pixels	100
汽车检测	320x240 pixels	120
文字检测	320x240 pixels	220



(a)



(b)

图 3-22 文字检测的实验结果：(a)MSRA 测试集和(b)一些其它图像

Figure 3-22 Some examples of text detection: (a) MSRA testing set, and (b) some other video frames

### 3.6 本章小结

本文利用联合直方图匹配和支持向量机等技术，提出了基于空间直方图特征的物体检测方法。该方法有效的解决了复杂背景下的物体检测问题，具有快速、鲁棒的特点。

与其它大多数物体检测方法相比，本文方法具有更好的可扩展性。基于空间直方图特征的物体检测方法不但可以处理具有显著空间结构关系的物体模式，比如侧面汽车和人脸；而且可以处理不具备固定空间结构关系的物体模式，比如视频文字。这表明空间直方图特征的表示方法能够适应不同种类的物体类型。

实验结果表明，通过基于类别可分性和特征相关性的特征选择方法，可以自动获取与目标物体类别相关联的分类特征。该方法适用于不同类型物体类别的特征选择，它不但可以选择出有效的分类特征，而且在保证分类性能的条件下降低了特征维数，提高了分类效率。

综合以上因素，可以看出空间直方图特征具有推广到表示更多种类型物体的潜力。因此，可以进一步对基于空间直方图特征的物体检测方法进行扩展，本文将在下一章探讨这方面的研究工作。

## 第4章 基于纠错码多类分类器的多姿态人脸检测

### 4.1 引言

在物体检测领域中，人脸检测是一个典型问题，它的解决方案对物体检测领域具有推动作用和借鉴意义。近年来，人脸检测研究领域取得了很大的进展，但是仍然面临着巨大挑战。究其原因人脸受诸多因素影响，模式变化复杂。这些因素主要包括复杂的成像背景和不同的光照条件、人脸的姿态变化、以及人脸的特殊状态(胡须、饰物和遮挡)等方面。类似这样的难题在大多数三维物体检测研究中同样存在。因此，人脸检测中很多问题的研究都有助于物体检测领域中关键技术的突破。

第3章讨论了基于空间直方图特征的物体检测方法，并且采用这种方法探讨了正面人脸检测问题的解决方案。在实际应用中，完善的人脸检测模型必须能够检测多姿态的人脸。本章针对人脸检测的姿态问题进行研究，不但提供多姿态人脸检测的新方法，而且在理论上为物体检测领域中的多类目标物体检测方面拓展研究广度。

本文采用基于纠错码的多类分类器，提出求解多姿态人脸检测问题的方法。该方法以纠错码多类分类器为框架，将基于空间直方图特征的物体检测算法推广至多姿态人脸检测。此外，为了构建更有效的纠错码多类分类器，本文提出了以最小化分类错误率为目标的联合训练算法。本文以深度旋转的人脸为实例，对所提出的基于纠错码多类分类器的多姿态人脸检测方法进行实验验证。

本章后续部分组织如下。第2节简单介绍了多姿态人脸检测方面的相关研究工作。第3节介绍基于纠错码的多类分类方法。第4节提出了结合空间直方图特征和纠错码多类分类方法的多姿态人脸检测框架。第5节提出了根据给定码本如何构建有效纠错码多类分类器的方法。第6节对所提出的多姿态人脸检测算法进行了实验评估。最后是本章小结。

### 4.2 相关研究工作介绍

假定摄像机镜头位置固定，那么多姿态人脸是由正面人脸作深度旋转和平面旋转而形成的，姿态空间由三个旋转方向表示，分别是上下旋转、左右

旋转和平面旋转，如图 4-1 中的示例图像所示。



图 4-1 人脸的多姿态： (a)正面 (b)上下旋转 (c)左右旋转 (d)平面旋转

Figure 4-1 Multi-view faces: (a) front view (b) pitch view (c) profile view (d) in-plane rotation

综合有关文献，目前的多姿态人脸检测解决方案主要有四种途径。

第一种是基于单种姿态检测器的方法。在训练过程中，首先建立离散姿态的人脸样本集，然后对每种姿态人脸都训练形成单独的检测器。在测试阶段中，首先把每种姿态的人脸检测器都应用于输入图像，得到各自的检测结果，然后将这些结果融合输出最后的多姿态人脸检测结果[97]。还有一种基于姿态估计策略的测试方法，即首先采用姿态估计器得到测试窗口的大致姿

态，然后通过相对应的某一类的单种姿态人脸检测器输出检测结果[98,99,100]。

第二种策略是把测试样本校正为正面向上姿态的人脸标准图像样本，通过正面人脸检测技术判断其是否为人脸。Rowley 等提出的平面旋转人脸检测器[101]属于这类方法。他们以正面向上人脸为标准图像样本，采用神经网络计算得到测试样本的旋转角度，并将其校正为标准图像样本作为正面人脸检测器的输入，来判断它是否为人脸。

第三种途径描述随姿态变化的人脸分布，以解决多姿态人脸检测问题。[102]采用平滑函数刻画人脸的姿态变化模式。[103]采用基于姿态空间为参数的流形表示人脸分布。这些方法实质上试图建立人脸姿态的参数化模型，从而把人脸检测和姿态估计在同一框架下进行求解，但是进展不太顺利。

第四种方法是在实际应用中采用的一种工程手段，尤其用于检测平面旋转人脸时比较有效。具体而言，即在正面人脸检测技术的基础上，将测试图像按一定的角度顺序依次旋转后再检测其中的正面人脸[104]。

与以上四种方法不同，本文提出一种新的解决思路。简而言之，就是将多姿态人脸检测看作为一个多类分类问题，通过多类分类算法来直接求解。这样做的动机是在人脸姿态空间中，各种姿态样本可以被认为是不同的类别；其中每类的类内具有相似性，而且各类之间具有差异性。

单个姿态的人脸检测（例如：正面人脸检测）属于单类目标物体检测的研究范围，而多姿态人脸检测是一种多类目标物体检测任务。与单类目标物体检测可以归结为二类模式分类问题不同，多类目标物体检测是更加复杂的多类别模式分类问题。

目前，解决多类分类的方法主要有三种类型。(1)直接的多类分类算法，例如决策树 C4.5、人工神经网络等。(2)将两类分类算法直接扩展至多类问题，每一类都有一个分类器来区分它本身和其它类，例如 one-to-one, one-against-others 等。(3)把多类问题通过编码方式转化为两类问题，例如纠错码的方式(Error-Correcting Output Codes, ECOC)。

### 4.3 基于纠错码的多类分类方法

基于纠错码的多类分类方法是信息论原理与机器学习理论交叉而形成的一种新方法。因此，本节简单概述相关的通信技术研究背景知识，在此基础上介绍将纠错码概念用于多类问题的模式分类方法的原理。



### 4.3.1 信息传输系统模型与纠错码

本小节仅对纠错码的基本理论进行简要说明，关于纠错码的详细论述，请参阅文献[87,88,89,90,91]。

提高信息传输的可靠性和有效性，是通信技术研究一直所追求的目标。信道编码是提高信息传输可靠性的一种重要手段，已经历了数十年的发展，在理论和应用中都取得了很大的进展。

香农(Shannon)在 1948 年提出了著名的信道编码定理，该定理指出：对于一个给定的有噪声的信道，只要发送端以低于信道容量的速率发送消息，则一定存在一种编码方法，使得译码错误率随着码长的增加按指数下降到任意小的值。根据这个定理，可以知道：如果采用信道编码技术，那么通过不可靠的信道可以实现可靠的信息传输。

信道编码定理为提高数据传输的可靠性奠定了理论基础，虽然定理本身并没有给出如何去构造具体的编码和译码的方法，但是为信道编码的发展指出了前进的方向。经过半个多世纪的努力，目前已经获得许多有效的编译码方法，并形成了一个新的技术：纠错码技术。纠错码技术就是信道编码技术，通过提高信息传输时的抗干扰能力增加信息传输可靠性。本文将不加区分的使用纠错码和信道编码这两个术语。

以纠错码为视角，信息传输系统可以简化为如图 4-2 所示的模型。在此模型中，发送端的信源输出消息  $u$ ，经过纠错码编码器编成码字  $x$ ，形成信号输入信道，由于信道传输中的噪声干扰，信道输出的信号  $y$  可能发生错误，经过纠错码译码器对  $y$  中的错误进行纠正，作为原始消息的恢复值  $\hat{u}$  由信宿接收。

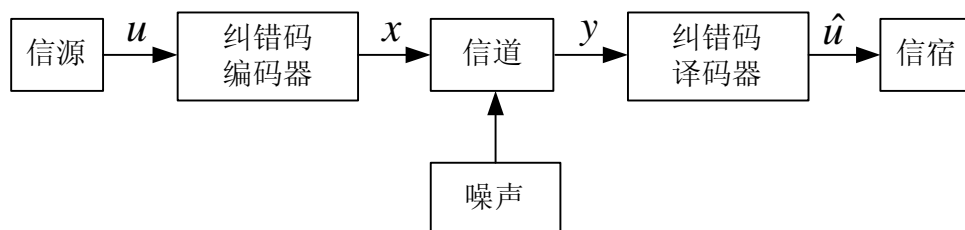


图 4-2 信息传输系统的简化模型

Figure 4-2 Model of information transmission



纠错码的基本思想是引入冗余度，即发送端在传输的信息码元后增加一些多余码元，通过检验和纠正信息传输过程的错误，以使信息损失或错误后仍能在接受端正确的恢复，从而提高通信可靠性[87]。常用的纠错码有汉明码、线性分组码、循环码和卷积码等。

目前，大多数信息系统中广泛采用二进制形式的码。表 4-1 给出了一个 (7,3)线性分组码的例子。该例子中的码本包括 8 个码字，每一个码字由 7 位二进制数表示，其中头 3 位表示信息元，后 4 位表示校验元。因此这个线性分组码可以用于  $2^3=8$  种消息的传输。

表 4-1 纠错码的一个例子

Table 4-1 An example of error correct code

信息组	码字
000	0000000
001	0011101
010	0100111
011	0111100
100	1001110
101	1010011
110	1101001
111	1110100

为了描述纠错码的性能，需要引入汉明(Hamming)距离的概念。两个码字  $x$ ， $y$  之间，对应位取值不同的个数，称为它们之间的汉明距离，用  $d(x,y)$  表示。例如， $x=0010000$ ， $y=0011101$ ，则  $d(x,y)=3$ 。在一个码本中，任两个码字之间汉明距离的最小值，称为该码本的最小汉明距离，用  $d_0$  表示。根据纠错码的基本理论，一个最小汉明距离为  $d$  的码，至多能纠正

个  $t = \left\lfloor \frac{d-1}{2} \right\rfloor$  ( $\lfloor a \rfloor$  是  $a$  的整数部分) 个错误。例如，表 4-1 所给出码本的最小汉明距离  $d_0=4$ ，它可以纠正 1 位错误。

### 4.3.2 将纠错码应用于多类分类方法

在信息论中，信道是指信息传输的通道。实际应用中通常所利用的信息传输通道包括通信中的各种物理通道（如电缆、光纤、微波线路等），以及磁带、光盘等信息存取设备。信道概念的关键是有一个输入以及一个与输入有关的输出，信道可以抽象的用输入与输出信号之间关系来表示。因此，任何为了某种目的而使信息不得不经过的通道都可以被看作为信道，例如一个分类器的输入到它的输出就可以看作是一个信道。

Dietterich 和 Bakiri[92,93]把机器学习中的分类器和信息传输模型进行类比分析。待测试样本的类别号通过分类器这种信道进行传输后，输出预测结果的过程，相当于信息传输过程。在此过程中，存在着诸多由于样本集合、特征选择等带来的噪声源，所以类别信息发生传输错误。因此，把纠错码理论引入机器学习领域，有可能提高分类器的可靠性。

在此分析的基础上，Dietterich 和 Bakiri 首先提出了将多类分类问题通过纠错码方法分解成多个两类模式分类问题，构建成纠错码的码本，并以此为基础获得多个单独的两类分类器；采用纠错码译码器来纠正两类分类器单独发生的错误。很多实验表明[94,95,96]，这个框架可以增加多类分类问题解的可靠性，提高分类性能。

假设  $S = \{(x_1, y_1), \dots, (x_m, y_m)\}$  是一个多类问题的训练集，其中每个样本来自特征向量集合  $X$ ，即  $x_i \in X$ ，样本类别号  $y_i$  不局限于二值集合，而是取值于离散值类别集合  $Y = \{0, \dots, k-1\}$ （ $k$  是类别数目），即  $y_i \in Y$ 。多类别的模式分类任务就是寻找一个函数  $H: X \rightarrow Y$ ，将任意样本  $x$  映射到类别集合中的一个标号  $y$ ， $y \in Y$ 。

在纠错码框架中，解决多类分类问题的第一步是分解成多个两类模式分类问题，构建成纠错码的码本。一般的做法是根据类别数目，采用相应纠错码的码本。考虑一个取值为 0 和 1 的码本矩阵  $Z$ ，其大小为  $k \times n$ ， $k$  是类别数目。码本矩阵的每一行对应一个类别，分配为一个长度为  $n$  的二值(0 和 1)符号串，称之为码字。码本矩阵的每一列根据列中的二值符号，将样本集合重新划分成两个新类别，称之为超类。

基于纠错码的多类分类方法的过程包括两个步骤。在第一个步骤中，针

对每一列码本训练一个两类分类器来区分所对应的超类，从而形成了 $n$ 个两类分类器，这些两类分类器被称为单分类器。第二个步骤是样本测试过程，将待分类的新样本分别输入到 $n$ 个单分类器进行测试，将分类结果串接成为一个长度为 $n$ 的输出向量；最后，根据输出向量和码字的距离进行译码，从而判断样本的类别。通常而言，译码采用最小汉明距离原则，即寻找与输出向量的汉明距离最小的码字所对应的类别号，作为分类器的输出。

如果信道的输入输出是离散的符号，而且信道的输出只与当前时刻的输入有关，则这种信道称为离散无记忆信道。可以证明[91]：在离散无记忆信道中，最小汉明距离译码原则与最大似然译码原则等价。如果单分类器之间彼此不相关，那么基于纠错码的多类分类方法所采用的最小汉明距离原则译码就等同于最大似然分类原则。

## 4.4 基于纠错码的多姿态人脸检测

本文将第3章讨论的以空间直方图特征为基础的单类物体检测方法和基于纠错码的多类分类方法相结合，提出了基于纠错码的多姿态人脸检测方法。多姿态人脸检测过程的系统框架概括为如图4-3所示，分为两个阶段：训练阶段和检测阶段。

在训练阶段，首先采集大量的多姿态人脸图像样本的和非人脸模式的图像样本，并提取其空间直方图特征，然后构建基于空间直方图特征的纠错码分类器。其中，每个单分类器采用空间直方图特征作为输入，通过直方图匹配和支持向量机进行超类的类别判断。

和第3章中的物体检测过程类似，在输入图像中进行多姿态人脸检测的阶段也包括三个步骤：(1)构建图像金字塔、(2)物体分类、(3)检测结果的合并。第一步和第三步的过程类似于3.2节中论述的相同过程。不同之处在于第二步中这里采用的是纠错码多类分类器，能够判断扫描窗口是否为物体目标，并且确定其属于哪个目标物体类别。

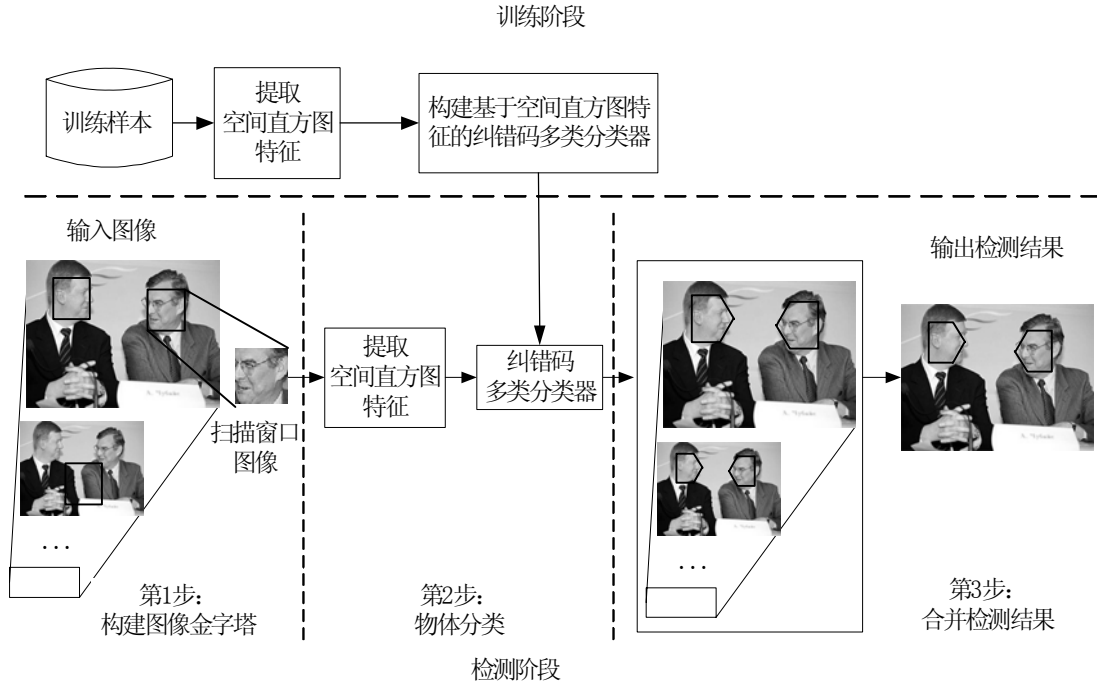


图 4-3 多姿态人脸检测的系统框架

Figure 4-3 System overview of multi-view face detection

针对多姿态人脸检测问题，假设要检测  $k-1$  类目标的物体，码字长度为  $n$ ，那么加上非目标类，总计为  $k$  类。因此首先要构造一个  $k \times n$  的码本矩阵  $Z$ ，然后在训练阶段获得  $n$  个单分类器，最后形成纠错码分类器。

在测试阶段，检测方法的过程如图 4-4 所示，其中扫描窗口图像用  $x$  表示。对于每个单分类器，在扫描窗口图像上提取空间直方图特征，然后采用直方图匹配方法和支持向量机分类器判断  $x$  是否属于相应的超类。单分类器的二值输出构成长度为  $n$  的输出向量  $V = [v_0, v_1, \dots, v_{n-1}]$ ，其中  $v_j$  是第  $j$  个单分类器的输出， $v_j \in \{0, 1\}$ ， $j = 0, 1, \dots, n-1$ ，输出向量和码字的距离采用汉明距离进行计算，定义如下：

$$L_i = \sum_{j=0}^{n-1} |Z_{ij} - v_j|, (i = 0, 1, \dots, k-1) \quad (4-1)$$

根据纠错码的译码原则， $x$  的类别是根据与输出向量距离最小的码字来决定的，即：

$$H(x) = \arg \max_i \{L_i \mid i = 0, 1, \dots, k-1\} \quad (4-2)$$

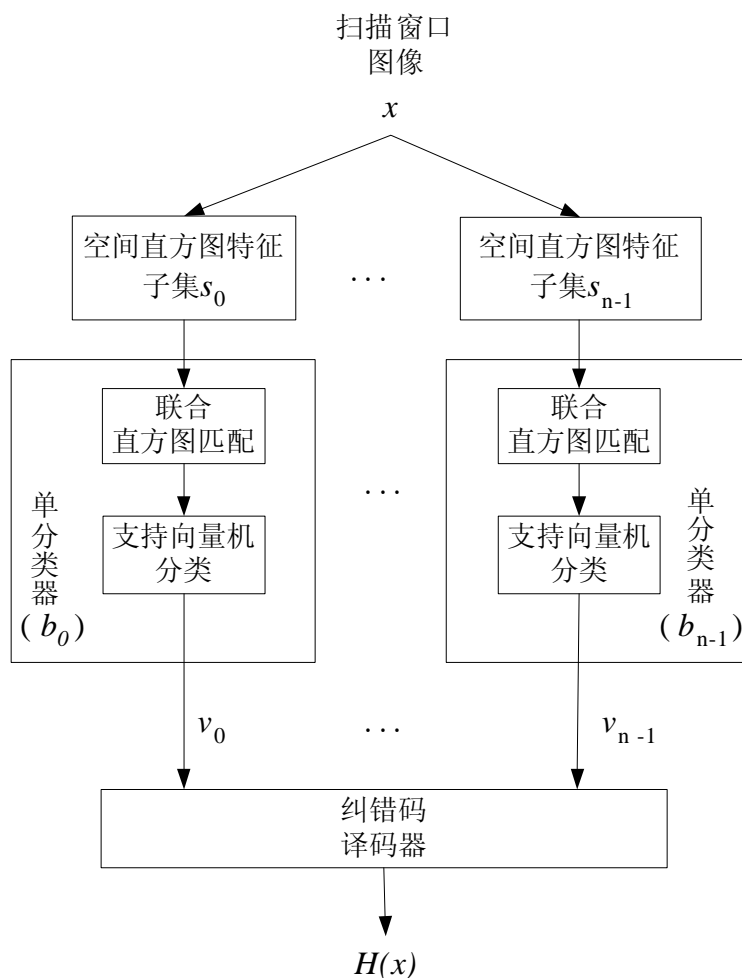


图 4-4 基于纠错码分类器的多姿态人脸检测

Figure 4-4 Architecture of the ECOC-based multi-view face detection

## 4.5 构建基于空间直方图特征的 ECOC 分类器

确定了码本矩阵和单分类器方法之后，最重要的问题是构建一个有效的 ECOC 分类器。文献中很少介绍如何建造一个有效的 ECOC 分类器。通常只是独立训练单分类器，然后直接用于 ECOC 分类器，这样往往引起单分类器训练困难的问题[105]。

本文结合空间直方图特征，提出了一个联合训练单分类器的方法。这种方法以整体 ECOC 多类分类器的错误率最小为目标，依次选择空间直方图特征和训练单分类器。

假设要检测  $k-1$  类目标的物体，并且具有以下输入条件：(1) 空间直方图特征集合  $F = \{f_1, \dots, f_m\}$ ，(2) 训练样本集合  $S = \{(x_1, y_1), \dots, (x_N, y_N)\}$  和测试样本集合  $V = \{(x'_1, y'_1), \dots, (x'_L, y'_L)\}$ ，其中  $x_i$  和  $x'_i$  是具有  $m$  维空间直方图特征的样本， $y_i \in \{0, 1, \dots, k-1\}$  和  $y'_i \in \{0, 1, \dots, k-1\}$  是样本类别标号，(3) 码本矩阵  $Z$ ，其大小为  $k \times n$ 。联合训练单分类器的过程如下所述。

(1) 采用联合直方图匹配训练方法(见 3.3.2 小节)，为码本矩阵中的每一列训练联合直方图匹配分类器，作为单分类器。这些单分类器  $\{b_0, \dots, b_{n-1}\}$  构成了初始的 ECOC 多类分类器  $H$ ；

(2) 设置分类正确率初始为 0，即  $Acc(pre) = 0$  和  $Acc(cur) = 0$ ；对于每一列，求出 Fisher 准则函数值最大的空间直方图特征  $f^i$ ，

$$F_{select}^i = \{f^i\}, \quad F_{ori}^i = F \setminus \{f^i\}, \quad i = 0, 1, \dots, n-1;$$

(3) 计算每一列单分类器的分类正确率，找到分类正确率最小的单分类器  $b_t$  ( $0 \leq t \leq n-1$ )。按照下列步骤为  $b_t$  选择特征；

(a) 对于每个空间直方图特征  $f \in F_{ori}^t$ ，在训练样本集合  $S$  上，计算

其 Fisher 准则函数  $J(f)$  和特征相关性度量  $Corr(f, F_{ori}^t)$ ；

(b) 计算特征相关性阈值  $Thre$ ：

$$\begin{cases} MinCorr = \min\{Corr(f, F_{select}^t) \mid f \in F_{ori}^t\} \\ MaxCorr = \max\{Corr(f, F_{select}^t) \mid f \in F_{ori}^t\} \\ Thre = MinCorr * (1 - \alpha) + MaxCorr * \alpha \end{cases} \quad (4-3)$$

这里， $\alpha$  是权值 ( $0 < \alpha < 1$ )，本文在实验中选取  $\alpha = 0.2$ 。

(c) 求出 Fisher 准则函数值较大的空间直方图特征  $f' \in F_{ori}^t$ ，即：

$$f' = \arg \max_{f_j} \{J(f_j) \mid Corr(f_j, F_{select}^t) \leq Thre\}$$

(d) 使用  $f'$  和  $F_{select}^t$  作为特征，在  $b_i$  所划分的超类样本集合上训练得

到一个分类器  $C$ ，这里采用支持向量机分类器；把  $b_i$  更新为联合直方图匹配和支持向量机  $C$  的串行组合，得到一个测试 ECOC 多类分类器  $H^*$ 。

(4) 计算测试样本集合  $V$  上的分类正确率，即

$$Acc(cur) = 1 - \frac{1}{L} \sum_{i=1}^L sign(H^*(x'_i), y'_i),$$

其中  $H^*(x)$  是对样本  $x$  的分类结果， $H^*(x) \in \{0, 1, \dots, k-1\}$ ，

$$sign(a, b) = \begin{cases} 1 & a \neq b \\ 0 & a = b \end{cases};$$

(5) 如果分类正确率满足条件： $Acc(cur) - Acc(pre) \geq \varepsilon$  ( $\varepsilon$  是一个小的正数)，那么执行以下步骤：

(a)  $Acc(pre) = Acc(cur)$ ， $F_{select}^t = F_{select}^t \cup \{f'\}$ ， $F_{ori}^t = F_{ori}^t \setminus \{f'\}$ ；

(b) 把 ECOC 多类分类器  $H$  更新为  $H^*$ ；

(c) 跳转至(3)，继续下一次选择特征的循环；

(6) 训练过程结束，返回特征子集  $F_{select}^i$ ，单分类器  $b_i$  ( $0 \leq i \leq n-1$ )，以及 ECOC 多类分类器  $H$ 。

构建基于空间直方图特征的 ECOC 分类器的流程如图 4-5 所示。

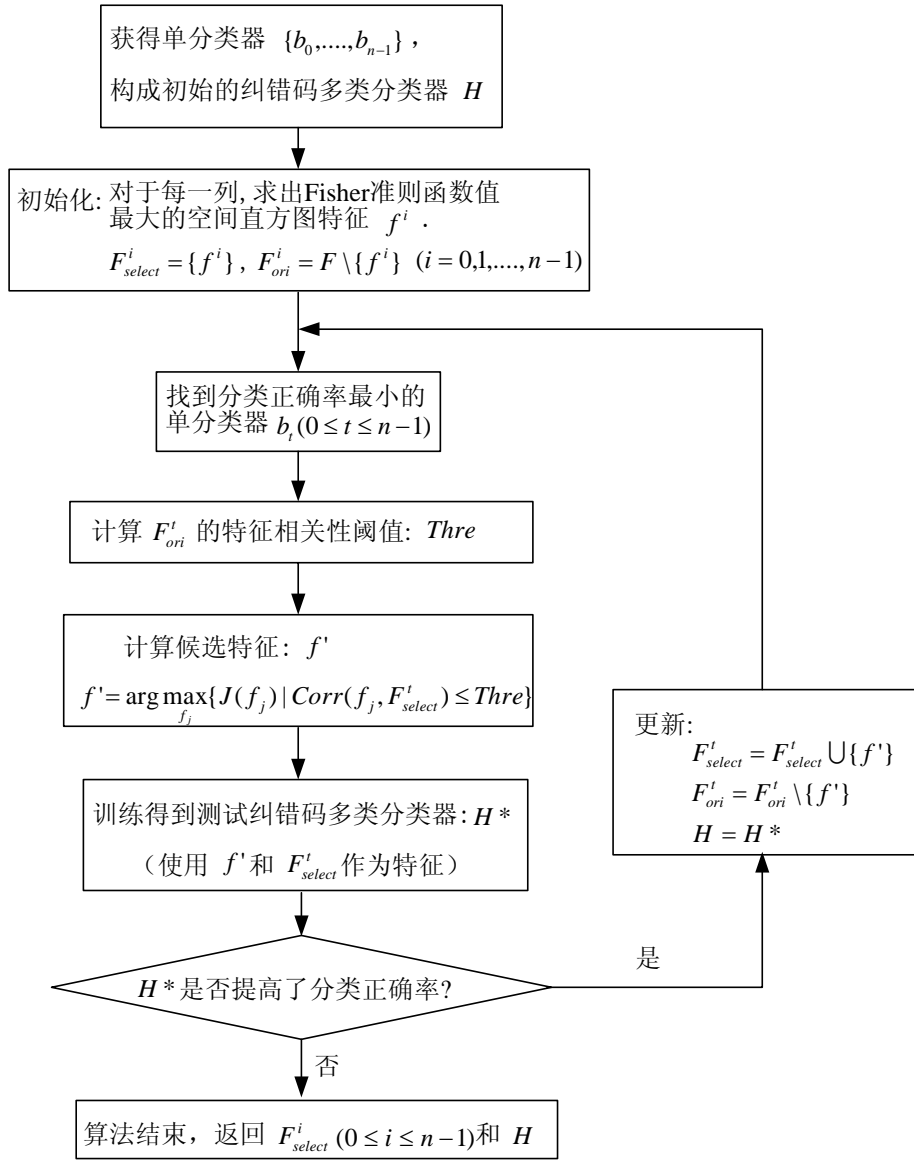


图 4-5 构建基于空间直方图特征的 ECOC 分类器的流程

Figure 4-5 Flowchart of constructing spatial histogram features based ECOC multi-class object detector

## 4.6 实验评估

本文以深度旋转人脸为例，验证基于纠错码多类分类器的多姿态人脸检测方法的有效性。本文把深度旋转的人脸划分成为左侧人脸、右侧人脸和正



面人脸等三个姿态。加上非人脸，本文要解决的多姿态人脸检测问题总共有四类目标物体。

采用纠错码对多姿态人脸检测的四类目标进行编码，码本 $Z$ 为 $4 \times 7$ 的矩阵，如表 4-2 所示。这组编码的最小汉明距离是 4，根据纠错码理论，该编码能够纠正任意一位的分类错误。

表 4-2 多姿态人脸检测 ECOC 编码的码本

Table 4-2 ECOC codebook of multi-view face detection

单分类器 类别	$b_0$	$b_1$	$b_2$	$b_3$	$b_4$	$b_5$	$b_6$
非人脸( $C_0$ )	0	0	0	0	0	0	0
正面人脸( $C_1$ )	1	1	1	1	0	0	0
左侧人脸( $C_2$ )	1	1	0	0	1	1	0
右侧人脸( $C_3$ )	1	0	1	0	1	0	1

为了训练 ECOC 多类分类器，我们采集了一个包含 11400 个正面人脸图像，4260 个左侧人脸图像，4080 个右侧人脸图像和 17285 个非人脸图像的训练样本集合(样本尺寸是  $32 \times 32$ )。还收集了大量人脸图像和非人脸图像作为训练分类器时的校验测试集合。3000 幅不含人脸的自然图像用于产生反例集合。人脸图像训练样本示例见图 4-6 所示。

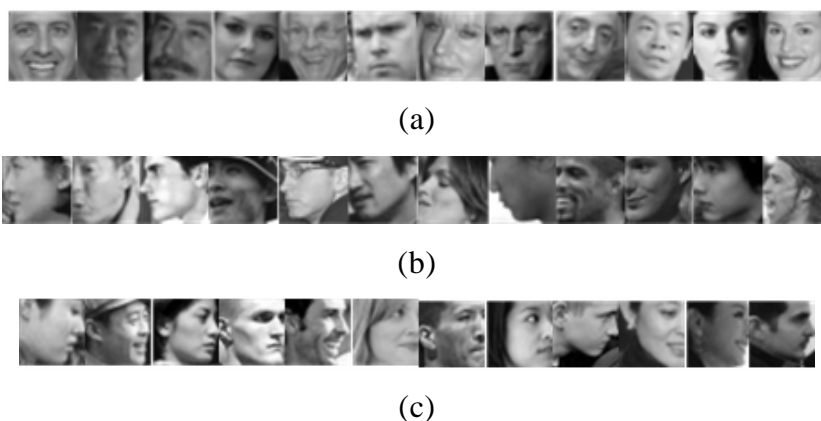


图 4-6 人脸图像训练样本示例：(a)正面人脸，(b)左侧人脸，(c)右侧人脸

Figure 4-6 Some examples of face samples: (a) front view, (b) left profile, and (c) right profile

经过训练得到了基于纠错码的多姿态人脸检测器，以及单分类器所采用的空间直方图特征子集。对于每个单分类器而言，大约 9~15 个空间直方图特征用于联合直方图匹配，大约 20~25 个空间直方图特征用于支持向量机分类器，支持向量机采用径向基核函数。多姿态人脸检测器由 7 个单分类器组成，通过汉明距离最小原则译码产生检测结果。

#### 4.6.1 纠错码性能分析

为了分析所得到的基于纠错码的多姿态人脸检测器的性能，本文收集了一个测试集合进行评价。这个集合包含 5400 个正面人脸图像、3011 个左侧人脸图像、3546 个右侧人脸图像和 6257 个非人脸图像的样本(样本尺寸是 32x32)。本文称此集合为纠错码性能测试集合。

表 4-3 报告了单分类器在纠错码性能测试集合上的分类错误率。另一方面，我们将这些单分类器进行组合，构成了基于纠错码的多姿态人脸检测器。表 4-4 给出了基于纠错码的多姿态人脸检测器在性能测试集合上的分类性能。从这两个表中，可以看出，将单分类器组合构成基于 ECOC 的多姿态人脸检测器之后，总分类错误率从 17.9% 下降到 4.0%。这个结果表明，基于 ECOC 的多姿态人脸检测器具有对单分类器的容错能力，可以从单分类器的有错误的输出中恢复正确的分类结果。

表 4-3 单分类器的分类错误率

Table 4-3 Classification error rates of the basic classifiers

单分类器	正确率	错误率
$b_0$	81.6%	18.4%
$b_1$	81.6%	18.4%
$b_2$	81.7%	18.3%
$b_3$	82.1%	17.9%
$b_4$	95.3%	4.7%
$b_5$	75.0%	25.0%
$b_6$	77.1%	22.9%
平均	82.1%	17.9%

表 4-4 基于纠错码的多姿态人脸检测器的分类性能

Table 4-4 Classification performance of the ECOC-based multi-view face detector

类别	样本数目	正确率	错误率
C <sub>0</sub>	6257	95.2%	4.8%
C <sub>1</sub>	5400	98.4%	1.6%
C <sub>2</sub>	3011	94.5%	5.5%
C <sub>3</sub>	3546	95.3%	4.7%
总计	18214	96.0%	4.0%

#### 4.6.2 在标准测试集合上的实验结果

本文以 CMU\_PROFILE 测试集[30]作为测试数据。这个集合有 208 幅图像，包含 441 个从左侧到右侧深度旋转的人脸，其中 347 个属于侧面人脸。

图 4-7 列出了本文的多姿态人脸检测系统在 CMU\_PROFILE 测试集上的 ROC 曲线。

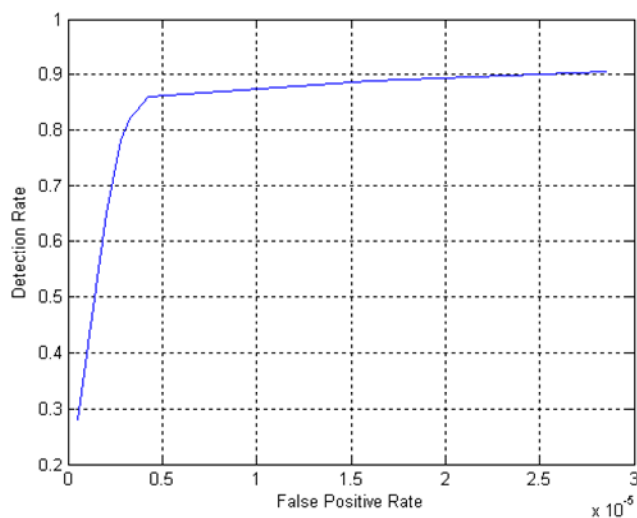


图 4-7 多姿态人脸检测实验的 ROC 曲线：CMU\_PROFILE 测试集

Figure 4-7 ROC of multi-view face detection on CMU\_PROFILE set

图 4-8 和图 4-9 显示了多姿态人脸检测的一些实例。这些实例表明空间直方图特征对于多姿态人脸检测是有效的，本文方法可以检测到复杂背景下的各种深度旋转姿态的人脸。



图 4-8 多姿态人脸检测的实例: CMU\_PROFILE 测试集

Figure 4-8 Examples of multi-view face detection on CMU\_PROFILE set



图 4-9 多姿态人脸检测的实例：其它图像

Figure 4-9 Examples of multi-view face detection on other images

表 4-5 列出了不同方法在测试集合上的实验结果。本文方法取得了比 [100,103] 更好的性能，在相同误检人脸数目的情况下，获得了更高的检测率。本文方法性能接近于 [30] 中目前最好的性能。和 [30] 中基于单类分类器的方法不同，本文采用了基于多类分类器的解决框架，可以扩展到更多姿态的人脸目标。

表 4-5 不同方法在 CMU\_PROFILE 测试集的人脸检测率

Table 4-5 Comparison of multi-view face detection rates of different methods

方法 \ 误检人脸数目	91	700
Jones 和 Viola [100]	70.0%	83.0%
Schneiderman 和 Kanade [30]	86.0%	93.0%
Osadchy et al. [103]	67.0%	83.0%
本文方法	82.0%	90.0%

#### 4.6.3 纠错码与一对多编码的比较

传统的多类分类算法有很多其它编码方式，本文选取了“一对多”(one-against-others)作为对比方法。如表 4-6 所示，针对三种姿态的人脸类别分别设计一个单独的分类器来区分其本身与其它类（包括非人脸类）。这组编码的最小汉明距离是 1，根据纠错码理论，该编码没有能力来纠正任意一位的分类错误。

表 4-6 多姿态人脸检测的一对多编码

Table 4-6 One-against-others codebook of multi-view face detection

类别 \ 单分类器	$b_0$	$b_1$	$b_2$
非人脸( $C_0$ )	0	0	0
正面人脸( $C_1$ )	1	0	0
左侧人脸( $C_2$ )	0	1	0
右侧人脸( $C_3$ )	0	0	1

为了比较基于 ECOC 的多姿态人脸检测算法与一对多编码方式多类分类算法的性能，我们实现了一个按照表 4-6 中编码方式构成的多姿态人脸检

测算法。

图 4-10 给出了两种编码方法的一些检测实例。图 4-11 报告了这两种编码方法具体的性能比较，列出了它们在 CMU\_PROFILE 测试集合上的 ROC 曲线比较结果。从中可以看出，ECOC 编码方式取得了比一对多编码更优越的性能，获得了更高的检测精度。



图 4-10 不同编码方式的多姿态人脸检测结果示例：(a) one-against-others, (b) ECOC

Figure 4-10 Examples of results with different codes for face detection: (a) one-against-others, and (b) ECOC

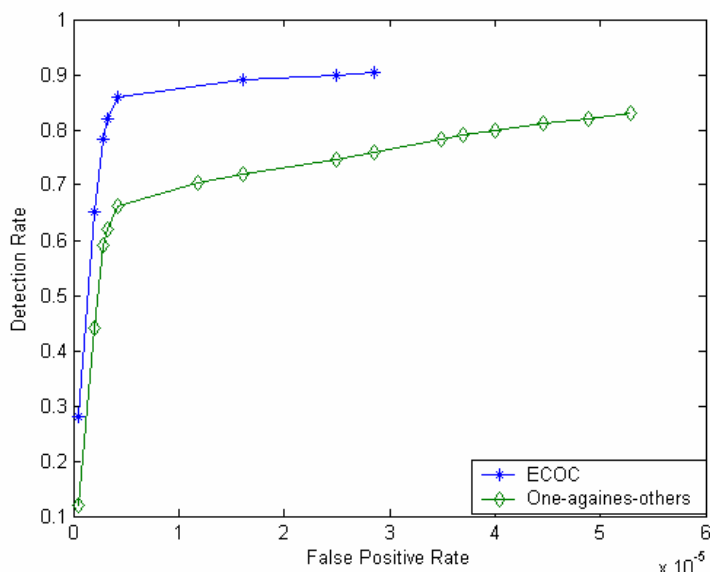


图 4-11 多姿态人脸检测算法在 CMU\_PROFILE 测试集上的 ROC 曲线比较：ECOC vs. One-against-others

Figure 4-11 Comparison of ROC curve of multi-view face detection on CMU\_PROFILE set: ECOC vs. One-against-others

#### 4.6.4 实验结论

综合以上多姿态人脸检测的实验结果，得到如下几个结论：

1. 纠错码方法可以解决深度旋转的多姿态人脸检测，在标准测试图像集合上获得了和当前最好的研究算法相媲美的结果；
2. 性能分析试验表明，纠错码方法在组合单分类器输出方面，采用汉明距离最小原则译码的方式，明显提高了分类器的正确率；
3. 与其它多类问题解决方法相比(例如一对多编码的方法)，纠错码多类分类方法取得了更好的性能。

#### 4.7 本章小结

本章采用纠错码理论，提出了一种基于纠错码多类分类器的多姿态人脸检测算法。该算法通过纠错码编码方式把多姿态人脸检测问题分解成为多个单独的二类分类问题，采用纠错码译码得到检测结果。每个单分类器利用空间直方图特征作为物体表示方式，通过直方图匹配和支持向量机来分类。本



章还提出了一种在给定纠错码码本的前提下如何训练有效纠错码多类分类器的方法，该方法以整体纠错码多类分类器的错误率最小为目标，依次选择有效特征和训练单分类器。实验结果表明，本文所提出的算法有效的解决了多姿态人脸检测问题，取得了良好的检测性能。

## 结论

由于物体检测技术在理论上的研究意义和在应用中的重要价值，它越来越受到研究者的重视，研究深度和广度在不断地增强。在计算机视觉学科中，物体检测一直是极具挑战性的理论研究课题。在实际应用中，物体检测技术对图像或视频进行模拟人类视觉感知行为的解析，在高层语义层次上为各种视觉应用提供了更有效的内容表达形式，以及更具灵活性的处理对象和处理机制。物体检测技术在多媒体内容分析与搜索、场景分析与理解、视频监控、视频分析、模式识别等领域得到了广泛的应用。

本文针对物体检测中的主要问题，在物体表示方法和物体检测算法两个层面进行了研究，主要研究目的是探索一种有效的物体特征表示方法并且应用于物体检测，从而提供物体检测的新方法。

论文的创新与主要贡献总结如下：

本文基于物体的局部纹理分析，提出了改进的空间直方图特征，用于物体检测任务中的特征表示。概括而言，本文在物体特征表示方面的主要贡献体现在三个方面：首先，以刻画物体的纹理信息和结构信息为着眼点，建立了与目标物体类别相关联的空间直方图特征提取方法；其次，对空间直方图特征进行了物体判别能力分析；最后，提出了空间直方图特征的度量方法，使用 Fisher 准则函数和互信息分别度量空间直方图特征的类别可分性和统计相关性。

提出了一种基于空间直方图特征的层次化物体检测方法。该方法采用空间直方图特征作为物体表示，并且通过由粗到精的策略，将直方图匹配和支持向量机分类器构成分层结构的分类器进行物体检测。实验结果表明该算法可以有效的解决物体检测问题，具有快速、鲁棒的特点。同时，该方法具有良好的可扩展性，可以处理典型的物体模式，包括：(1)具有显著空间结构关系的刚体物体模式，例如侧面汽车；(2)可变形的柔性非刚体结构类型三维物体，例如人脸；(3)不具备部件结构及空间关系的纹理模式类型物体，例如文字。

在特征选择方面，提出了一种结合类别可分性和特征相关性的特征选择方法，用于构建有效的分类特征。该方法采用分类器评价特征子集的性能，选择分类性能高而互相之间相关性弱的特征，构成分类特征子集。本文将该方法应用于构造紧致的空间直方图特征子集，作为物体检测中分类器算

法的输入特征。实验结果表明该方法能够自动获取与目标物体类别相关联的分类特征，它不但可以选择出有效的分类特征，而且提高了分类效率。

将基于空间直方图特征的物体检测方法进行扩展，提出了基于纠错码多类分类器的多姿态人脸检测方法。该方法把多姿态人脸检测当作一个多类分类问题，通过多类分类算法求解。首先，通过纠错码编码方式把多姿态人脸检测问题分解成为多个单独的二类分类问题；其次，每个二类分类问题利用空间直方图特征作为物体表示方式，通过直方图匹配和支持向量机来分类；最后，采用汉明距离最小原则的纠错码译码方式得到检测结果。针对基于纠错码多类分类器的训练困难的问题，提出了以最小化分类错误率为目标的联合训练算法。实验结果表明，该方法可以有效的解决多姿态人脸检测问题，取得了良好的检测性能。

总之，本文围绕物体检测进行了研究和探索，取得了一定的研究成果。主要成果包括：研究了基于局部纹理特征的物体表示方法，提出了改进的空间直方图特征；提出了基于空间直方图特征的层次化物体检测算法；提出了基于类别可分性和特征相关性的特征选择方法；将基于空间直方图特征的物体检测算法进行推广，提出了一种基于空间直方图特征和纠错码分类器的多姿态人脸检测方法；针对基于纠错码多类分类器的训练困难的问题，提出了以最小化分类错误率为目标的联合训练算法。

展望未来，物体检测领域还充满了各种挑战性的课题。除了对现有工作做进一步完善外，今后可以从以下几个方面着重进行物体检测的研究：

1. 在数据方面，需要建立更大规模、物体种类更丰富的数据库；
2. 将空间直方图特征和其它各种特征进行融合；同时更加深入的研究特征的自动提取，包括语义层面上的表示特征和判别意义下的分类特征；
3. 在物体检测方法方面，将基于学习理论的方法和基于部件的方法进行融合；同时可以考虑将物体检测和图像分割相结合，从而提高对图像场景理解的准确性；还可以结合场景中的上下文知识，提高物体分割的精度；
4. 在机器学习理论层面，可以考虑将半监督的在线学习方法和增量式的学习方法引入物体检测的研究领域。

## 参考文献

- 1 H.A.Rowley, S.Baluja, T.Kanade. Neural network-based face detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(1): 29~38
- 2 C.Garcia, M.Delakis. Convolution face finder: a neural architecture for fast and robust face detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004,26(11): 1408~1423
- 3 K.I.Kim, K.Jung, J.H.Kim. Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003,25(12): 1631~1639
- 4 M.Turk and A.Pentland. Face recognition using Eigenfaces. Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, 1991:586~591
- 5 P. N. Belhumeur, J. P. Hespanha, D. J. Kriegman. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(7): 711~720
- 6 M.Swain, D.Ballard. Color indexing. International Journal of Computer Vision, 1991, (7): 11~32
- 7 J. Huang, S. R. Kumar, M. Mitra, W.J. Zhu, R.Zabih.Spatial Color Indexing and Applications.International Journal of Computer Vision, 1999,35(3): 245~268
- 8 Q.Zhao, H.Tao. Object Tracking using Color Correlogram.The Second Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, Beijing, China, oct.15-16, 2005
- 9 B.Schiele. Object Recognition using Multidimensional Receptive Field Histograms, PhD Thesis, I.N.P.Grenoble (English translation), 1997
- 10 R.L.Hsu, M.A.Mottaleb, A.K.Jain.Face Detection in Color Images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(5): 696~706

- 11 P.F. Felzenszwalb. Learning Models for Object Recognition. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001, (1): 1056~1062
- 12 M. A. Smith, T. Kanade. Video skimming for quick browsing based on audio and image characterization. Carnegie Mellon Univ., Pittsburgh, PA, Tech. Rep. CMU-CS-95-186, July, 1995
- 13 T. Sato, T. Kanade, E. K. Hughes, M. A. Smith. Video OCR for Digital News Archives. In IEEE Workshop on Content Based Access of Image and Video Databases, 1998, Bombay:52~60
- 14 M Cai, J.Song, M.R Lyu. A New Approach for Video Text Detection. Proceedings of 2002 International Conference on Image Processing, 2002: 117~120
- 15 M.R. Lyu, J.Song, M.Cai. A Comprehensive Method for Multilingual Video Text Detection. IEEE Transactions on Circuits and Systems for Video Technology, 2005, 15(2):243~255
- 16 F. Bernhard, K. Christian, Real-time Face Detection using Edge-orientation Matching.Proceedings of Third International Conference Audio and Video Based Biometric Person Authentication, 2001: 78~83
- 17 M.Zobel, A.Gebhard, D.Paulus, J.Denzler, H.Niemann. Robust Facial Feature Localization by Coupled Features. Proceedings of the Fourth International Conference on Automatic Face and Gesture Recognition, 2000:2~7
- 18 C.S.Burrus, R.A.Gopinath, H.T.Guo. Introduction to Wavelets and Wavelet Transforms, Prentice Hall press, 1998
- 19 C. P. Papageorgiou, M. Oren, T. Poggio. A General Framework for Object Detection. Proceedings of Sixth International Conference on Computer Vision, 1998:555~562
- 20 C. Garcia, G. Tziritias. Face Detection Using Quantized Skin Color Regions Merging and Wavelet Packet Analysis. IEEE Transactions on Multimedia, 1999,1(3):264~277
- 21 P.Viola, M.Jones. Rapid object detection using a boosted cascade of simple features. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2001:511~518

- 22 D.G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 2004, 60(2): 91~110
- 23 A. Hadid, M. Pietikainen, T. Ahonen. A discriminative feature space for detecting and recognizing faces. *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2004:797~804
- 24 E.Hadjidemetriou, M.D.Grossberg, S.K.Nayar.Multiresolution Histograms and Their Use for Recognition.*IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, 26(7): 831~847
- 25 E.Hadjidemetriou, M.D.Grossberg, S.K.Nayar. Resolution Selection Using Generalized Entropies of Multiresolution Histograms. *The Seventh European Conference on Computer Vision*, 2002: 220~235
- 26 W.Hsu, T.Chua, H.Pung.An Integrated Color-Spatial Approach to Content-Based Image Retrieval. *ACM Multimedia*, 1995:305~313
- 27 M.Stricker, A.Dimai. Color Indexing with Weak Spatial Constraints. *Proceedings of SPIE Conference on Storage and Retrieval for Image Databases*, 1996:29~39
- 28 E.Osuna, R.Freund, F.Girosi.Training support vector machines: an application to face detection. *Proceedings of 1997 IEEE Conference on Computer Vision and Pattern Recognition*, 1997:130~136
- 29 C.Papageorgiou, T.Poggio. A training object system: car detection in static images. *MIT AI Memo No.180*, 1999
- 30 H. Schneiderman, T. Kanade. A Statistical Method for 3D Object Detection Applied to Faces and Cars. *IEEE International Conference on Computer Vision and Pattern Recognition*, 2000:746~751
- 31 P.Viola, M.Jones.Fast and Robust Classification using Asymmetric AdaBoost and a Detector Cascade. *Proceedings of 2001 Neural Information Processing Systems*, 2001: 1311~1318
- 32 P.Viola, M.Jones.Robust Real-Time Face Detection. *International Journal of Computer Vision*, 2004,57(2): 137~154
- 33 S.Z.Li, L.Zhu, Z.Q.Zhang, A.Blake, H.J.Zhang, H.Shum. Statistical learning of multi-view face detection. *Proceedings of the 7th European Conf. on Computer Vision*, 2002, (4): 67~81

- 34 X.R.Chen, A.Yuille. Detecting and Reading Text in Natural Scenes. Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, 2004:366~373
- 35 K.K.Sung, T.Poggio.Example-based learning for view-based human face detection. IEEE Transactions on Pattern Analysis And Machine Intelligence, 1998,20(1): 39~50
- 36 C.J.Liu, A Bayesian discriminating features method for face detection. IEEE Transactions on Pattern Analysis And Machine Intelligence, 2003, (25): 725~740
- 37 B.Menser, F.Muller.Face detection in color images using principal component analysis. Proceedings of 7th International Congress on Image Processing and its Applications, 1999:13~15
- 38 M.H.Yang, N.Ahuja, D.Kriegman. Face detection using mixtures of linear subspaces. Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition, 2000:70~76
- 39 A.Mohan, C. Papageorgiou, T.Poggio. Example-based object detection in images by components. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001,23(4): 349~361
- 40 M.V.Naquest, S.Ullman. Object recognition with informative features and linear classification. Proceedings of 9th International Conference on Computer Vision, 2003: 281~288
- 41 S.Agarwal, A.Awan, D.Roth.Learning to detect objects in images via a sparse, part-based representation. IEEE Transactions on Pattern Analysis And Machine Intelligence, 2004, 26(11): 1475~1490
- 42 R.Fergus, P.Perona, A.Zisserman. Object class recognition by unsupervised scale-invariant learning. Proceedings of 9th International Conference on Computer Vision and Pattern Recognition, 2003(2): 264~271
- 43 B.Leibe, A.Leonardis, B.Schiele.Combined object categorization and segmentation with an implicit shape model. ECCV2004 Workshop on Statistical Learning in Computer Vision, 2004
- 44 B.Leibe, B.Schiele. Scale-invariant object categorization using a scale-adaptive mean-shift search. Proceedings of DAGM'04 Annual Pattern Recognition Symposium, Springer LNCS, 3175, 2004: 145~153

- 45 Y.Y. Lin, T.L.Liu. Robust Face Detection with Multi-Class Boosting. Proceedings of International Conference on Computer Vision and Pattern Recognition, 2005: 680~687
- 46 Y.Amit, D.Geman, X.D.Fan. A Coarse-to-Fine Strategy for Multiclass Shape Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26(12): 1606~1621
- 47 F.F.Li, P.Perona. A Bayesian Hierarchical Model for Learning Natural Scene Categories. Proceedings of International Conference on Computer Vision and Pattern Recognition, 2005: 524~531
- 48 Fei-Fei Li, R. Fergus, P. Perona. A Bayesian Approach to Unsupervised One-Shot Learning of Object Categories. Proceedings of 9<sup>th</sup> International Conference on Computer Vision, 2003: 1134~1141
- 49 L. Fei-Fei, R. Fergus, P. Perona. Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories. International Conference on Computer Vision and Pattern Recognition 2004, Workshop on Generative-Model Based Vision, 2004:178~186
- 50 A. Torralba, K. P. Murphy, W. T. Freeman. Sharing features: efficient boosting procedures for multiclass object detection. Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004:762~769
- 51 Z.W.Tu, X.G.Chen, A.L.Yuille, S.C.Zhu. Image Parsing: Unifying Segmentation, Detection, and Recognition. Proceedings of 9<sup>th</sup> International Conference on Computer Vision, 2003:18~25
- 52 M.H. Yang, D.J. Kriegman, N. Ahuja. Detecting Faces in Images: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002,24(1): 34~58
- 53 C.Y.Kin, R.Cipolla. Feature-Based Human Face Detection. CUED/F-INFENG/TR 249, August 1996
- 54 C.Morimoto, M.Flickner. Real-Time Multiple Face Detection Using Active Illumination. Proceedings of the Fourth International Conference on Automatic Face and Gesture Recognition, 2000: 8~13
- 55 P.S Penio, J.A. Joseph. Local Feature Analysis: A general statistical theory for



- object representation, <http://venezia.Rockefeller.edu>
- 56 M.H.Yang, N.Ahuja, D.Kriegman. Face Detection Using Mixtures of Linear Subspaces. Proceedings of the Fourth International Conference on Automatic Face and Gesture Recognition, 2000:70~76
- 57 R.Feraud, O.Bernier, J-E.Viallet, M.Colobert. A Fast and Accurate Face Detector for Indexation of Face Images. Proceedings of the Fourth International Conference on Automatic Face and Gesture Recognition, 2000: 77~82
- 58 S.H.Lin, S.Y.Kung, L.J.Lin. Face Recognition/Detection by Probabilistic Decision-Based Neural Network. IEEE Transactions on Neural Networks. 1997,8(1): 114~132
- 59 L.R. Rabiner,B.H. Jung. Fundamentals of Speech Recognition. Prentice Hall, 1993
- 60 F.Samaria, S.Young. HMM based architecture for face identification. Image and Vision Computing, 1994(12): 537~583
- 61 R.Xiao, M.J.Li, H.J.Zhang. Robust Multi-Pose Face Detection in Images. IEEE Transactions on Circuits and Systems for Video Technology, 2004,14(1): 31~41
- 62 R.Xiao, L.Zhu, H.J.Zhang. Boosting Chain Learning for Object Detection. Proceedings of 9<sup>th</sup> International Conference on Computer Vision, 2003: 709~715
- 63 C.Liu, H.Shum.Kullback-Leibler Boosting and Its Application to Face Detection.Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003: 587~594
- 64 H.Martin, H.Hunke. Locating and Tracking of Human Faces with Neural Networks. Technical Report of CMU.CMU-Cs-94-155, 1994
- 65 S.Tony, A.Pentland. Parametrized Structure from Motion for 3D Adaptive Feedback Tracking of Faces. MIT Media Laboratory, Perceptual Computing Technical Report #401,1996
- 66 Y.Dai, Y.Nakano. Face-Texture Model Based on SGLD and Its Application in Face Detection in a Color Scene. Pattern Recognition, 1996, 29(6): 1007~1017

- 67 H. Wu, Q. Chen, M. Yachida. Face Detection from Color Images Using a Fuzzy Pattern Matching Method. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1999, 21(6): 557~563
- 68 R.L.Hsu, M.A.Mottaleb, A.K.Jain. Face detection in color images. Proceedings of International Conference on Image Processing, 2001: 1046~1049
- 69 M.Pietikainen, T.Ojala, Z.Xu. Rotation-invariant texture classification using feature distributions. Pattern Recognition, 2000, (33): 43~52
- 70 T.Ojala, M.Pietikainen, T.Maenpa. Multiresolution ray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(7): 971~987
- 71 V.Takala, T.Ahonen, M.Pietikainen. Block-based methods for image retrieval using local binary patterns. SCIA 2005 Proceedings of Image Analysis, Lecture Notes in Computer Science 3540, Springer, 2005: 882~891
- 72 边肇祺, 张学工. 模式识别. 清华大学出版社, 2000
- 73 荆其诚, 焦书兰, 喻柏林, 胡维生. 色度学. 科学出版社, 1979
- 74 羊昌尧. 电视原理. 人民邮电出版社, 1988
- 75 V.Vapnik. Statistical Learning Theory. Wiley, New York, 1998
- 76 T.W.S.Chow, D.Huang. Estimating optimal feature subsets using efficient estimation of high-dimensional mutual information. IEEE Transactions on Neural Networks, 2005, 16(1): 213~224
- 77 Y.M.Wu, A.D.Zhang. Feature selection for classifying high-dimensional numerical data. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004: 251~258
- 78 C. R. Rao. Linear statistical inference and its applications. John Wiley, 1973
- 79 P.Langley. Selection of relevant features in machine learning. Proceedings of the AAAI Fall Symposium on Relevance, 1994: 1~5
- 80 H.Liu, R.Setiono. A probabilistic approach to feature selection-a filter solution. Proceedings of the 13th International Conference on Machine Learning. 1996: 319~327
- 81 R.Kohavi, H.J.George. Wrappers for feature subset selection. Artificial Intelligence, 1997, 97(1-2): 273~324

- 82 J.H.Yang, V.Honavar, Feature subset selection using a genetic algorithm. IEEE Intelligent Systems, 1998, (13): 44~49
- 83 UIUC Car Image Database, <http://l2r.cs.uiuc.edu/~cogcomp/Data/Car/>, 2004
- 84 R.Lienhart, A.Wernicked.Localizing and segmenting text in images and videos. IEEE Transactions on Circuits and Systems for Video Technology. 2002,12(4): 236~268
- 85 X.S.Hua, W.Y.Liu, H.J.Zhang.An automatic performance evaluation protocol for video text detection algorithms. IEEE Transactions on Circuits and Systems for Video Technology, 2004,14(4): 498~507
- 86 X.S.Hua, W.Y.Liu, H.J.Zhang.Automatic performance evaluation protocol for video text detection algorithms. Proceedings of International Conference on Document Analysis and Recognition, 2001: 545~550
- 87 D.J.C.MacKay. Information Theory, Inference and Learning Algorithms. Cambridge University Press, 2003
- 88 朱雪龙.应用信息论基础.清华大学出版社, 2001
- 89 傅祖芸.信息论-基础理论与应用.电子工业出版社, 2002
- 90 姜丹.信息论与编码.中国科学技术大学出版社, 2004
- 91 王新梅,肖国镇.纠错码-原理与方法.西安电子科技大学出版社,2003
- 92 T.G.Dietterich, G.Bakiri.Error-correcting output codes: A general method for improving multiclass inductive learning programs. Proceedings of the Ninth National Conference on Artificial Intelligence (AAAI-91), 1991:572~577
- 93 T.G.Dietterich, G.Bakiri.Solving multi-class learning problems via error correcting output codes. Journal of Artificial Intelligence Research, 1995, (2): 263~286
- 94 R.Ghani.Using error-correcting codes for text classification. Proceedings of 17th International Conference on Machine Learning, 2000: pp303~310
- 95 J.Kittler, R.Ghaderi, T.Windeatt, J.Matas.Face verification via error correcting output codes. Image and Vision Computing, 2003, 21(13-14): 1163~1169
- 96 S.Mahamud, M.Hebert, J.Shi. Object recognition using boosted discriminants. IEEE Conference on Computer Vision and Pattern Recognition, 2001:551~558
- 97 H. Schneiderman, T. Kanade.Object Detection Using the Statistics of Parts.

- International Journal of Computer Vision, 2004, 56(3): 151~177
- 98 C. Huang, B. Wu, H.Z.Ai, S.H.Lao. Omni-Directional Face Detection based on Real AdaBoost. IEEE International Conference on Image Processing 2004, 2004: 24~27
- 99 S.Z.Li, Z.Q.Zhang. FloatBoost Learning and Statistical Face Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26(9): 1112~1123
- 100 M. Jones, P. Viola. Fast Multi-view face detection. Technical Report TR2003-96, Mitsubishi Electric Research Laboratories, 2003
- 101 H.A. Rowley, S. Baluja, T. Kanade. Rotation Invariant Neural Network-Based Face Detection. IEEE International Conferences on Computer Vision and Pattern Recognition, 1998: 38~44
- 102 Y. Li, S. Gong, H. Liddell. Support vector regression and classification based multi-view face detection and recognition. IEEE International Conference on Automatic Face & Gesture Recognition, Grenoble, France, 2000: 300~305
- 103 M. Osadchy, M.L. Miller, Y. LeCun. Synergistic Face Detection and Pose Estimation with Energy-Based Models. The proceedings of 17th Advances in Neural Information Processing Systems, 2004: 1017~1024
- 104 J. Miao, B. Yin, K. Wang, L. Shen, X. Chen. A hierarchical multiscale and multiangle system for human face detection in a complex background using gravity-center template. Pattern Recognition, 1999, 32(7): 1237~1248.
- 105 O. Dekel, Y. Singer. Multiclass Learning by Probabilistic Embeddings. The proceedings of 15th Advances in Neural Information Processing Systems, 2002: 945~952
- 106 T. Randen, J.H. Husoy. Filtering for texture classification: a comparative study. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1999, 21(4): 291~310
- 107 Z.W. Tu. Probabilistic Boosting-Tree: Learning Discriminative Models for Classification, Recognition, and Clustering. Proceedings of Tenth IEEE International Conference on Computer Vision, 2005: 1589~1596
- 108 张晓华, 山世光, 曹波, 高文, 周德龙, 赵德斌. CAS-PEAL 大规模中国人脸图像数据库及其基本评测介绍. 计算机辅助设计与图形学学报, 2005, 17(1): 9~17

## 攻读学位期间发表的学术论文

### 第一作者论文:

- 1 Hongming Zhang, Wen Gao, Debin Zhao, Xilin Chen. Object Detection Using Spatial Histogram Features. Image and Vision Computing. Elsevier. Accepted. [SCI 源国际期刊]
- 2 张洪明, 赵德斌, 高文. 基于肤色模型、神经网络和人脸结构模型的平面旋转人脸检测. 计算机学报, 2002, 25(11): 1250~1256 [EI 源国内一级期刊]
- 3 Hongming Zhang, Debin Zhao. Spatial Histogram Features for Face Detection in Color Images. Advances in Multimedia Information Processing - PCM 2004: 5th Pacific Rim Conference on Multimedia, Lecture Notes in Computer Science 3331, 2004: 377~384 [SCI]
- 4 Hongming Zhang, Wen Gao, Debin Zhao, Xilin Chen. Learning Informative Features for Spatial Histogram-Based Object Detection. Proceedings of 2005 International Joint Conference on Neural Networks, Montreal, Canada, Jul.31-August.4, 2005: 1806~1811
- 5 Hongming Zhang, Wen Gao, Xilin Chen, Shiguang Shan, Debin Zhao. Robust Multi-View Face Detection Using Error Correcting Output Codes. 9th European Conference on Computer Vision, 2006 (Accepted as Oral)
- 6 Hongming Zhang, Debin Zhao, Wen Gao, Xilin Chen. Combining Skin Color Model and Neural Network for Rotation Invariant Face Detection. ICMI'2000, Lecture Notes in Computer Science 1948, 2000: 237~244 [SCI]
- 7 张洪明, 高文, 赵德斌, 陈熙霖. 空间直方图特征及其在物体检测中的应用. (已投计算机学报)

### 合作论文:

- 8 张文超, 山世光, 张洪明, 陈杰, 陈熙霖, 高文. 基于局部 Gabor 变化直方图序列的人脸描述与识别方法. 软件学报 (已录用)
- 9 Jun Miao, Hongming Zhang, Wen Gao, et al. FaceTracker: A Human Face Tracking and Facial Organ Localizing System. The Proc. of the 8th International Conference on Computer Vision, 2001: 743~743
- 10 Lijuan Duan, Guoqin Cui, Wen Gao, Hongming Zhang. Adult Image

- Detection Method Base-On Skin Color Model and Support Vector Machine. Fifth Asian Conference on Computer Vision (ACCV2002). Jan 22-25, 2002.
- 11 Shiguang Shan, Wen Gao, Jie Yan, Hongming Zhang, and Xilin Chen. Individual 3D Face Synthesis Based on Orthogonal Photos and Speech-Driven Facial Animation. Proceeding of 2000 International Conference on Image Processing, 2000, (2): 238~241
  - 12 Jun Miao, Hong Liu, Wen Gao, Hongming Zhang, Gang Deng, Xilin Chen. A System for Human Face and Facial Feature Location. International Journal of Image And Graphics (IJIG), 2003,13(3): 461~479
  - 13 Wenchao Zhang, Shiguang Shan, Hongming Zhang, Wen Gao, Xilin Chen. Multi-resolution Histograms of Local Variation Patterns (MHLVP) for Robust Face Recognition. Audio- and Video-based Biometric Person Authentication 2005, July 20 - 22, 2005: 937~944 [EI, SCI]
  - 14 Wenchao Zhang, Shiguang Shan, Wen Gao, Xilin Chen, Hongming Zhang. Local Gabor Binary Pattern Histogram Sequence (LGBPHS): a Novel Non-Statistical Model for Face Representation and Recognition. Proceedings of Tenth IEEE International Conference on Computer Vision, 2005:786~791[EI]

## 哈尔滨工业大学博士学位论文原创性声明

本人郑重声明：此处所提交的博士学位论文《基于局部纹理特征的物体检测方法》，是本人在导师指导下，在哈尔滨工业大学攻读博士学位期间独立进行研究工作所取得的成果。据本人所知，论文中除已注明部分外不包含他人已发表或撰写过的研究成果。对本文的研究工作做出重要贡献的个人和集体，均已在文中以明确方式注明。本声明的法律结果将完全由本人承担。

作者签字：

日期： 年 月 日

## 哈尔滨工业大学博士学位论文使用授权书

《基于局部纹理特征的物体检测方法》系本人在哈尔滨工业大学攻读博士学位期间在导师指导下完成的博士学位论文。本论文的研究成果归哈尔滨工业大学所有，本论文的研究内容不得以其它单位的名义发表。本人完全了解哈尔滨工业大学关于保存、使用学位论文的规定，同意学校保留并向有关部门送交论文的复印件和电子版本，允许论文被查阅和借阅。本人授权哈尔滨工业大学，可以采用影印、缩印或其他复制手段保存论文，可以公布论文的全部或部分内容。

保密☐，在 年解密后适用本授权书。

本学位论文属于

不保密☐。

作者签名：

日期： 年 月 日

导师签名：

日期： 年 月 日

## 致谢

值此论文完成之际，我体会到了辛苦劳动后的喜悦。在此，我衷心感谢在求学过程中关心、帮助和支持我的所有人。

首先，我要衷心感谢我的导师高文教授。高老师不但给予我学业上的指导，而且鼓励我积极探索新思想和新方法。每当我在课题选择、实验进展、论文写作各方面遇到疑难问题时，总是能够得到导师认真的指导和富有建设性的建议。作为实验室的开创者、领导者和学术带头人，高老师广阔的视野、渊博的学识、严谨民主的学风、高超的领导才能和在事业上不断追求的精神，值得学生终生学习。

感谢实验室的陈熙霖、赵德斌老师对我的悉心培养。论文工作自始至终都获得了陈老师和赵老师的认真指点。两位老师广博的学识、敏锐的思维、无私的科学精神、乐观的生活态度，是我一直钦佩和学习的榜样。和两位老师的学习和交流的过程，使我受益匪浅。

感谢实验室的刘岩、姚鸿勋老师。他们无私奉献的工作精神和硕果累累的科研成就，是我学习的榜样。

感谢山世光博士的合作与帮助，他和刘岩老师编写的图像处理平台多年以来一直是我完成研究工作的基础。感谢曾炜、吕岩、孙晓艳、王瑞等师兄师姐的帮助。感谢王建宇、苗军、闫胜业、张文超、蔡秀娟、卿来云、陈杰、叶齐祥等对我工作的启发和帮助。感谢 JDL 联合实验室所有的老师和同学们，大家共同营造了团结友好、互相帮助、轻松愉快的良好环境。

向我的父母、我的妻子、以及我的亲人表示最诚挚的谢意，他们多年来对我始终如一的爱和鼓励是我不断进取的动力。

本文受到国家自然科学基金(项目编号：60332010)、国家 863 计划(2002AA118010)、中科院百人计划的资助、上海市科委项目(03DZ15013)、以及上海银晨智能识别科技有限公司的资助。



## 个人简历

张洪明(1976.4)，男，江西萍乡人。主要研究领域为物体检测、计算机视觉、图像处理、模式识别、机器学习。在攻读博士学位期间，发表和录用国际期刊论文 1 篇、国内期刊论文 1 篇、国际会议论文 4 篇，另有与其他作者合作论文 7 篇、申请国内专利 1 项。

### 教育背景

- 1994.9~1998.7 哈尔滨工业大学计算机科学与工程系，工学学士。
- 1998.9~2000.7 哈尔滨工业大学计算机科学与工程系，工学硕士，论文题目：人脸检测技术的研究与实现。
- 2000.3~2006.2 哈尔滨工业大学计算机科学与工程系，工学博士，论文题目：基于局部纹理特征的物体检测方法。

### 攻读博士学位期间参与的科研项目

- 1 2004~2005 基于生物特征的身份识别研究(国家自然科学基金重点项目，项目编号：60332010)：主要工作：物体检测算法研究
- 2 2002~2004 多媒体掌上电脑与通信掌上电脑(国家 863 计划，项目编号：2002AA118010)：主要工作：移动平台下的人脸检测
- 3 2001~2003 生物特征识别核心技术与关键问题研究(国家 863 计划，项目编号：2001AA114190)：主要工作：对光照鲁棒的快速人脸检测算法研究与实现
- 4 2000~2006 面像识别研究(上海银晨智能识别科技有限公司资助项目)：主要工作：人脸检测算法和运动物体检测算法的研究与实现
- 5 2004~2005 中科院百人计划资助项目