

清华大学

综合论文训练

题目：道路交通标志检测

系 别：计算机科学与技术系

专 业：计算机科学与技术

姓 名：王鹏

指导教师：李建民 助理研究员

2010年7月

关于学位论文使用授权的说明

本人完全了解清华大学有关保留、使用学位论文的规定，即：学校有权保留学位论文的复印件，允许该论文被查阅和借阅；学校可以公布该论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存该论文。

(涉密的学位论文在解密后应遵守此规定)

签 名：_____ 导 师 签 名：_____ 日 期：_____

中文摘要

本文通过颜色阈值法以及基于Haar特征的快速级层分类器实现道路交通标志的检测，并对交通标志相对摄像机的运动建立了概率图模型和基于该模型的一种快速推理算法，从而对交通标志在图像中可能出现的位置进行预测，以此增强检测的准确度。本论文提出了三种方法：颜色阈值法，层级分类器法，以及运动概率模型。他们分别适用于大色块的不规则标识、形状规则的标志，以及视频。颜色阈值的原理是在HSV颜色空间的各个通道上取阈值来进行标识识别。层级分类器法使用Haar特征来构建层状的逐级分类器。运动概率模型对标志在图像上的运动建立概率模型。这三个方法彼此互补，能够自由组合。

本文的创新点主要有：

- 使用HSV颜色空间进行颜色阈值；
- 提出了交通标志运动的概率图模型；
- 提出了一种快速推理算法。

关键词： 交通标志检测 Haar特征 级层分类器 概率图模型 颜色
阈值

ABSTRACT

This paper implements a fast road sign detector based on color thresholding and a fast cascade classifier using Haar wavelets. It also describes a probability graph modeling the motion of road signs, as well as an efficient inference algorithm on this graphic model. This model is used to predict the probability of a road sign appearing at any given position on the image, thus improving the performance of detection. In this paper, three major methods are explored, which are color-thresholding based on HSV color space, cascade classifier and motion probability model. They are suitable for signs with large single color region, signs with regular shape, and videos respectively. Color-thresholding method sets upper and lower thresholds for each channel in the HSV color space. Cascade classifier classifies samples with a stage-by-stage manner. Motion probability model captures the motion of the sign in image. These three methods complementary and can be freely composed.

The contribution of this paper includes:

- Using HSV color space to calculate color thresholding;
- Introducing a graph model describing road signs' motion;
- Introducing an efficient inference algorithm.

Key words: road sign detection Haar-wavelets cascade classifier graph model color thresholding

目 录

第 1 章 引言	1
1.1 研究背景	1
1.2 研究现状	1
1.3 内容概述	3
第 2 章 实验数据	4
2.1 数据集1	4
2.2 数据集2	5
第 3 章 HSV空间的颜色阈值	7
3.1 算法描述	7
3.1.1 HSV颜色空间	7
3.1.2 检测目标	8
3.1.3 阈值选择	9
3.1.4 噪声过滤	9
3.1.5 包围框的生成	9
3.2 实验结果	10
3.2.1 HSV各通道的区分效果	10
3.2.2 阈值筛选后的效果	11
3.2.3 噪音过滤的效果	11
3.2.4 包围框生成的效果	13
3.3 Histogram Equalization算法及其效果	13
3.4 不足及讨论	13
第 4 章 Haar特征层级分类器	17
4.1 算法描述	17
4.1.1 层级分类器	17
4.1.2 每一层的Haar检测器	18

4.1.3 Haar特征	19
4.1.3.1 基本Haar特征	19
4.1.3.2 Haar特征的快速算法	19
4.1.3.3 倾斜Haar特征	20
4.1.4 级层分类器的训练算法	20
4.2 实验结果	21
4.2.1 只使用基本Haar特征	21
4.2.2 使用扩展Haar特征	22
第5章 运动概率模型	27
5.1 算法描述	27
5.1.1 交通标志的运动公式	27
5.1.2 最大似然估计	29
5.1.3 概率图模型	30
5.1.4 推理算法	32
5.1.5 在线版本的推理算法	36
5.1.6 混合高斯扩展	38
5.2 实验结果	43
5.2.1 最大似然估计	43
5.2.2 在模拟数据上的测试	43
5.2.3 在实际数据上的测试	45
第6章 总结与讨论	47
6.1 各方法优缺点	47
6.2 主要贡献	47
插图索引	49
表格索引	51
公式索引	52
参考文献	55

致 谢	57
声 明	58
附录 A 外文资料的调研阅读报告（英文）	59

主要符号对照表

$P(\mathbf{X})$	事件 \mathbf{X} 的概率
$E(\mathbf{X})$	随机变量 \mathbf{X} 的期望
$\mathcal{N}(\mathbf{X} \mu, \sigma^2)$	随机变量 \mathbf{X} 服从期望为 μ 标准差为 σ 的高斯分布
$\mathcal{U}(a, b)$	$[a, b]$ 区间上的均匀分布
$D(x, y)$	$\sum_n xy - \frac{\sum_n x \sum_n y}{n}$

第 1 章 引言

1.1 研究背景

道路交通标志检测的研究是无人驾驶智能车的子项目。其目标是能够在无人驾驶汽车行驶时拍摄的实时道路视频中自动检测出交通标志，并识别交通标志的类别，以此信息来指导无人驾驶汽车后续的行驶。本项目只实现交通标志的检测，不包含标志的识别以及利用标志信息对车辆进行操纵。交通标志的检测是指在视频的每一帧中，确定出现的每个交通标志的位置及大小（即确定其包围框）。

该项目的难度主要有两个方面：一是要求快速实时，因为它应用于车辆行驶中的实时视频上，并且需要第一时间检测到出现的交通标志并报告给车辆的控制器；而且要求精度比较高，因为它实际应用于车辆的控制。

本实验中所说的交通标志指如图1.1的20类。

1.2 研究现状

交通标志识别既作为一项子任务被自动控制、机器人领域广泛研究，也作为一项特定物体识别任务被计算机视觉、图像处理领域研究。已有的交通标志检测算法大体上可以分为两大类：基于颜色的检测算法^[1-6]，和基于形状的检测

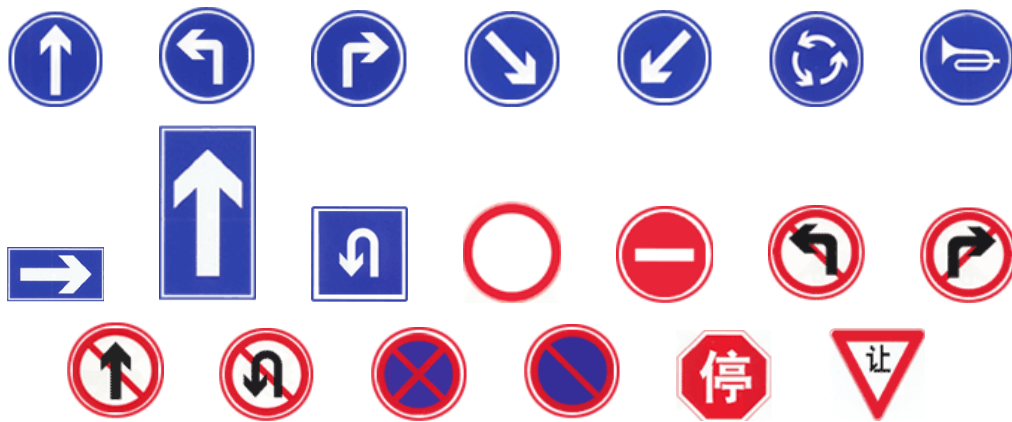


图 1.1 标准20类交通标志

测算法^[7-11]。[1-3]都使用了颜色阈值方法来检测出可能构成交通标志的各个像素，再通过后续的图像处理方法来得到交通标志的包围框。他们所使用的颜色空间和后续图像处理方法各不相同。在基于形状的算法中，大部分文章的算法都借鉴了人脸识别中使用广泛的特征，即Haar小波^[8]。在基本的Haar小波之上，[9]提出了加入颜色信息的Haar小波，即把彩色图像分成7个颜色通道，在每个通道上单独生成Haar小波，再将所有7的通道上的Haar小波合起来作为特征集。[10]引入了对角线型的Haar小波及其快速计算方法，从而使得 Haar小波可以对45度角的颜色变化进行更好的描述。[12]讨论了任何一个角度的Haar小波都可以用两个基本Haar小波的线性和等价表示，作者将这一对基本Haar小波称为“pair of equivalent features”。

在同样都使用Haar小波的情况下，各文章使用的分类器并不相同。[7]引入了经典的快速级层（cascade）分类器。它使得各个训练样例依次通过级层分类器的各层，当样例在任何一层被判为负例时就会被排除，无法进入下一层。最后通过所有层的样例即认为是正例。每一层都保证在尽量不误判正例的情况下，尽可能多地将负例排除。各层的分类能力由弱到强，大部分的负例都会在早期即被排除，只有正例和少数负例会进入后期的分类器。各层分类器都通过Adaboost算法训练得到，每层的训练样例是所有正例和通过了之前所有层的负例，训练方法是利用Adaboost在所有Haar特征中选择足够的特征组合起来，使其能够成功将训练集分类。这样的级层分类器框架特别适用于负例相对于正例多得多的情况下，而物体检测由于使用的是滑动窗口，正是这样一种情况。

除了使用Haar特征和级层分类器，[11]提出了使用6个局部特征部件来形成椭圆检测器，对交通标志进行检测。[13]使用了Laplace Kernel作为分类器。[9]使用了LDA作为分类器。

在构建了检测器之后，由于图片场景中往往包含空间环境信息（如地平线、天空、地面），视频中往往包含物体的运动信息，所以我们可以利用这些信息建立概率模型还预测各个位置出现交通标志的可能性，从而提高检测准确率。[14]提出了通过将图片进行粗分割（分成天空、地面、竖立平面）和对地平线等环境参数的估计来加强检测效果，其检测器使用了HOG特征^[15]和SVM分类器。[16]提出了一个概率图模型来建模图像经分割后各个区块的几何关系。[17]通过交通标志的运动方程来设计Kalman滤波器对标志进行跟踪。[18]使用了一种 Knowledge-Fusion方法来建立交通标志的运动模型。

1.3 内容概述

本文分成四个部分，第一部分介绍了实验所用数据及实验方法；第二部分介绍基于HSV颜色空间的颜色阈值算法及其实验结果；第三部分介绍基于Haar特征的快速级层分类器及其加入颜色信息的扩展，以及实验结果；第四部分介绍了基于交通标志运动方程的概率图模型及其推理算法，以及实验结果；第五部分为讨论及总结。

第 2 章 实验数据

本论文使用了两个数据集。

2.1 数据集1

数据集1是一段大约1小时长的道路视频，由6段视频拼成，从中手工标出了大约 500个交通标志的包围框（没有覆盖所有出现）。可以用来测试检测器的性能，但由于标注的交通标志并非来自连续帧，所以无法用于跟踪或运动模型。

数据集1的帧如图2.1

数据集1中标注出的交通标识如图2.2。

数据集1中的一段200帧左右的连续序列经过了彻底标注，可以用于跟踪运动。



图 2.1 数据集1的帧示例

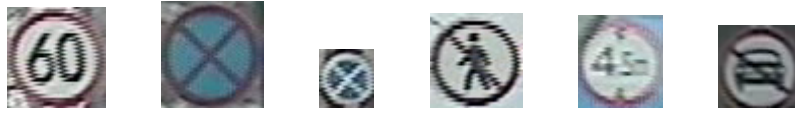


图 2.2 数据集1中标注出的交通标识示例



图 2.3 数据集2的帧示例

2.2 数据集2

数据集2也是一段道路视频，这段视频大约4分钟，由8个小段拼成。这段视频经过了彻底标注，大约出现了5千次的交通标志观测，由大约由200个交通标志产生，即每个交通标志大约产生25次观测。可以用于跟踪运动。

数据集2的帧如图2.3

数据集2中标注出的交通标志在训练分类器时分成了三大类，第一类如图2.4，第二类如图2.5，第三类如图2.6。



图 2.4 数据集2中标注出的交通标识示例（第一类）



图 2.5 数据集2中标注出的交通标识示例（第二类）



图 2.6 数据集2中标注出的交通标识示例（第三类）

第3章 HSV空间的颜色阈值

3.1 算法描述

本算法使用HSV颜色空间的颜色阈值算法，将其作为对某些类型的交通标志的检测器。它将彩色图像转换到HSV颜色空间，在三个颜色通道上选取上下界作为阈值，从而提取出可能组成交通标志的像素点。然后通过Erode、Close等图像处理基本操作对这些像素点进行过滤，取出噪声。最后通过联通块合并形成交通标志的包围框。

3.1.1 HSV颜色空间

HSV是众多颜色空间中的一种。它的H、S、V三个通道分别表示色调、饱和度、亮度，其中S和V为0到1的实数，H一般表示为360度的角度。整个HSV颜色空间可以表示为一个圆锥形，如图3.1。

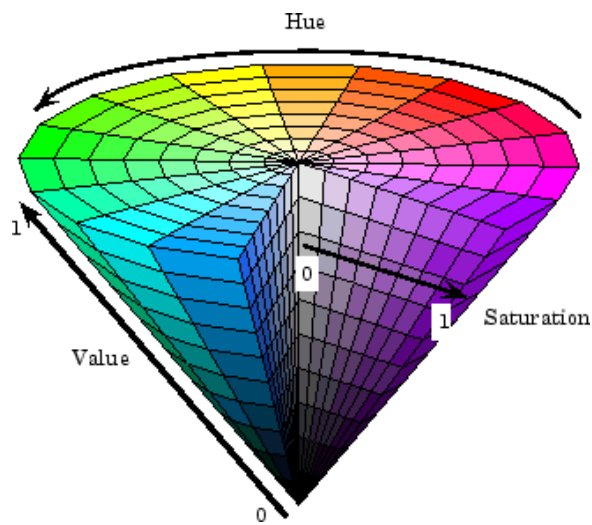


图 3.1 HSV的圆锥示意图。引自Matlab™帮助文件

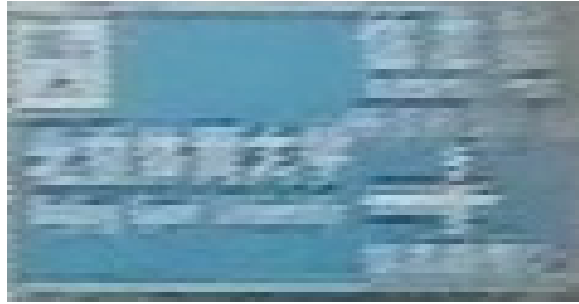


图 3.2 大面积、实色填充的矩形指路标志

HSV与RGB颜色空间的转换公式为：

$$\begin{cases} V = M \\ S = \frac{M-m}{M} \\ H = \begin{cases} 60h, & h \geq 0 \\ 60h + 360, & h < 0 \end{cases} \end{cases} \quad (3-1)$$

其中，

$$M = \max(R, G, B) \quad (3-2)$$

$$m = \min(R, G, B) \quad (3-3)$$

$$h = \begin{cases} \frac{G-B}{M-m}, & M = R \\ 2 + \frac{B-R}{M-m}, & M = G \\ 4 + \frac{R-G}{M-m}, & M = B \end{cases} \quad (3-4)$$

3.1.2 检测目标

在本实验中，HSV阈值算法主要用来识别一些大面积的、实色填充的矩形指路标志，例如图3.2。

这些标志的特点是大小及长宽比不固定、内容变化较大、外观缺乏统一模式或没有明显的模式，所以不适合用后文所述的基于形状和外观模式的检测器进行检测。而由于它们的色块较大、颜色较明显，更适合于基于颜色的检测器。

注意这类交通标志不属于前文所述的20种基本标志，属于扩展实验。但它会被用于后续的智能车任务。

3.1.3 阈值选择

阈值选择的方法分为手动和自动两种。自动阈值选择可以使用某种搜索策略来在阈值空间中搜索达到最佳分类效果的阈值，分类效果可以由训练集上的运行结果确定。

本实验使用手工调整阈值。

3.1.4 噪声过滤

对图片通过阈值检测，其结果是在各个通道内，图片成为二值图片，即每个像素只有0（表示没有在阈值范围内，不是目标对象，即交通标志）和1（表示在阈值范围内，是交通标志）。由于阈值检测是以像素为单位进行的，所有结果图片中可能存在有许多孤立的值为1的点，以下简称为“白点”。另外，在白色区块的边缘，也可能出现许多锯齿。而且，本来属于同一交通标志的白点，可能被分割为两块不连通的区域，影响后续用联通域算法对包围框的计算。所以，需要使用一些处理来过滤掉这些孤立的白点，并让被割断的、非常接近的白色区块连成一整块。

Erode、Dilate、Close等操作都属于卷积过滤操作。它们的基本操作模式都是用一个小的模板（或称“核”）在图片上进行滑动，对模板所覆盖的像素进行某种操作，将操作结果赋给模板中心所在的像素。Erode使用的操作是min（取最小值），Dilate是max（取最大值），而Close是指先进行Erode，然后进行Dilate。

3.1.5 包围框的生成

包围框使用联通域算法生成。经过上面的处理后，图像仍是一堆单个的像素点，没有它们之间是否衔接、是否组成区块的信息。联通域域算法从单个的像素点群中计算出各个联通域。所谓“联通”的概念有两种：强连通，或“4联通”，是指将左右、上下相邻的像素定义为联通；弱连通，或“8联通”，是在强连通的基础上，将对角线方向相邻的像素也视为联通。

计算联通域的算法有很多，本文使用洪泛（Flooding）算法，即从某点出发，将其连通的点加入域，在依次从这些点出发，将联通的点加入域。深度优先或广度优先算法均可。

在得到了各个联通域后，由联通域生成包围框的方法是：每个联通域对应一个包围框，联通域内最左、最右、最上、最下的像素确定了包围框左、右、上、下四条边的位置。

生成的包围框还可以进行一些预处理，比如两个包围框可能出现重叠，而交通标志一般不会出现重叠，它更有可能来自同一个交通标志被割裂开的两部分。所以可以把它们合并成一个标志。

3.2 实验结果

3.2.1 HSV各通道的区分效果

本节展示HSV的各通道对于图片的现实效果，从中可以定性判断出各个通道能不能把交通标志从背景中很好的凸显出来。

作为例子，实验展示使用的图片见图3.3。

进行HSV空间转换后，

H通道下的图片见图3.5。

S通道下的图片见图3.4。

V通道下的图片见图3.6。

从以上三幅图的结果可以看出，H和S通道对交通标志的区分能力比较好，尤其是S通道。回想三个通道的直观含义，H通道表示色调，即红、黄、蓝、绿等等“基本颜色”，S通道表示饱和度，即颜色的鲜艳程度。而我们考虑到交通标志颜色使用的原则，就是通过特定颜色的、鲜艳的色块来尽量明显地让司机把它从其背景中辨识出来。而另一方面，V通道代表亮度，它往往由环境决定，交通标志的亮度并不一定比背景高。由此可以看出H、S、V各通道的实验效果可以通过交通标志的设计原则得到解释。

鉴于此，本论文只对H和S通道取阈值，然后将二者做与运算。

本实验使用手工调整阈值，根据本组训练数据，选择的阈值分别为：

- H（色调）通道，判断为交通标志所允许的最小值为7，判断为交通标志所允许的最大值为22；
 - S（饱和度）通道，判断为交通标志所允许的最小值为100，判断为交通标志所允许的最大值为256；
- （各通道的取值范围是[0, 256)）

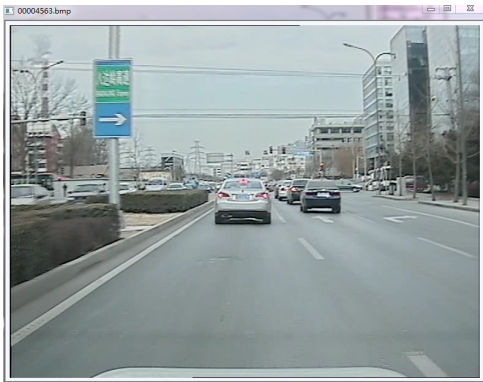


图 3.3 原图像



图 3.5 H通道。白色表示值最大，黑色表示值最小。



图 3.4 S通道。白色表示值最大，黑色表示值最小。



图 3.6 V通道。白色表示值最大，黑色表示值最小。

3.2.2 阈值筛选后的效果

在对H、S通道应用了上文所述的阈值后，H、S通道下在阈值范围内的像素组成的图片如下：

H通道见图3.7。

S通道见图3.10。

对H和S做与运算后的结果见图3.8。

3.2.3 噪音过滤的效果

经过Erode算法处理后，效果见图3.11。

经过Close算法处理后，效果见图3.9。



图 3.7 H通道取阈值后的结果。白色表示表示在阈值范围内，可能是交通标志的一部分。

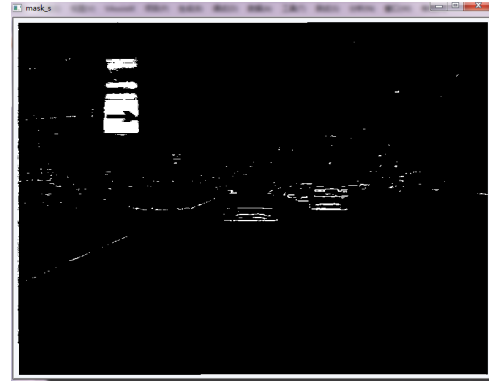


图 3.10 S通道取阈值后的结果。白色表示表示在阈值范围内，可能是交通标志的一部分。



图 3.8 取与运算后的Mask。白色表示表示在阈值范围内，可能是交通标志的一部分。



图 3.11 Erode算法处理后的图像。白色表示表示在阈值范围内，可能是交通标志的一部分。

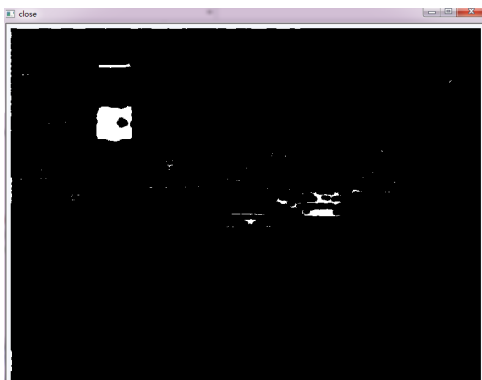


图 3.9 Close算法处理后的图像。白色表示表示在阈值范围内，可能是交通标志的一部分。

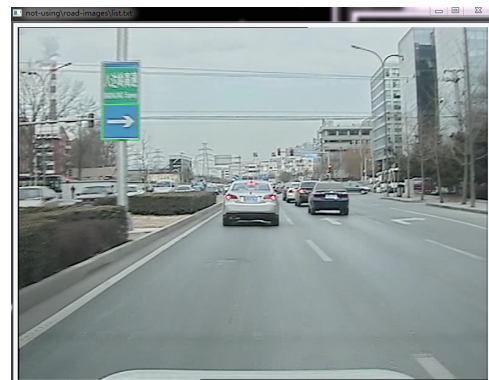


图 3.12 生成的包围框

3.2.4 包围框生成的效果

经过使用洪泛算法生成联通域，进而由联通域生成包围框后，各包围框在原图片中的位置见图3.12。 s

3.3 Histogram Equalization算法及其效果

Histogram Equalization算法是一个对颜色进行规范化的典型方法。它的数学含义可以描述如下。

定义 3.1: 对于一个分布 $P(\mathbf{X})$ ，其对应的Histogram Equalization是一个实值函数 $f_P(\cdot; a, b)$ ，其中 a 、 b 为参数，满足：

$$f_P(x; a, b) \equiv P(\mathbf{X} < x)(b - a) + a \quad (3-5)$$

Histogram Equalization具有如下性质：

定理 3.1: 若 $\mathbf{X} \sim P(\mathbf{X})$ ，则 $f_P(X; a, b) \sim \mathcal{U}(a, b)$ ，其中 $\mathcal{U}(a, b)$ 表示 $[a, b]$ 区间上的均匀分布。

Histogram Equalization操作可以使图片各像素的值分布更加均匀，从而增大整个图片的对比度；并且它使整个图片的平均像素值居中，从而归一化了整个图片的亮度。Histogram Equalization操作一般只能在单个通道上进行，本实验中在RGB通道上进行。

进行Histogram Equalization后，图片如图3.13。其H通道为图3.16，S通道为图3.14，V通道为图3.17。H通道经阈值后为图3.15，S通道经阈值后为图3.18。取与运算后为图3.19。经Erode操作后为图3.22，经Close操作后为图3.20。最后的包围框结果为图3.23。

3.4 不足及讨论

颜色阈值法有两个主要缺点。一是手动设置阈值对场景的依赖比较大，当光照强度、摄像机曝光参数发生变化时，往往需要重新调整参数。这个缺点可以用前文所述的自动参数调整来克服。

第二个缺点是交通标志的颜色特征并非其特有，其它的物体也可能产生相同的颜色，例如图3.21和图3.24所示。图中的公共汽车的挡板具有与交通标志

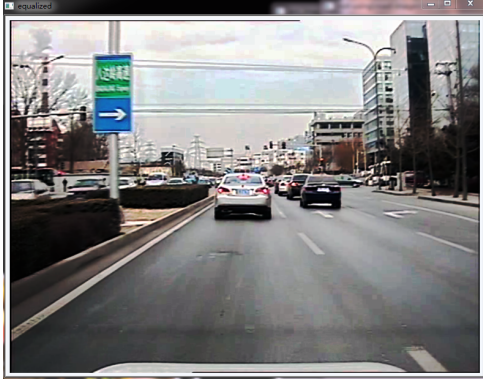


图 3.13 Histogram Equalization后的原图像



图 3.16 Histogram Equalization后的H通道。白色表示值最大，黑色表示值最小。



图 3.14 Histogram Equalization后的S通道。白色表示值最大，黑色表示值最小。



图 3.17 Histogram Equalization后的V通道。白色表示值最大，黑色表示值最小。



图 3.15 Histogram Equalization后的H通道取阈值后的结果。白色表示表示在阈值范围内，可能是交通标志的一部分。



图 3.18 Histogram Equalization后的S通道取阈值后的结果。白色表示表示在阈值范围内，可能是交通标志的一部分。



图 3.19 Histogram Equalization后和取与运算后的Mask。白色表示表示在阈值范围内，可能是交通标志的一部分。

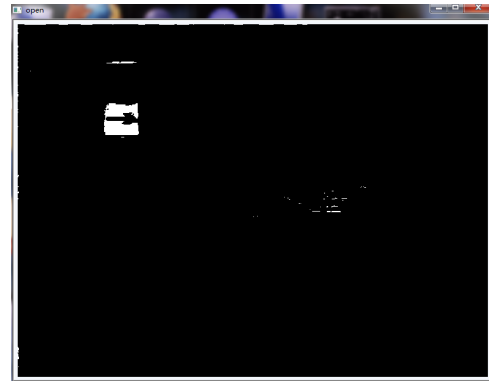


图 3.22 Histogram Equalization后和Erode算法处理后的图像。白色表示表示在阈值范围内，可能是交通标志的一部分。



图 3.20 Histogram Equalization后合Close算法处理后的图像。白色表示表示在阈值范围内，可能是交通标志的一部分。

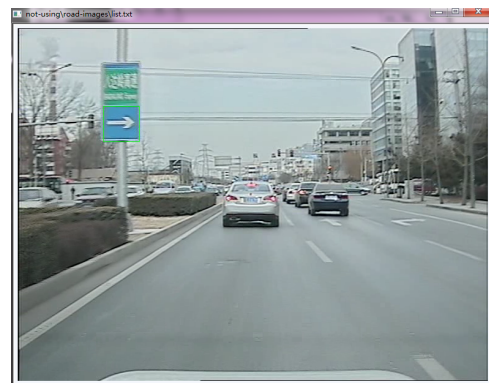


图 3.23 Histogram Equalization后生成的包围框

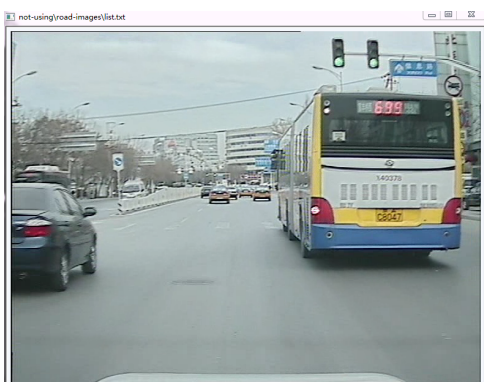


图 3.21 一幅出现误判的图片



图 3.24 误判图片最后的处理结果。经过所有处理后，巴士的后挡板仍很难与交通标志区分。

相同的颜色，很难区分。减缓这个问题的方法有两种，一是可以限制交通标志的长宽比，例如通常如此扁长的矩形不太可能是交通标志；二是利用后文所述的概率模型对交通标志可能出现的位置做出估计。

第 4 章 Haar特征层级分类器

4.1 算法描述

本算法基于[7]提出的利用Haar小波的快速层级（Cascade）分类器，并参考[9]将Haar特征拓展到具有颜色信息。

4.1.1 层级分类器

层级分类器的基本思想是在负例远远多于正例、而且大多数负例很容易被排除的情况下，利用简单的、快速计算的分类器排除大部分负例，仅剩少数较难区分的负例才动用复杂的、计算量较大的分类器。

层级分类器的每层都是一个单独的分类器，分类器的复杂度及分类能力由低向高逐渐增加。在使用它进行分类时，样例从低向高逐级通过各个分类器，如果在某一层该样例被判定为负例，该样例将被排除而不再通往下一层。只有那些通过了所有层的样例，才被判断为正例。如图4.1所示。

通过这样的设计，每个样例平均通过的层数将大为减少。假设样例通过各层的概率都为 f ，那么每个样例平均通过的层数为：

$$\begin{aligned} & 1 + f + f^2 + \dots + f^{N-1} \\ = & \frac{1 - f^N}{1 - f} \end{aligned} \quad (4-1)$$

其中 N 为总层数。

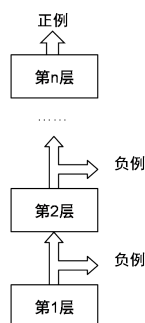


图 4.1 层级分类器原理

而且，处在低级的层，其计算复杂度比高级的层要低得多。假设第 n 层的计算量为 n ，那么如果直接使用最复杂的第 N 层，每个样本的平均计算量为 N 。而如果使用层级分类器，则每个样本的平均计算量为：

$$\begin{aligned}
 & 1 + 2f + 3f^2 + \dots + Nf^{N-1} \\
 = & \frac{1 - f^N}{(1 - f)^2} - \frac{Nf^N}{1 - f} \\
 \approx & \frac{1}{(1 - f)^2} \tag{4-2}
 \end{aligned}$$

下面计算层级分类器的精度。假设每一层的命中率（判断为正例且判断正确的样本数/正例总数）为 h ，每一层的误报率（判断为正例且判断错误的样本数/负例总数）为 f ，那么整个层级分类器的命中率 H 和误报率 F 为：

$$H = h^N \tag{4-3a}$$

$$F = f^N \tag{4-3b}$$

从中可以看出，随着层数的增加，命中率和误报率都会降低。所以，为了保证整个层级分类器有较高的命中率和较低的误报率，需要是 h 足够大，而 f 并不需要很小（因为乘幂后会迅速降低）。本论文设定的 h 为0.995， f 为0.5。

4.1.2 每一层的Haar检测器

对于每一层的分类器的设计，本论文沿用[7]的设计，使用Haar分类器。这种分类器由一系列单特征Haar分类器组成。每个单特征Haar分类器使用自己的特征对样本进行分类，然后整个Haar分类器将这些单特征的分类结果进行加权相加，然后输出总的结果。

Haar分类器的训练是从大量的Haar特征中选出少量，并计算出它们相加时的权。这个过程通过AdaBoost算法完成。AdaBoost算法一般用来将一系列弱分类器组成一个强分类器，但它也可以视为是一种特征选择算法，用来在大量特征中选取少量效果好的进行组合。

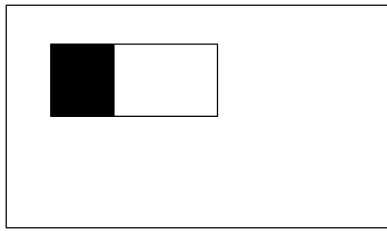


图 4.2 Haar特征示意图

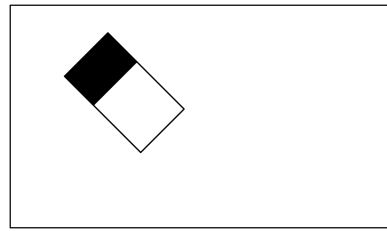


图 4.3 倾斜Haar特征示意图

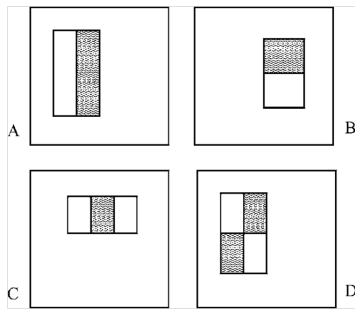


图 4.4 基本Haar特征。引自[7]

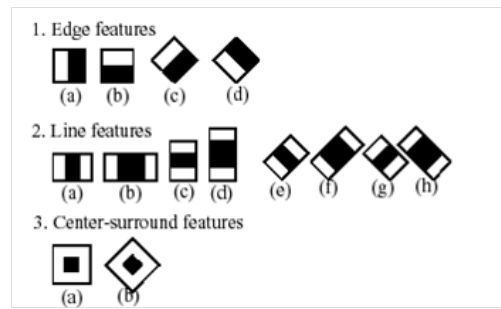


图 4.5 扩展Haar特征。引自[7]

4.1.3 Haar特征

4.1.3.1 基本Haar特征

一个基本的Haar特征，以左右型为例，可以表示成图4.2。

图中外围的大矩形框表示图像，内部的黑白相邻的矩形框表示一个位于特定位置的Haar特征。这个特征的值的计算规则为白色矩形覆盖的图像的像素值和减去黑色矩形覆盖的图像的像素值和。

除了左右型外，基本Haar特征还包括如图4.4所示的几种（包括它们的黑白镜像）。

4.1.3.2 Haar特征的快速算法

Haar特征的计算要用到一个矩形区域的像素值的和，这里有一个快速计算的方法，称为“积分图像法”。设 $I(x, y)$ 表示矩形 $(0, 0) - (x, y)$ （左上角坐标-右下角坐标）的像素和，那么矩形 $(x_1, y_1) - (x_2, y_2)$ 的像素和 $S(x_1, y_1, x_2, y_2)$ 的计算公式为：

$$S(x_1, y_1, x_2, y_2) = I(x_2, y_2) + I(x_1, y_1) - I(x_1, y_2) - I(x_2, y_1) \quad (4-4)$$

由此，只要对每一幅图像计算出它对于的积分图像 $I(x, y)$ ，就可以快速计算出该幅图像中任何位置的Haar特征。

4.1.3.3 倾斜Haar特征

[10]提出了倾斜的Haar特征，如图4.3。

它也可以通过相似的四个定点的快速算法得到。

4.1.4 级层分类器的训练算法

训练算法见算法1。

<p>Input: <i>PosSamples</i>, 训练所用的正例集</p> <p>Input: <i>NegSamples</i>, 训练所用的负例集</p> <p>Input: <i>Fmax</i>, 整个分类器所允许的最大误报率</p> <p>Input: <i>fmax</i>, 每层所允许的最大误报率</p> <p>Input: <i>hmin</i>, 每层所允许的最小命中率</p> <p>Output: <i>Cascade</i>, 层级分类器</p> <pre>1 $F \leftarrow 1$ 2 while $F > Fmax$ do 3 $f \leftarrow 1$ 4 while $f > fmax$ do 5 用Adaboost算法在训练集{<i>PosSamples</i>, <i>NegSamples</i>}上找到一个最好的特征，加入单层分类器 6 调节这个特征的阈值，使得单层分类器的命中率大于<i>hmin</i> 7 $f \leftarrow$ 单层分类器的误报率 8 end 9 将单层分类器加入级层分类器<i>Cascade</i>，作为最后一层 10 $F \leftarrow F \cdot f$ 11 end 12 return $\{S_i\}$</pre>
--

Algorithm 1: 训练层级分类器

4.2 实验结果

4.2.1 只使用基本Haar特征

本实验使用的Haar特征如图4.4所示。实验中一共训练出了24层分类器。在保留不同层的情况下，分类器的表现如下：

表 4.1 基本Haar特征性能（数据集1）

层数	命中率	误报数/帧
10	0.785171	189.768
11	0.764259	101.17
12	0.749049	60.642
13	0.731939	29.0667
14	0.709125	15.0321
15	0.671103	8.31358
16	0.657795	5.06667
17	0.60076	2.91358
18	0.553232	1.86667
19	0.515209	1.21975
20	0.501901	0.837037
21	0.376426	0.269136
22	0.372624	0.237037
23	0.370722	0.22716
24	0.368821	0.212346

表 4.2 基本Haar特征性能（数据集2，第一类）

层数	命中率	误报数/帧
10	0.840498	253.366
11	0.833022	127.609
12	0.801246	61.3773
13	0.793146	36.157
14	0.786293	19.3015
15	0.759502	9.72346
16	0.736449	5.71543
17	0.723988	3.31222
18	0.703427	2.55665
19	0.683489	2.03479

20	0.663551	1.52721
21	0.649221	1.3818
22	0.639875	1.29973

表 4.3 基本Haar特征性能（数据集2，第二类）

层数	命中率	误报数/帧
24	0.612045	0.234084
23	0.612745	0.238002
22	0.632353	0.341822
21	0.647059	0.411361
20	0.656863	0.612145
19	0.67577	0.989226
18	0.705882	1.8952
17	0.733894	3.36435
16	0.75	6.04016
15	0.771008	10.0069
14	0.795518	15.2772
13	0.817227	28.0872
12	0.834034	46.3986
11	0.866947	76.4574
10	0.882353	122.147

检测结果的示例图片如图4.12到图4.13。

4.2.2 使用扩展Haar特征

本实验使用的Haar特征如图4.5所示。实验中一共训练出了20层分类器。在保留不同层的情况下，分类器的表现如下：

表 4.4 扩展Haar特征性能（数据集1）

层数	命中率	误报数/帧
6	0.749049	126.43
7	0.659696	31.1877
8	0.65019	20.3432
9	0.587452	7.4
10	0.54943	2.55802

11	0.479087	0.967901
12	0.45057	0.523457
13	0.410646	0.251852
14	0.39924	0.177778
15	0.389734	0.123457
16	0.380228	0.0888889
17	0.374525	0.0839506
18	0.370722	0.0814815
19	0.368821	0.0765432
20	0.368821	0.0765432

表 4.5 扩展Haar特征性能（数据集2，第一类）

层数	命中率	误报数/帧
12	0.537695	16.7154
13	0.521495	10.8243
14	0.449844	4.18376
15	0.441121	0.786798
16	0.427414	0.768956
17	0.418692	0.728814
18	0.416822	0.719893
19	0.416822	0.718109
20	0.401869	0.654773

表 4.6 扩展Haar特征性能（数据集2，第二类）

层数	命中率	误报数/帧
21	0.581232	0.330069
20	0.588936	0.364349
19	0.598739	0.409403
18	0.621849	0.478942
17	0.641457	0.608227
16	0.663866	0.854065
15	0.67437	1.46327
14	0.719188	2.34574
13	0.727591	3.99608
12	0.759804	9.00294

11	0.769608	16.24
10	0.829832	46.6641

从图4.9和图4.10的结果看，在相同的命中率下，基本特征能够达到比扩展特征更好的效果。也就是说更多的特征选择反而降低了整体的性能。具体原因有待进一步分析。

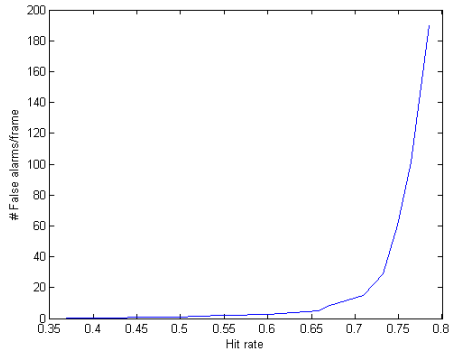


图 4.6 基本Haar特征性能 (数据集1)

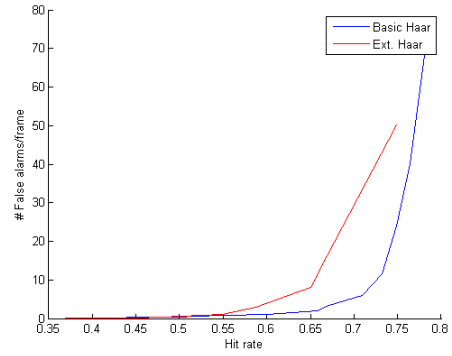


图 4.9 扩展Haar特征性能 (数据集1)

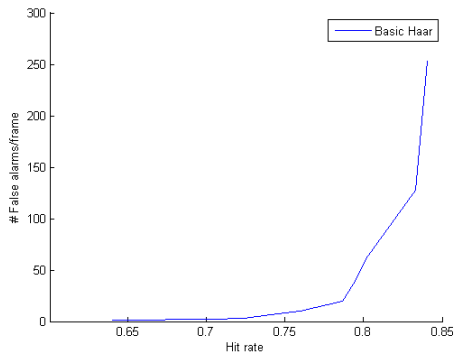


图 4.7 基本Haar特征性能 (数据集2, 第一类)

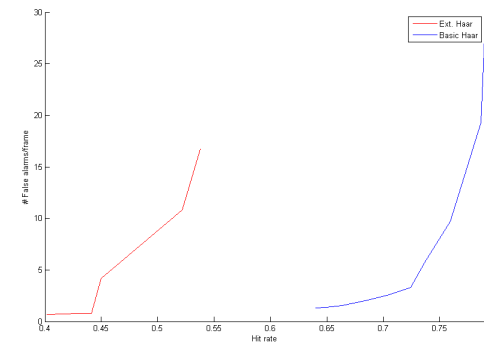


图 4.10 扩展Haar特征性能 (数据集2, 第一类)

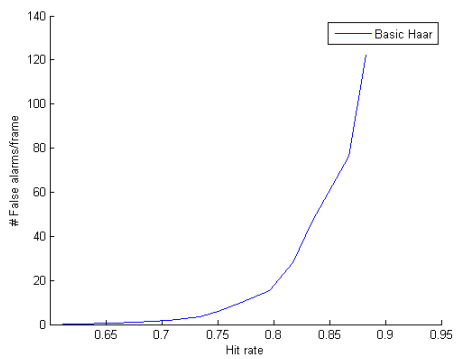


图 4.8 基本Haar特征性能 (数据集2, 第二类)

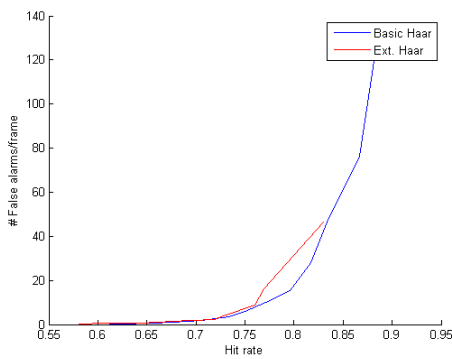


图 4.11 扩展Haar特征性能 (数据集2, 第二类)



图 4.12 基本Haar特征检测结果示例1（数据集1）



图 4.14 基本Haar特征检测结果示例2（数据集1）



图 4.13 基本Haar特征检测结果示例3（数据集2，第一类）



图 4.15 基本Haar特征检测结果示例4（数据集2，第二类）

第5章 运动概率模型

5.1 算法描述

由于车辆以近似匀速直线向前运动，交通标志在图像中的运动也存在一定规律。通过这种规律可以建立概率模型，从而根据之前交通标志的位置统计，对之后的交通标志出现在各位置的概率做出估计。这样可以缩小滑动窗口的搜索范围，或对检测器的输出进行选择 and 过滤。

5.1.1 交通标志的运动公式

驾驶车辆匀速沿 z 轴运动，所有交通标志都与 $x-y$ 平面平行，且其各边平行于 x 或 y 轴。对于每一个交通标志，它在 t 时刻时在图像中的位置，其计算公式为：

$$\begin{cases} x = \frac{X_R\alpha}{\tau-t} + x_0 \\ y = \frac{Y_R\alpha}{\tau-t} + y_0 \\ w = \frac{W_R\alpha}{\tau-t} \\ h = \frac{H_R\alpha}{\tau-t} \end{cases} \quad (5-1)$$

其中， x, y, w, h 分别表示标志的左上角在图片上的横、纵坐标以及标志在图片上的宽和高，这4个值和 t 一同组成了5个观测值。 X_R, Y_R, W_R, H_R 分别标志标志在真实世界中左上角距离摄像机中轴线的横、纵距离以及真实的标志的宽与高。 α 可以视为车辆运动的速度的倒数（需要进行一系列单位换算）， τ 表示从现在时刻到摄像机经过标志所需的时间，它反映了现在时刻标志与摄像机在运动方向的距离。 x_0, y_0 表示视线消失点（即摄像机中轴线与屏幕的交点）在屏幕上的 x, y 坐标。

运动公式的证明如下，以 x 方向为例。如图5.1所示， AB 为摄像机的投影面， F 为摄像机的焦点， FB 为摄像机的中轴线。 E 为沿中轴线运动的一点。各线段的长度为： $|GF| = X_R, |EG| = z, |BF| = d, |BC| = x - x_0$ 。

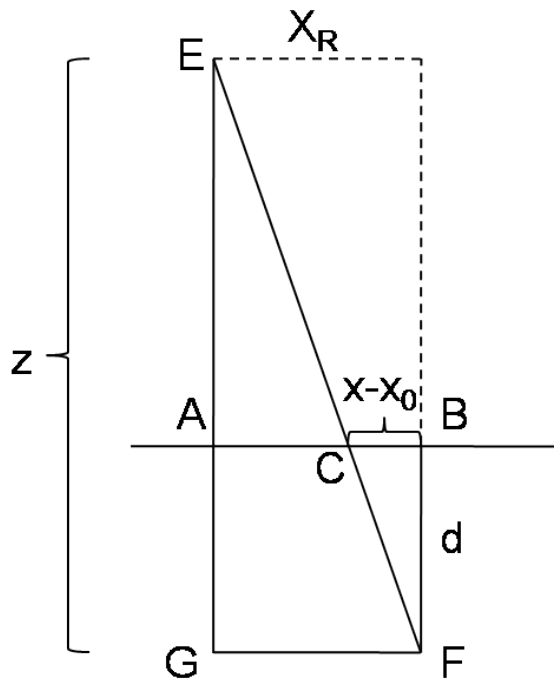


图 5.1 交通标志运动公式证明。 AB 为摄像机的投影面， F 为摄像机的焦点， FB 为摄像机的中轴线。 E 为沿中轴线运动的一点。各线段的长度为： $|GF| = X_R$ ， $|EG| = z$ ， $|BF| = d$ ， $|BC| = x - x_0$

由比例关系易知：

$$\frac{x - x_0}{d} = \frac{X_R}{z} \quad (5-2)$$

设 E 点运动的速度为 v ，它从 E 运动到 G 的时间为 $\tau - t$ ，则由 $z = v(\tau - t)$ 。从而：

$$x = \frac{X_R d}{v(\tau - t)} + x_0 \quad (5-3)$$

证毕。

$X_R, Y_R, W_R, H_R, x_0, y_0$ 可以作为常数预先给定，也可以作为随机变量进行估计。当它们都作为随即变量进行估计时， $X_R, Y_R, W_R, H_R, \alpha$ 可以压缩成四个随机

变量 $\beta_X, \beta_Y, \beta_W, \beta_H$ ，其定义为：

$$\begin{cases} \beta_X = X_R \alpha \\ \beta_Y = Y_R \alpha \\ \beta_W = W_R \alpha \\ \beta_H = H_R \alpha \end{cases} \quad (5-4)$$

5.1.2 最大似然估计

从式5-1中可以看出， x, y, w, h 对于时间 t 的变化曲线都是抛物线，在已知一系列 (x, y, w, h, t) 的情况下，对运动函数各参数的估计，等价于对二参数或三参数正双曲线的估计。我们可以参考 [19]给出一种最大似然的估计方法。

假设有一组数据 $\{(x_n, y_n)\}$ 符合下面的关系：

$$(x + a)(y + b) = c \quad (5-5)$$

我们想通过 $\{(x_n, y_n)\}$ 估计出 a, b, c 三个参数。先将5-5变形为：

$$xy = -bx - ay + c - ab \quad (5-6)$$

令 $z = xy, A = -b, B = -a, C = c - ab$ ，则5-6可写为：

$$z = Ax + By + C \quad (5-7)$$

5-7式是一个标准的二维线性拟合问题。它的一种解法如下：

设误差函数为平方误差，即

$$\mathcal{L} = \sum_n (Ax + By + C - z)^2 \quad (5-8)$$

通过将 \mathcal{L} 对 A, B, C 三者分别取偏导，并令三个偏导均为0，得到 A, B, C 的计算

公式为：

$$\begin{cases} A = \frac{\mathbf{D}(x,z)\mathbf{D}(y,y) - \mathbf{D}(y,z)\mathbf{D}(x,y)}{\mathbf{D}(x,x)\mathbf{D}(y,y) - \mathbf{D}(x,y)\mathbf{D}(x,y)} \\ B = \frac{\mathbf{D}(y,z) - A\mathbf{D}(x,y)}{\mathbf{D}(y,y)} \\ C = \frac{\sum_n z - A \sum_n x - B \sum_n y}{n} \end{cases} \quad (5-9)$$

其中，

$$\mathbf{D}(x,y) \equiv \sum_n xy - \frac{\sum_n x \sum_n y}{n} \quad (5-10)$$

相关系数的计算为：

$$r = \sqrt{1 - \frac{\mathbf{D}(z,z) - A\mathbf{D}(x,z) - B\mathbf{D}(y,z)}{\mathbf{D}(z,z)}} \quad (5-11)$$

进一步可得到 a,b,c 的计算公式。

5.1.3 概率图模型

本论文建立的运动模型可以用概率图模型表示，见图5.2。

图中的似然函数为：

$$\begin{aligned} h(\alpha, X_R, Y_R, W_R, H_R, x_0, y_0) &= P(x, y, w, h, t | \alpha, X_R, Y_R, W_R, H_R, x_0, y_0) \\ &= \mathcal{N}(xt | \tau x + x_0 t - \beta_X - \tau x_0, \Lambda_x^{-1}) \\ &\quad \mathcal{N}(yt | \tau y + y_0 t - \beta_Y - \tau y_0, \Lambda_y^{-1}) \\ &\quad \mathcal{N}(wt | \tau w - \beta_W, \Lambda_w^{-1}) \mathcal{N}(ht | \tau h - \beta_H, \Lambda_h^{-1}) \end{aligned} \quad (5-12)$$

其中 $\mathcal{N}(x|\mu, \sigma^2)$ 表示 x 服从期望为 μ 标准差为 σ 的高斯分布。式中的 $\Lambda_x, \Lambda_y, \Lambda_w, \Lambda_h$ 分别表示 x, y, w, h 四个观测值的准确度，是方差的倒数。 m, n 等下标忽略了。

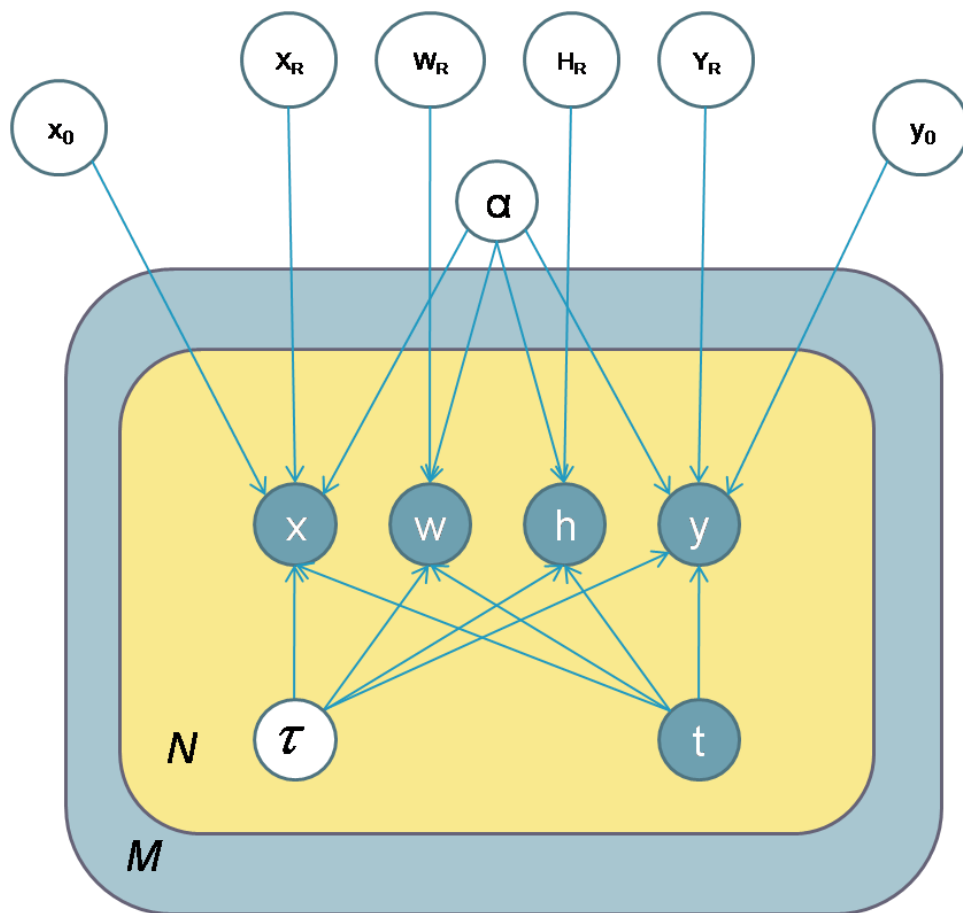


图 5.2 概率图模型。实心圆圈表示观测值，空心圆圈表示待推理的随机变量。圆角矩形表示维度板，维度在左下角标出。维度板表示其内部的每个圆圈代表D个随机变量，且这N个随机变量互相间没有连线，D为维度。

5.1.4 推理算法

对这个概率图模型的推理方法有许多，例如sum-product、Belief Propagation、Sampling等等。各算法的精确度和运算量不同。这里本论文使用Variational Inference框架^[20]推导出一个相对比较快速的推理算法。

记 $\mathbf{X} = \{(x_{m,n_m}, y_{m,n_m}, w_{m,n_m}, h_{m,n_m}, t_{m,n_m}) | 1 \leq m \leq M, 1 \leq n_m \leq N_m\}$ ， M 为交通标志的个数， N_m 为第 m 号交通标志的观测个数（即历史轨迹的长度）；

记 $\mathbf{Z} = \{\beta_X, \beta_Y, \beta_W, \beta_H, x_0, y_0, \tau_m | 1 \leq m \leq M\}$

由5-12知：

$$\begin{aligned} P(\mathbf{X}|\mathbf{Z}) &= \prod_{1 \leq m \leq M} \prod_{1 \leq n \leq N_m} \mathcal{N}(x_{m,n} t_{m,n} | \tau x_{m,n} + x_0 t_{m,n} - \beta_X - \tau x_0, \Lambda_x^{-1}) \\ &\quad \mathcal{N}(y_{m,n} t_{m,n} | \tau y_{m,n} + y_0 t_{m,n} - \beta_Y - \tau y_0, \Lambda_y^{-1}) \\ &\quad \mathcal{N}(w_{m,n} t_{m,n} | \tau w_{m,n} - \beta_W, \Lambda_w^{-1}) \mathcal{N}(h_{m,n} t_{m,n} | \tau h_{m,n} - \beta_H, \Lambda_h^{-1}) \end{aligned} \quad (5-13)$$

我们推理的目的求出 \mathbf{Z} 的后验概率，即 $P(\mathbf{Z}|\mathbf{X})$ 。令 $q(\mathbf{Z})$ 为对 $P(\mathbf{Z}|\mathbf{X})$ 的估计。

现在做两点假设：

(1) $q(\mathbf{Z})$ 存在如下分解：

$$q(\mathbf{Z}) = q(\beta_X, \beta_Y, \beta_W, \beta_H) q(x_0, y_0) q(\{\tau_m\})$$

(2) \mathbf{Z} 的先验 $P(\mathbf{Z})$ 为：

$$\begin{aligned} P(\mathbf{Z}) &= P(\beta_X) P(\beta_Y) P(\beta_W) P(\beta_H) P(x_0) P(y_0) \prod_{1 \leq m \leq M} P(\tau_m) \\ &= \mathcal{N}(\beta_X | \mu_{\beta_X}, \Lambda_{\beta_X}^{-1}) \mathcal{N}(\beta_Y | \mu_{\beta_Y}, \Lambda_{\beta_Y}^{-1}) \mathcal{N}(\beta_W | \mu_{\beta_W}, \Lambda_{\beta_W}^{-1}) \mathcal{N}(\beta_H | \mu_{\beta_H}, \Lambda_{\beta_H}^{-1}) \\ &\quad \mathcal{N}(x_0 | \mu_{x_0}, \Lambda_{x_0}^{-1}) \mathcal{N}(y_0 | \mu_{y_0}, \Lambda_{y_0}^{-1}) \prod_{1 \leq m \leq M} \mathcal{N}(\tau_m | \mu_{\tau_m}, \Lambda_{\tau_m}^{-1}) \end{aligned} \quad (5-14)$$

那么，根据[20]，对 $P(\mathbf{Z}|\mathbf{X})$ 的最佳估计 $q^*(\mathbf{Z})$ 可以表示成：

$$q^*(\mathbf{Z}) = q^*(\beta_X) q^*(\beta_Y) q^*(\beta_W) q^*(\beta_H) q^*(x_0) q^*(y_0) \prod_{1 \leq m \leq M} q^*(\tau_m) \quad (5-15)$$

其中，对于 $q^*(\beta_X)$,

$$\begin{aligned}
\ln q^*(\beta_X) &= \mathbf{E}_{\mathbf{Z}-\beta_X}[\ln P(\mathbf{X}, \mathbf{Z})] + \text{const} \\
&= \mathbf{E}_{\mathbf{Z}-\beta_X}[\ln P(\mathbf{X}|\mathbf{Z}) + \ln P(\mathbf{Z})] + \text{const} \\
&= \mathbf{E}_{\mathbf{Z}-\beta_X}[\ln P(\mathbf{X}|\mathbf{Z}) + \ln P(\beta_X)] + \text{const} \\
&= \mathbf{E}_{\mathbf{Z}-\beta_X}[\Lambda_x \sum_{1 \leq m \leq M} \sum_{1 \leq n \leq N_m} ((\tau_m - t_{m,n})(x_{m,n} - x_0) - \beta_X)^2 + \Lambda_{\beta_X}(\beta_X - \mu_{\beta_X})^2] + \text{const}
\end{aligned} \tag{5-16}$$

即：

$$q^*(\beta_X) = \mathcal{N}(\beta_X | \mu_{\beta_X}^*, \Lambda_{\beta_X}^{*-1}) \tag{5-17}$$

其中，

$$\Lambda_{\beta_X}^* = D\Lambda_X + \Lambda_{\beta_X} \tag{5-18}$$

$$\mu_{\beta_X}^* = \frac{\Lambda_x \sum_{1 \leq m \leq M} \sum_{1 \leq n \leq N_m} (\mathbf{E}[\tau_m]x_{m,n} + \mathbf{E}[x_0]t_{m,n} - \mathbf{E}[\tau_m]\mathbf{E}[x_0] - x_{m,n}t_{m,n}) + \Lambda_{\beta_X}\mu_{\beta_X}}{\Lambda_{\beta_X}^*} \tag{5-19}$$

其中，

$$D = \sum_{1 \leq m \leq M} N_m \tag{5-20}$$

为所有观测数。

类似地，对于 $q^*(\beta_Y)$,

$$q^*(\beta_Y) = \mathcal{N}(\beta_Y | \mu_{\beta_Y}^*, \Lambda_{\beta_Y}^{*-1}) \tag{5-21}$$

其中，

$$\Lambda_{\beta_Y}^* = D\Lambda_Y + \Lambda_{\beta_Y} \tag{5-22}$$

$$\mu_{\beta_Y}^* = \frac{\Lambda_Y \sum_{1 \leq m \leq M} \sum_{1 \leq n \leq N_m} (\mathbf{E}[\tau_m] y_{m,n} + \mathbf{E}[y_0] t_{m,n} - \mathbf{E}[\tau_m] \mathbf{E}[y_0] - y_{m,n} t_{m,n}) + \Lambda_{\beta_Y} \mu_{\beta_Y}}{\Lambda_{\beta_Y}^*} \quad (5-23)$$

对于 $q^*(\beta_W)$,

$$q^*(\beta_W) = \mathcal{N}(\beta_W | \mu_{\beta_W}^*, \Lambda_{\beta_W}^{*, -1}) \quad (5-24)$$

其中,

$$\Lambda_{\beta_W}^* = D\Lambda_W + \Lambda_{\beta_W} \quad (5-25)$$

$$\mu_{\beta_W}^* = \frac{\Lambda_W \sum_{1 \leq m \leq M} \sum_{1 \leq n \leq N_m} (\mathbf{E}[\tau_m] w_{m,n} - w_{m,n} t_{m,n}) + \Lambda_{\beta_W} \mu_{\beta_W}}{\Lambda_{\beta_W}^*} \quad (5-26)$$

对于 $q^*(\beta_H)$,

$$q^*(\beta_H) = \mathcal{N}(\beta_H | \mu_{\beta_H}^*, \Lambda_{\beta_H}^{*, -1}) \quad (5-27)$$

其中,

$$\Lambda_{\beta_H}^* = D\Lambda_H + \Lambda_{\beta_H} \quad (5-28)$$

$$\mu_{\beta_H}^* = \frac{\Lambda_h \sum_{1 \leq m \leq M} \sum_{1 \leq n \leq N_m} (\mathbf{E}[\tau_m] h_{m,n} - h_{m,n} t_{m,n}) + \Lambda_{\beta_H} \mu_{\beta_H}}{\Lambda_{\beta_H}^*} \quad (5-29)$$

对于 $q^*(x_0)$,

$$q^*(x_0) = \mathcal{N}(x_0 | \mu_{x_0}^*, \Lambda_{x_0}^{*, -1}) \quad (5-30)$$

其中,

$$\Lambda_{x_0}^* = \Lambda_x \sum_{1 \leq m \leq M} \sum_{1 \leq n \leq N_m} (\mathbf{E}[\tau_m^2] - 2t_{m,n} \mathbf{E}[\tau_m] + t_{m,n}^2) + \Lambda_{x_0} \quad (5-31)$$

$$\mu_{x_0}^* = \frac{\Lambda_x \sum_{1 \leq m \leq M} \sum_{1 \leq n \leq N_m} (\mathbf{E}[\tau_m^2] x_{m,n} - \mathbf{E}[\tau_m] \mathbf{E}[\beta_X] - 2\mathbf{E}[\tau_m] x_{m,n} t_{m,n} + \mathbf{E}[\beta_X] t_{m,n} + x_{m,n} t_{m,n}^2) + \Lambda_{x_0} \mu_{x_0}}{\Lambda_{x_0}^*} \quad (5-32)$$

对于 $q^*(y_0)$,

$$q^*(y_0) = \mathcal{N}(y_0 | \mu_{y_0}^*, \Lambda_{y_0}^{*, -1}) \quad (5-33)$$

其中,

$$\Lambda_{y_0}^* = \Lambda_y \sum_{1 \leq m \leq M} \sum_{1 \leq n \leq N_m} (\mathbf{E}[\tau_m^2] - 2t_{m,n} \mathbf{E}[\tau_m] + t_{m,n}^2) + \Lambda_{y_0} \quad (5-34)$$

$$\mu_{y_0}^* = \frac{\Lambda_y \sum_{1 \leq m \leq M} \sum_{1 \leq n \leq N_m} (\mathbf{E}[\tau_m^2] y_{m,n} - \mathbf{E}[\tau_m] \mathbf{E}[\beta_Y] - 2\mathbf{E}[\tau_m] y_{m,n} t_{m,n} + \mathbf{E}[\beta_Y] t_{m,n} + y_{m,n} t_{m,n}^2) + \Lambda_{y_0} \mu_{y_0}}{\Lambda_{y_0}^*} \quad (5-35)$$

对于 $q^*(\tau_m)$,

$$q^*(\tau_m) = \mathcal{N}(\tau_m | \mu_{\tau_m}^*, \Lambda_{\tau_m}^{*, -1}) \quad (5-36)$$

其中,

$$\begin{aligned} \Lambda_{\tau_m}^* &= \Lambda_x \sum_{1 \leq n \leq N_m} (\mathbf{E}[x_0^2] - 2x_{m,n} \mathbf{E}[x_0] + x_{m,n}^2) + \Lambda_y \sum_{1 \leq n \leq N_m} (\mathbf{E}[y_0^2] - 2y_{m,n} \mathbf{E}[y_0] + y_{m,n}^2) \\ &\quad + \Lambda_w \sum_{1 \leq n \leq N_m} w_{m,n}^2 + \Lambda_h \sum_{1 \leq n \leq N_m} h_{m,n}^2 + \Lambda_{\tau_m} \end{aligned} \quad (5-37)$$

$$\mu_{\tau_m}^* = \frac{\Lambda_x \sum_{1 \leq n \leq N_m} (\mathbf{E}[x_0^2] t_{m,n} - \mathbf{E}[x_0] \mathbf{E}[\beta_X] - 2\mathbf{E}[x_0] x_{m,n} t_{m,n} + \mathbf{E}[\beta_X] x_{m,n} + t_{m,n} x_{m,n}^2)}{\Lambda_{\tau_m}^{*, -1}}$$

$$\begin{aligned}
& + \frac{\Lambda_y \sum_{1 \leq n \leq N_m} (\mathbf{E}[y_0^2] t_{m,n} - \mathbf{E}[y_0] \mathbf{E}[\beta_Y] - 2\mathbf{E}[y_0] y_{m,n} t_{m,n} + \mathbf{E}[\beta_Y] y_{m,n} + t_{m,n} y_{m,n}^2)}{\Lambda_{\tau_m}^{*, -1}} \\
& + \frac{\Lambda_w \sum_{1 \leq n \leq N_m} (\mathbf{E}[\beta_W] w_{m,n} + w_{m,n}^2 t_{m,n})}{\Lambda_{\tau_m}^{*, -1}} \\
& + \frac{\Lambda_h \sum_{1 \leq n \leq N_m} (\mathbf{E}[\beta_H] h_{m,n} + h_{m,n}^2 t_{m,n})}{\Lambda_{\tau_m}^{*, -1}} \\
& + \frac{\Lambda_{\tau_m} \mu_{\tau_m}}{\Lambda_{\tau_m}^*} \tag{5-38}
\end{aligned}$$

(以上式子中有两个不同的变量的乘积的期望，如： $\mathbf{E}[\tau_m x_0]$ ，之所以能够拆分成 $\mathbf{E}[\tau_m] \mathbf{E}[x_0]$ ，是因为 τ_m 和 x_0 不相关。

另外， $\mathbf{E}[\tau_m^2] = \mathbf{E}^2[\tau_m] + \Lambda_{\tau_m}^{-1}$)

从5-17到5-38可以看出，要计算某一个变量的分布，需要知道其它变量的期望。所以它们互相耦合，没有显式解。但我们可以通过迭代来求解，即用当前计算出的分布得到需要的期望，从而更新某一个变量的分布，再用更新后的分布计算期望，去更新其它变量的分布。注意这种更新需要是“异步”的，即某个变量的分布得到更新后需要马上应用到后续其它变量的分布，而不能等到一轮后再投入使用。

迭代算法的收敛性由[20]保证。

5.1.5 在线版本的推理算法

假设我们拥有了一系列观测值 $\{\mathbf{X}_1, \dots, \mathbf{X}_n\}$ ，想要推测 \mathbf{Z} 的后验概率。可以推出：

$$\begin{aligned}
P(\mathbf{Z}|\mathbf{X}_1, \dots, \mathbf{X}_n) &= \frac{P(\mathbf{X}_1, \dots, \mathbf{X}_n|\mathbf{Z})P(\mathbf{Z})}{P(\mathbf{X}_1, \dots, \mathbf{X}_n)} \\
&= \frac{P(\mathbf{X}_1, \dots, \mathbf{X}_{n-1}|\mathbf{Z})P(\mathbf{X}_n|\mathbf{X}_1, \dots, \mathbf{X}_{n-1}, \mathbf{Z})P(\mathbf{Z})}{P(\mathbf{X}_1, \dots, \mathbf{X}_n)} \\
&= P(\mathbf{Z}|\mathbf{X}_1, \dots, \mathbf{X}_{n-1}) \frac{P(\mathbf{X}_1, \dots, \mathbf{X}_{n-1})}{P(\mathbf{Z})} \frac{P(\mathbf{X}_n|\mathbf{X}_1, \dots, \mathbf{X}_{n-1}, \mathbf{Z})}{P(\mathbf{X}_1, \dots, \mathbf{X}_n)} P(\mathbf{Z}) \\
&\propto P(\mathbf{Z}|\mathbf{X}_1, \dots, \mathbf{X}_{n-1})P(\mathbf{X}_n|\mathbf{X}_1, \dots, \mathbf{X}_{n-1}, \mathbf{Z}) \\
&= P(\mathbf{Z}|\mathbf{X}_1, \dots, \mathbf{X}_{n-1})P(\mathbf{X}_n|\mathbf{Z}) \tag{5-39}
\end{aligned}$$

(式中的 \propto 是指与 \mathbf{Z} 无关的项都忽略。最后一步的根据是 \mathbf{X}_n 与 $\{\mathbf{X}_1, \dots, \mathbf{X}_{n-1}\}$ 关于 \mathbf{Z} 条件独立。)

对于上式，我们可以把 $P(\mathbf{Z}|\mathbf{X}_1, \dots, \mathbf{X}_{n-1})$ 理解成观测 \mathbf{X}_n 前的先验，而把 $P(\mathbf{X}_n|\mathbf{Z})$ 看成观测 \mathbf{X}_n 的似然函数。由此，上式指出我们可以设计一个在线的推理算法，每观测一个新数据时，都把基于以前的所有观测的后验概率当做先验，而把产生这个新数据的条件概率作为似然函数。

在线算法如下：

设在观测到 \mathbf{X}_n 之前，

$$\begin{aligned}
P(\beta_X|\mathbf{X}_1, \dots, \mathbf{X}_{n-1}) &= \mathcal{N}(\beta_X|\mu_{\beta_X}, \Lambda_{\beta_X}^{-1}) \\
P(\beta_Y|\mathbf{X}_1, \dots, \mathbf{X}_{n-1}) &= \mathcal{N}(\beta_Y|\mu_{\beta_Y}, \Lambda_{\beta_Y}^{-1}) \\
P(\beta_W|\mathbf{X}_1, \dots, \mathbf{X}_{n-1}) &= \mathcal{N}(\beta_W|\mu_{\beta_W}, \Lambda_{\beta_W}^{-1}) \\
P(\beta_H|\mathbf{X}_1, \dots, \mathbf{X}_{n-1}) &= \mathcal{N}(\beta_H|\mu_{\beta_H}, \Lambda_{\beta_H}^{-1}) \\
P(x_0|\mathbf{X}_1, \dots, \mathbf{X}_{n-1}) &= \mathcal{N}(x_0|\mu_{x_0}, \Lambda_{x_0}^{-1}) \\
P(y_0|\mathbf{X}_1, \dots, \mathbf{X}_{n-1}) &= \mathcal{N}(y_0|\mu_{y_0}, \Lambda_{y_0}^{-1}) \\
P(\tau_m|\mathbf{X}_1, \dots, \mathbf{X}_{n-1}) &= \mathcal{N}(\tau_m|\mu_{\tau_m}, \Lambda_{\tau_m}^{-1})
\end{aligned} \tag{5-40}$$

在观测到 \mathbf{X}_n 之后，

$$\begin{aligned}
P(\beta_X|\mathbf{X}_1, \dots, \mathbf{X}_n) &= \mathcal{N}(\beta_X|\mu_{\beta_X}^*, \Lambda_{\beta_X}^{*-1}) \\
P(\beta_Y|\mathbf{X}_1, \dots, \mathbf{X}_n) &= \mathcal{N}(\beta_Y|\mu_{\beta_Y}^*, \Lambda_{\beta_Y}^{*-1}) \\
P(\beta_W|\mathbf{X}_1, \dots, \mathbf{X}_n) &= \mathcal{N}(\beta_W|\mu_{\beta_W}^*, \Lambda_{\beta_W}^{*-1}) \\
P(\beta_H|\mathbf{X}_1, \dots, \mathbf{X}_n) &= \mathcal{N}(\beta_H|\mu_{\beta_H}^*, \Lambda_{\beta_H}^{*-1}) \\
P(x_0|\mathbf{X}_1, \dots, \mathbf{X}_n) &= \mathcal{N}(x_0|\mu_{x_0}^*, \Lambda_{x_0}^{*-1}) \\
P(y_0|\mathbf{X}_1, \dots, \mathbf{X}_n) &= \mathcal{N}(y_0|\mu_{y_0}^*, \Lambda_{y_0}^{*-1}) \\
P(\tau_m|\mathbf{X}_1, \dots, \mathbf{X}_n) &= \mathcal{N}(\tau_m|\mu_{\tau_m}^*, \Lambda_{\tau_m}^{*-1})
\end{aligned} \tag{5-41}$$

则有，

$$\left\{ \begin{aligned}
 \Lambda_{\beta_X}^* &= \Lambda_x + \Lambda_{\beta_X} \\
 \mu_{\beta_X}^* &= \frac{\Lambda_x(\mu_{\tau_m}^* - t)(x - \mu_{x_0}^*) + \mu_{\beta_X} \Lambda_{\beta_X}}{\Lambda_{\beta_X}^*} \\
 \Lambda_{\beta_Y}^* &= \Lambda_y + \Lambda_{\beta_Y} \\
 \mu_{\beta_Y}^* &= \frac{\Lambda_y(\mu_{\tau_m}^* - t)(y - \mu_{y_0}^*) + \mu_{\beta_Y} \Lambda_{\beta_Y}}{\Lambda_{\beta_Y}^*} \\
 \Lambda_{\beta_W}^* &= \Lambda_w + \Lambda_{\beta_W} \\
 \mu_{\beta_W}^* &= \frac{\Lambda_w w(\mu_{\tau_m}^* - t) + \mu_{\beta_W} \Lambda_{\beta_W}}{\Lambda_{\beta_W}^*} \\
 \Lambda_{\beta_H}^* &= \Lambda_h + \Lambda_{\beta_H} \\
 \mu_{\beta_H}^* &= \frac{\Lambda_h h(\mu_{\tau_m}^* - t) + \mu_{\beta_H} \Lambda_{\beta_H}}{\Lambda_{\beta_H}^*} \\
 \Lambda_{x_0}^* &= \Lambda_x(\mu_{\tau_m}^* - t)^2 + \frac{\Lambda_x}{\Lambda_{\tau_m}^*} + \Lambda_{x_0} \\
 \mu_{x_0}^* &= \frac{\Lambda_x[(\mu_{\tau_m}^* - t)^2 + \frac{1}{\Lambda_{\tau_m}^*} - \mu_{\beta_X}^*(\mu_{\tau_m}^* - t)] + \mu_{x_0} \Lambda_{x_0}}{\Lambda_{x_0}^*} \\
 \Lambda_{y_0}^* &= \Lambda_y(\mu_{\tau_m}^* - t)^2 + \frac{\Lambda_y}{\Lambda_{\tau_m}^*} + \Lambda_{y_0} \\
 \mu_{y_0}^* &= \frac{\Lambda_y[(\mu_{\tau_m}^* - t)^2 + \frac{1}{\Lambda_{\tau_m}^*} - \mu_{\beta_Y}^*(\mu_{\tau_m}^* - t)] + \mu_{y_0} \Lambda_{y_0}}{\Lambda_{y_0}^*} \\
 \Lambda_{\tau_m}^* &= \Lambda_x(x - \mu_{x_0}^*)^2 + \frac{\Lambda_x}{\Lambda_{x_0}^*} + \Lambda_y(y - \mu_{y_0}^*)^2 + \frac{\Lambda_y}{\Lambda_{y_0}^*} + \Lambda_w w^2 + \Lambda_h h^2 + \Lambda_{\tau_m} \\
 \mu_{\tau_m}^* &= \frac{\Lambda_x[(x - \mu_{x_0}^*)^2 + \frac{1}{\Lambda_{x_0}^*} + \mu_{\beta_X}^*(x - \mu_{x_0}^*)] + \Lambda_y[(y - \mu_{y_0}^*)^2 + \frac{1}{\Lambda_{y_0}^*} + \mu_{\beta_Y}^*(y - \mu_{y_0}^*)] + \Lambda_w(w^2 + \mu_{\beta_W}^* w) + \Lambda_h(h^2 + \mu_{\beta_H}^* h) + \mu_{\tau_m} \Lambda_{\tau_m}}{\Lambda_{\tau_m}^*}
 \end{aligned} \right. \quad (5-42)$$

5.1.6 混合高斯扩展

以上两个模型，无论是离线版本或是在线版本，都假设交通标志的大小和相对于摄像机主轴的位置（即 X_R, Y_R, W_R, H_R ）是固定的。但在实际使用中，交通标志的大小可能变化，位置也可能变化。我们称 X_R, Y_R, W_R, H_R 的一个赋值为一个“通道”，那么我们需要解决多通道的问题。本论文使用混合高斯模型，引入随即变量 $\{\pi_k\}$ 和 $\{Z_m\}$ ，前者表示混合高斯模型中每个高斯组件所占的比重，后者表示对于产生每个标志，每个高斯组件所承担的责任。 $\{\pi_k\}$ 的和为1。 $\{Z_m\}$ 中的每个变量是都是 k 维，每一维用 $Z_m^{(k)}$ 表示，每一维都是0-1变量。且有 $\sum_k Z_m^{(k)} = 1$ 。

扩展后的概率图模型如图5.3。

其中的似然函数为：

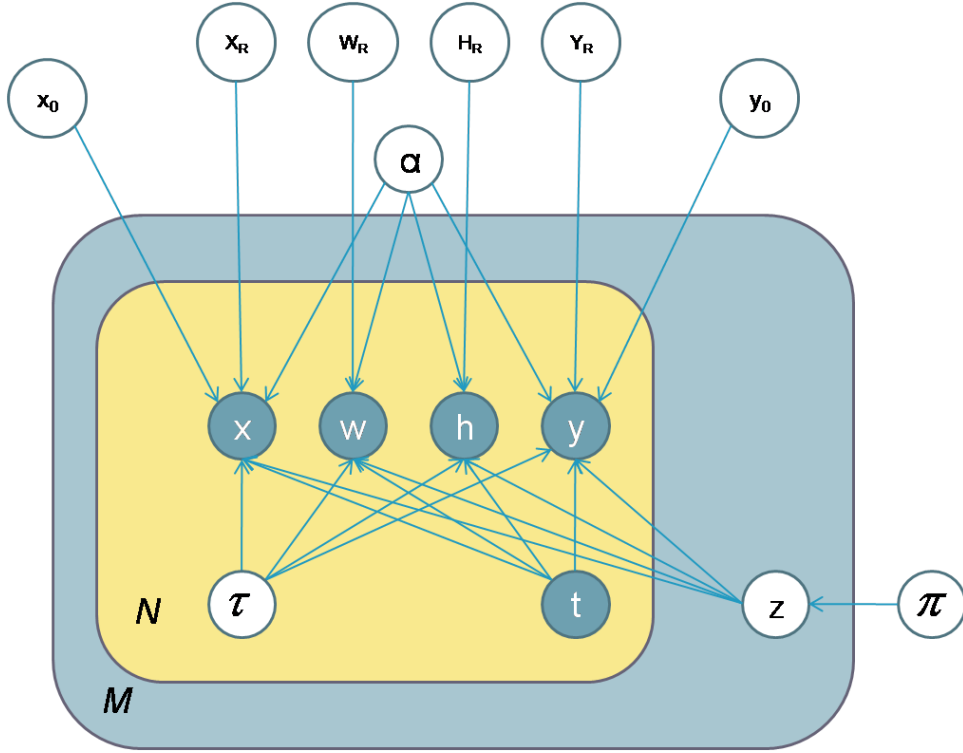


图 5.3 概率图模型。实心圆圈表示观测值，空心圆圈表示待推理的随机变量。圆角矩形表示维度板，维度在左下角标出。维度板表示其内部的每个圆圈代表D个随机变量，且这N个随机变量互相间没有连线，D为维度。

$$\begin{aligned}
 h(\alpha, X_R, Y_R, W_R, H_R, x_0, y_0) &= P(x, y, w, h, t | \alpha, X_R, Y_R, W_R, H_R, x_0, y_0) \\
 &= \prod_{1 \leq k \leq K} [\mathcal{N}(xt | \tau x + x_0 t - \alpha X_R^{(k)} - \tau x_0, \Lambda_x^{-1}) \\
 &\quad \mathcal{N}(yt | \tau y + y_0 t - \alpha Y_R^{(k)} - \tau y_0, \Lambda_y^{-1}) \\
 &\quad \mathcal{N}(wt | \tau w - \alpha W_R^{(k)}, \Lambda_w^{-1}) \mathcal{N}(ht | \tau h - \alpha H_R^{(k)}, \Lambda_h^{-1})]^{Z^{(k)}}
 \end{aligned} \tag{5-43}$$

$$\begin{aligned}
 h(\pi) &= P(Z | \pi) \\
 &= \prod_{1 \leq k \leq K} \pi_k^{Z^{(k)}}
 \end{aligned} \tag{5-44}$$

其中K为通道数， X_R, Y_R, W_R, H_R 都是k维变量，每一维用 $X_R^{(k)}$ 等表示。 m, n 等下标忽略了。

如果我们以 $\beta_x, \beta_y, \beta_w, \beta_h$ 作为一组通道，那么似然函数可以写为：

$$\begin{aligned}
h(\alpha, X_R, Y_R, W_R, H_R, x_0, y_0) &= P(x, y, w, h, t | \alpha, X_R, Y_R, W_R, H_R, x_0, y_0) \\
&= \prod_{1 \leq k \leq K} [\mathcal{N}(xt | \tau x + x_0 t - \beta_X^{(k)} - \tau x_0, \Lambda_x^{-1}) \\
&\quad \mathcal{N}(yt | \tau y + y_0 t - \beta_Y^{(k)} - \tau y_0, \Lambda_y^{-1}) \\
&\quad \mathcal{N}(wt | \tau w - \beta_W^{(k)}, \Lambda_w^{-1}) \mathcal{N}(ht | \tau h - \beta_H^{(k)}, \Lambda_h^{-1})]^{Z^{(k)}} \quad (5-45)
\end{aligned}$$

$\beta_X, \beta_Y, \beta_W, \beta_H$ 都是 k 维变量，每一维用 $\beta_X^{(k)}$ 等表示。

混合高斯模型的推导有很多种方法，这里我们使用EM算法，5-17到5-38 各式经修改后变为如下：

$$q^*(\beta_X^{(k)}) = \mathcal{N}(\beta_X | \mu_{\beta_X^{(k)}}^*, \Lambda_{\beta_X^{(k)}}^{*, -1}) \quad (5-46)$$

$$\Lambda_{\beta_X^{(k)}}^* = \Lambda_X \sum_{1 \leq m \leq M} \sum_{1 \leq k \leq K} N_m \mathbf{E}[Z_m^{(k)}] + \Lambda_{\beta_X} \quad (5-47)$$

$$\mu_{\beta_X^{(k)}}^* = \frac{\Lambda_X \sum_{1 \leq m \leq M} \sum_{1 \leq n \leq N_m} \mathbf{E}[Z_m^{(k)}] (\mathbf{E}[\tau_m] x_{m,n} + \mathbf{E}[x_0] t_{m,n} - \mathbf{E}[\tau_m] \mathbf{E}[x_0] - x_{m,n} t_{m,n}) + \Lambda_{\beta_X} \mu_{\beta_X}}{\Lambda_{\beta_X^{(k)}}^*} \quad (5-48)$$

$$q^*(\beta_Y^{(k)}) = \mathcal{N}(\beta_Y | \mu_{\beta_Y^{(k)}}^*, \Lambda_{\beta_Y^{(k)}}^{*, -1}) \quad (5-49)$$

$$\Lambda_{\beta_Y^{(k)}}^* = \Lambda_Y \sum_{1 \leq m \leq M} \sum_{1 \leq k \leq K} N_m \mathbf{E}[Z_m^{(k)}] + \Lambda_{\beta_Y} \quad (5-50)$$

$$\mu_{\beta_Y^{(k)}}^* = \frac{\Lambda_Y \sum_{1 \leq m \leq M} \sum_{1 \leq n \leq N_m} \mathbf{E}[Z_m^{(k)}] (\mathbf{E}[\tau_m] y_{m,n} + \mathbf{E}[y_0] t_{m,n} - \mathbf{E}[\tau_m] \mathbf{E}[y_0] - y_{m,n} t_{m,n}) + \Lambda_{\beta_Y} \mu_{\beta_Y}}{\Lambda_{\beta_Y^{(k)}}^*} \quad (5-51)$$

$$q^*(\beta_W^{(k)}) = \mathcal{N}(\beta_W | \mu_{\beta_W^{(k)}}^*, \Lambda_{\beta_W^{(k)}}^{*, -1}) \quad (5-52)$$

$$\Lambda_{\beta_W}^* = \Lambda_W \sum_{1 \leq m \leq M} \sum_{1 \leq k \leq K} N_m \mathbf{E}[Z_m^{(k)}] + \Lambda_{\beta_W} \quad (5-53)$$

$$\mu_{\beta_W}^* = \frac{\Lambda_W \sum_{1 \leq m \leq M} \sum_{1 \leq n \leq N_m} \mathbf{E}[Z_m^{(k)}] (\mathbf{E}[\tau_m] w_{m,n} - w_{m,n} t_{m,n}) + \Lambda_{\beta_W} \mu_{\beta_W}}{\Lambda_{\beta_W}^*} \quad (5-54)$$

$$q^*(\beta_H^{(k)}) = \mathcal{N}(\beta_H | \mu_{\beta_H}^*, \Lambda_{\beta_H}^{*, -1}) \quad (5-55)$$

$$\Lambda_{\beta_H}^* = \Lambda_H \sum_{1 \leq m \leq M} \sum_{1 \leq k \leq K} N_m \mathbf{E}[Z_m^{(k)}] + \Lambda_{\beta_H} \quad (5-56)$$

$$\mu_{\beta_H}^* = \frac{\Lambda_H \sum_{1 \leq m \leq M} \sum_{1 \leq n \leq N_m} \mathbf{E}[Z_m^{(k)}] (\mathbf{E}[\tau_m] h_{m,n} - h_{m,n} t_{m,n}) + \Lambda_{\beta_H} \mu_{\beta_H}}{\Lambda_{\beta_H}^*} \quad (5-57)$$

$$q^*(x_0) = \mathcal{N}(x_0 | \mu_{x_0}^*, \Lambda_{x_0}^{*, -1}) \quad (5-58)$$

$$\Lambda_{x_0}^* = \Lambda_x \sum_{1 \leq m \leq M} \sum_{1 \leq n \leq N_m} (\mathbf{E}[\tau_m^2] - 2t_{m,n} \mathbf{E}[\tau_m] + t_{m,n}^2) + \Lambda_{x_0} \quad (5-59)$$

$$\begin{aligned} \mu_{x_0}^* &= \frac{\Lambda_x \sum_{1 \leq m \leq M} \sum_{1 \leq n \leq N_m} \sum_{1 \leq k \leq K} \mathbf{E}[Z_m^{(k)}] (\mathbf{E}[\tau_m^2] x_{m,n} - \mathbf{E}[\tau_m] \mathbf{E}[\beta_X^{(k)}] - 2\mathbf{E}[\tau_m] x_{m,n} t_{m,n})}{\Lambda_{x_0}^*} \\ &+ \frac{\Lambda_x \sum_{1 \leq m \leq M} \sum_{1 \leq n \leq N_m} \sum_{1 \leq k \leq K} \mathbf{E}[Z_m^{(k)}] (\mathbf{E}[\beta_X^{(k)}] t_{m,n} + x_{m,n} t_{m,n}^2) + \Lambda_{x_0} \mu_{x_0}}{\Lambda_{x_0}^*} \end{aligned} \quad (5-60)$$

$$q^*(y_0) = \mathcal{N}(y_0 | \mu_{y_0}^*, \Lambda_{y_0}^{*, -1}) \quad (5-61)$$

$$\Lambda_{y_0}^* = \Lambda_y \sum_{1 \leq m \leq M} \sum_{1 \leq n \leq N_m} (\mathbf{E}[\tau_m^2] - 2t_{m,n} \mathbf{E}[\tau_m] + t_{m,n}^2) + \Lambda_{y_0} \quad (5-62)$$

$$\begin{aligned}
\mu_{y_0}^* &= \frac{\Lambda_y \sum_{1 \leq m \leq M} \sum_{1 \leq n \leq N_m} \sum_{1 \leq k \leq K} \mathbf{E}[Z_m^{(k)}] (\mathbf{E}[\tau_m^2] y_{m,n} - \mathbf{E}[\tau_m] \mathbf{E}[\beta_Y^{(k)}] - 2\mathbf{E}[\tau_m] y_{m,n} t_{m,n})}{\Lambda_{y_0}^*} \\
&+ \frac{\Lambda_y \sum_{1 \leq m \leq M} \sum_{1 \leq n \leq N_m} \sum_{1 \leq k \leq K} \mathbf{E}[Z_m^{(k)}] (\mathbf{E}[\beta_Y^{(k)}] t_{m,n} + y_{m,n} t_{m,n}^2) + \Lambda_{y_0} \mu_{y_0}}{\Lambda_{y_0}^*} \quad (5-63)
\end{aligned}$$

$$q^*(\tau_m) = \mathcal{N}(\tau_m | \mu_{\tau_m}^*, \Lambda_{\tau_m}^{*, -1}) \quad (5-64)$$

$$\begin{aligned}
\Lambda_{\tau_m}^* &= \Lambda_x \sum_{1 \leq n \leq N_m} (\mathbf{E}[x_0^2] - 2x_{m,n} \mathbf{E}[x_0] + x_{m,n}^2) + \Lambda_y \sum_{1 \leq n \leq N_m} (\mathbf{E}[y_0^2] - 2y_{m,n} \mathbf{E}[y_0] + y_{m,n}^2) \\
&+ \Lambda_w \sum_{1 \leq n \leq N_m} w_{m,n}^2 + \Lambda_h \sum_{1 \leq n \leq N_m} h_{m,n}^2 + \Lambda_{\tau_m} \quad (5-65)
\end{aligned}$$

$$\begin{aligned}
\mu_{\tau_m}^* &= \frac{\Lambda_x \sum_n \sum_k \mathbf{E}[Z_m^{(k)}] (\mathbf{E}[x_0^2] t_{m,n} - \mathbf{E}[x_0] \mathbf{E}[\beta_X^{(k)}] - 2\mathbf{E}[x_0] x_{m,n} t_{m,n} + \mathbf{E}[\beta_X^{(k)}] x_{m,n} + t_{m,n} x_{m,n}^2)}{\Lambda_{\tau_m}^{*, -1}} \\
&+ \frac{\Lambda_y \sum_n \sum_k \mathbf{E}[Z_m^{(k)}] (\mathbf{E}[y_0^2] t_{m,n} - \mathbf{E}[y_0] \mathbf{E}[\beta_Y^{(k)}] - 2\mathbf{E}[y_0] y_{m,n} t_{m,n} + \mathbf{E}[\beta_Y^{(k)}] y_{m,n} + t_{m,n} y_{m,n}^2)}{\Lambda_{\tau_m}^{*, -1}} \\
&+ \frac{\Lambda_w \sum_{1 \leq n \leq N_m} \sum_{1 \leq k \leq K} \mathbf{E}[Z_m^{(k)}] (\mathbf{E}[\beta_W^{(k)}] w_{m,n} + w_{m,n}^2 t_{m,n})}{\Lambda_{\tau_m}^{*, -1}} \\
&+ \frac{\Lambda_h \sum_{1 \leq n \leq N_m} \sum_{1 \leq k \leq K} \mathbf{E}[Z_m^{(k)}] (\mathbf{E}[\beta_H^{(k)}] h_{m,n} + h_{m,n}^2 t_{m,n})}{\Lambda_{\tau_m}^{*, -1}} \\
&+ \frac{\Lambda_{\tau_m} \mu_{\tau_m}}{\Lambda_{\tau_m}^*} \quad (5-66)
\end{aligned}$$

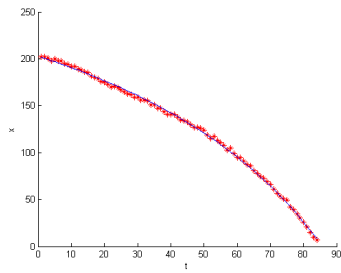
对于 $\mathbf{E}[Z_m^{(k)}]$, 计算公式如下:

$$\mathbf{E}[Z_m^{(k)}] = \frac{\mathbf{E}[\pi_k] \mathbf{P}(\mathbf{X}_n | \phi^{(k)})}{\prod_{1 \leq k \leq K} \mathbf{E}[\pi_k] \mathbf{P}(\mathbf{X}_n | \phi^{(k)})} \quad (5-67)$$

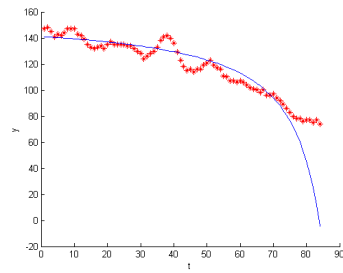
其中 $\phi^{(k)} \equiv \{\beta_X^{(k)}, \beta_Y^{(k)}, \beta_W^{(k)}, \beta_H^{(k)}, x_0, y_0, \tau_m\}$

而

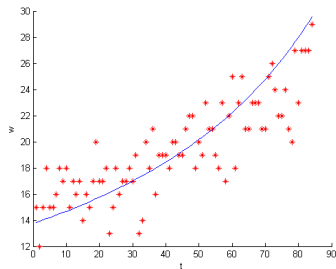
$$\mathbf{E}[\pi_k] = \sum_{1 \leq m \leq M} \mathbf{E}[Z_m^{(k)}] \quad (5-68)$$



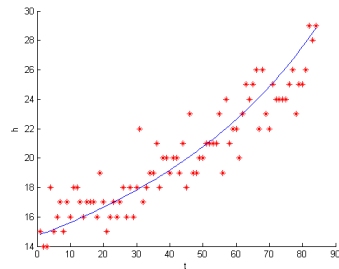
(a) x 随 t 的运动



(b) y 随 t 的运动



(c) w 随 t 的运动



(d) y 随 h 的运动

图 5.4 最大似然估计法结果1

5.2 实验结果

5.2.1 最大似然估计

如5.1.2节所述， x, y, w, h 随 t 变化的曲线都是双曲线，这些双曲线可以用最大似然估计出。一段来自数据集1的连续视频（大约200帧）的最大似然估计如图5.4。

但最大似然估计法的效果也有不好的情况，如图5.5。

5.2.2 在模拟数据上的测试

首先我们使用加入混合高斯扩展的运动概率模型在一个模拟的测试场景上运行，如图5.6。该场景仿真三条通道，每个通道上有30个交通标志，每个交通标志可以观测到100次。交通标志的观测值（即 (x, y, w, h, t) ）在模拟计算得出后加上高斯噪声，输入进概率模型。

在模拟场景下进行概率推测的结果如图5.7。图中，上面4行表示三个通道

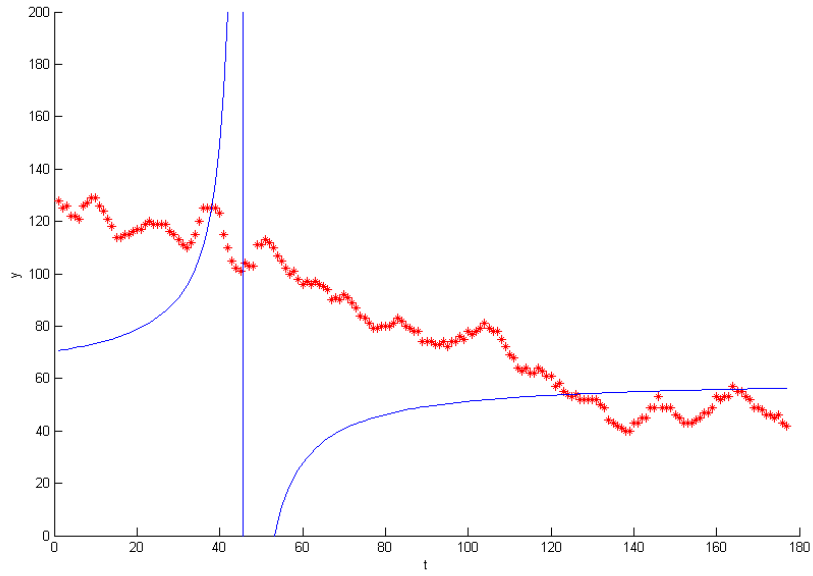


图 5.5 最大似然估计法结果2

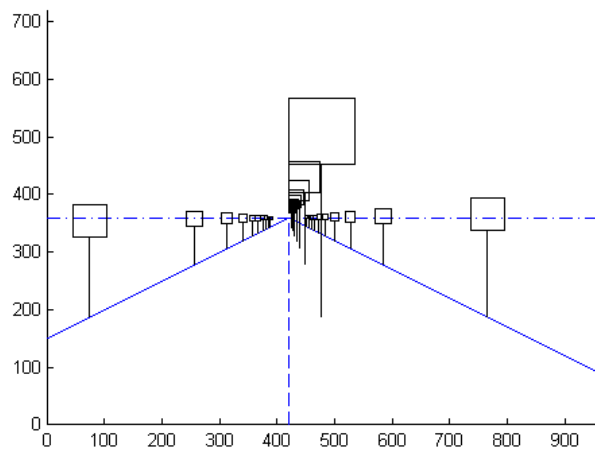


图 5.6 模型场景

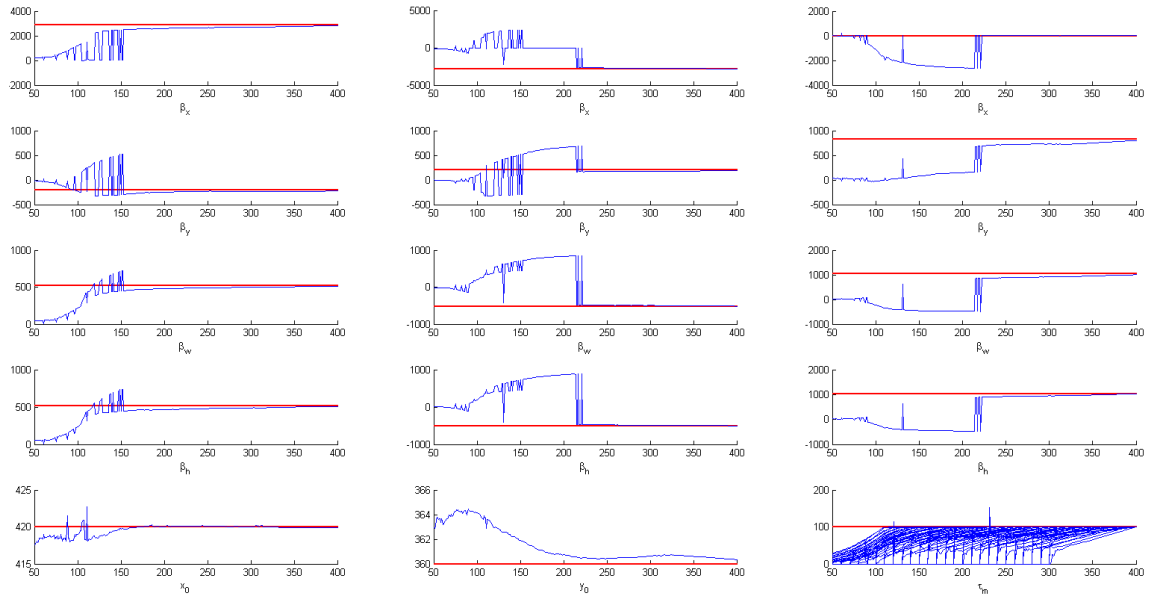


图 5.7 模型场景上的推理结果。上面4行表示三个通道各自的 $(\beta_X, \beta_Y, \beta_W, \beta_H)$ 的推理结果，最后一行表示 (x_0, y_0) 以及各个标识的 τ_m 的推理结果。红线表示真实值。每个通道有30个标识，分布在0到300帧之间依次出现，每个标识持续100帧。我们看到在大约230帧以后，即在大约每个通道出现了23个标识以后，各参数的推理结果基本达到真实值。

各自的 $(\beta_X, \beta_Y, \beta_W, \beta_H)$ 的推理结果，最后一行表示 (x_0, y_0) 以及各个标识的 τ_m 的推理结果。红线表示真实值。每个通道有30个标识，分布在0到300帧之间依次出现，每个标识持续100帧。我们看到在大约230帧以后，即在大约每个通道出现了23个标识以后，各参数的推理结果基本达到真实值。

5.2.3 在实际数据上的测试

使用加入混合高斯扩展的运动概率模型在实际的一段道路视频上的测试如图5.8。这段视频大约4分钟，由8个小段拼成。大约出现了5千次的交通标志观测，由大约由200个交通标志产生，即每个交通标志大约产生25次观测。这些观测的包围框全部由手工标注。

图中红色粗框为交通标志的包围框。包围框上方的数字为交通标志的编号，下方数字为这个交通标志所最有可能属于的通道编号及其确信度（一个标志属于各个通道的确信度的和为1）。包围框拖出的粉色矩形表示之前该交通标志的位置，给出了它的轨迹。蓝色表示先验给出的参数，绿色表示经推理后的参数。参数中，横竖点划线的焦点表示视线消失点，即 (x_0, y_0) ，从这个点辐

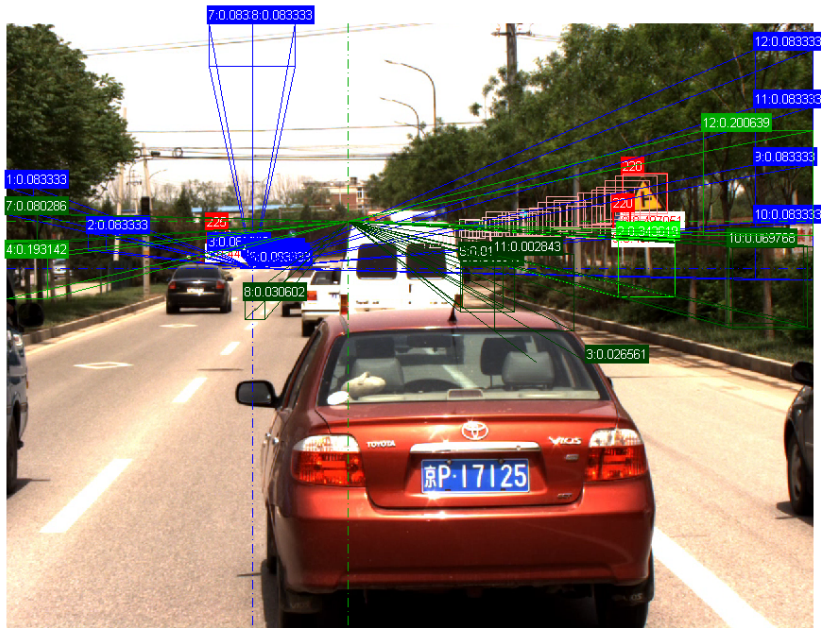


图 5.8 运动概率模型在实际数据上的测试。图中红色粗框为交通标志的包围框。包围框上方的数字为交通标志的编号，下方数字为这个交通标志所最有可能属于的通道编号及其确信度（一个标志属于各个通道的确信度的和为1）。包围框拖出的粉色矩形表示之前该交通标志的位置，给出了它的轨迹。蓝色表示先验给出的参数，绿色表示经推理后的参数。参数中，横竖点划线的焦点表示视线消失点，即 (x_0, y_0) ，从这个点辐射出的每个矩形表示一条通道，即 $(\beta_x, \beta_y, \beta_w, \beta_h)$ ；通道上方的数字表示属于该通道的所有交通标志占总交通标志的比例，即 π_k ；通道颜色也由 π_k 决定，值越大颜色越先验。

射出的每个矩形表示一条通道，即 $(\beta_x, \beta_y, \beta_w, \beta_h)$ ；通道上方的数字表示属于该通道的所有交通标志占总交通标志的比例，即 π_k ；通道颜色也由 π_k 决定，值越大颜色越先验。

从图中的结果看出，220号交通标志的轨迹基本符合12号通道，2、4、12三条通道是交通标志的主要通道。 x_0 的估计基本正确， y_0 的估计偏高，各个通道的分布基本符合直观。

第6章 总结与讨论

6.1 各方法优缺点

本文介绍了三种交通标志检测的方法，分别是颜色阈值法、Haar级层分类器法以及运动概率模型。

颜色阈值法的优点是能够避免滑动窗口的低效率，缺点是过分依赖颜色，无法适应光照环境变化，无法区分与交通标志颜色接近的物体。

Haar级层分类器法的优点是同时依赖纹理与形状，从而能够检测到没有大的色块、但形状很有特点的物体，如圆圈类交通标志，且速度比较快；其缺点是无法同时进行多个类别的检测，且在要求低误报率时，命中率会有较大损失。对于无法进行多类别检测的缺点，有两种解决办法：一是手工将正例分类，对每一类单独训练出一个层级分类器；二是扩展直线型的层级结构，将它变为树形，样例从树的根结点出发，只要沿任何一个路径到达叶结点，都算作正例。训练过程中，在训练单层分类器时，试着用KMeans算法对正例进行聚类，若聚类后对每一类单独训练出一个分类器后效果比较好，就使用这些分类器，并由此在这一点将树分叉，将正例分配到各个分支进行后续层的训练。

运动概率模型的特点是比较灵活，根据数据量的大小和已有信息的多少，可以将任何一个参数在随机变量与已知常量之间切换，或处于二者之间，即先验确信度比较大的随机变量；其缺点是需要的计算量较大，尤其是离线算法，而且匀速直线运动的假设比较强，如果车辆行驶证转弯很多，或者车辆变道、上下颠簸很严重，则会破坏假设。

6.2 主要贡献

本论文主要有三个方面的贡献，一是设计了一套颜色阈值法处理流程，组合运用了HSV颜色空间、Erode算法、Close算法、Flooding算法等；二是推导出了交通标志在图像投影上的运动公式，并设计了三种利用交通标志运动信息的模型，由弱到强依次是最大似然模型、全贝叶斯概率图模型和加入混合高斯扩展的全贝叶斯概率图模型；三是设计了三种推理算法，分别是在全贝叶斯概率图模型上的推理算法及其离线版本，以及混合高斯扩展模型上的推理算法。

建立这些模型，尤其是全贝叶斯模型，其意义是展示了怎样将运动几何分析与现今流行的概率方法，尤其是概率图方法结合起来。本论文分析的实际上是最简单的运动，其它的运动可以类似地设计出概率图模型。概率图的设计过程中，需要面对一组矛盾，这也是所有模型设计工作都需要面对的：灵活性与复杂性的矛盾。模型越复杂，越能涵盖更多变化，能力越强，但它需要的训练数据量也越多，在训练数据量不够的情况下，会产生过拟合。具体到概率图模型来说，作为随机变量的参数越多，模型越灵活；图的结构越复杂（边越多），模型越灵活。但大量的边（甚至是全联通），意味着敢于做出的条件独立假设越少；参数越多，意味着敢于做出的不变性假设越少。在缺少足够假设的情况下，实际上一个模型很难具有学习能力。但当做出的假设过多过强时，适用的情况就会变少，如果实际使用时真实环境打破了假设，效果就会下降。

设计推理算法的意义比较具体，它实际上是Variational Inference方法的一个应用，展示了Variational Inference，作为EM算法的一个贝叶斯扩展，在推理算法的设计上是强有力的。

在模型和推理算法的设计上，本文始终尽量做到拥有良好的概率解释，其结果就是将模型和算法推广的全贝叶斯的、全功能的规模。但实际上，根据速度和精度的要求，某些假设可以加强，某些模型或算法的步骤可以简化。这会导致算法失去概率依据，但能达到实验效果。比如，在加入混合高斯扩展后，我们在推理算法中并不一定要严格套用混合高斯推理的一般方法，而可以简化成类似KMeans算法，也就是把 $E[Z^{(k)}]$ 的取值从0到1之间的任何实数简化为只有0或1。另一个例子，在本文设计的混合高斯模型上的推理算法中，呈现出了E和M两个步骤，两个步骤中都要用到模型的似然函数，所以严格地说两个步骤中的似然函数应该相同。也就是说式5-67中的似然函数应该与式5-12相同。但实际实验中，如果原来的似然函数在式5-67中效果不好，例如区分度不够，可以尝试换一个似然函数，例如根据包围框与 (x_0, y_0) 的连线与各个通道（都是从 (x_0, y_0) 出发向外线性增大的矩形）的夹角。这种改变失去了概率模型依据，但具有直观意义，且不影响其它的步骤（例如5-67这一步只是用来决定交通标志属于哪个通道，改变它具体的实现方法，只要保证结果，就不影响模型的其它部分）。

插图索引

图 1.1	标准20类交通标志	1
图 2.1	数据集1的帧示例	4
图 2.2	数据集1中标注出的交通标识示例	5
图 2.3	数据集2的帧示例	5
图 2.4	数据集2中标注出的交通标识示例（第一类）	5
图 2.5	数据集2中标注出的交通标识示例（第二类）	6
图 2.6	数据集2中标注出的交通标识示例（第三类）	6
图 3.1	HSV的圆锥示意图	7
图 3.2	大面积、实色填充的矩形指路标志	8
图 3.3	原图像	11
图 3.4	S通道	11
图 3.5	H通道	11
图 3.6	V通道	11
图 3.7	H通道取阈值后的结果	12
图 3.8	取与运算后的Mask	12
图 3.9	Close算法处理后的图像	12
图 3.10	S通道取阈值后的结果	12
图 3.11	Erode算法处理后的图像	12
图 3.12	生成的包围框	12
图 3.13	Histogram Equalization后的原图像	14
图 3.14	Histogram Equalization后的S通道	14
图 3.15	Histogram Equalization后的H通道取阈值后的结果	14
图 3.16	Histogram Equalization后的H通道	14
图 3.17	Histogram Equalization后的V通道	14
图 3.18	Histogram Equalization后的S通道取阈值后的结果	14
图 3.19	Histogram Equalization后和取与运算后的Mask	15
图 3.20	Histogram Equalization后合Close算法处理后的图像	15
图 3.21	一幅出现误判的图片	15

图 3.22	Histogram Equalization后和Erode算法处理后的图像	15
图 3.23	Histogram Equalization后生成的包围框	15
图 3.24	误判图片最后的处理结果	15
图 4.1	层级分类器原理	17
图 4.2	Haar特征示意图	19
图 4.3	倾斜Haar特征示意图	19
图 4.4	基本Haar特征	19
图 4.5	扩展Haar特征	19
图 4.6	基本Haar特征性能 (数据集1)	25
图 4.7	基本Haar特征性能 (数据集2, 第一类)	25
图 4.8	基本Haar特征性能 (数据集2, 第二类)	25
图 4.9	扩展Haar特征性能 (数据集1)	25
图 4.10	扩展Haar特征性能 (数据集2, 第一类)	25
图 4.11	扩展Haar特征性能 (数据集2, 第二类)	25
图 4.12	基本Haar特征检测结果示例1 (数据集1)	26
图 4.13	基本Haar特征检测结果示例3 (数据集2, 第一类)	26
图 4.14	基本Haar特征检测结果示例2 (数据集1)	26
图 4.15	基本Haar特征检测结果示例4 (数据集2, 第二类)	26
图 5.1	交通标志运动公式证明	28
图 5.2	概率图模型	31
图 5.3	概率图模型	39
图 5.4	最大似然估计法结果1	43
图 5.5	最大似然估计法结果2	44
图 5.6	模型场景	44
图 5.7	模型场景上的推理结果	45
图 5.8	运动概率模型在实际数据上的测试	46

表格索引

表 4.1	基本Haar特征性能 (数据集1)	21
表 4.2	基本Haar特征性能 (数据集2, 第一类)	21
表 4.3	基本Haar特征性能 (数据集2, 第二类)	22
表 4.4	扩展Haar特征性能 (数据集1)	22
表 4.5	扩展Haar特征性能 (数据集2, 第一类)	23
表 4.6	扩展Haar特征性能 (数据集2, 第二类)	23

公式索引

公式 3-1	8
公式 3-2	8
公式 3-3	8
公式 3-4	8
公式 3-5	13
公式 4-1	17
公式 4-2	18
公式 4-3a.....	18
公式 4-3b.....	18
公式 4-4	19
公式 5-1	27
公式 5-2	28
公式 5-3	28
公式 5-4	29
公式 5-5	29
公式 5-6	29
公式 5-7	29
公式 5-8	29
公式 5-9	30
公式 5-10.....	30
公式 5-11.....	30
公式 5-12.....	30
公式 5-13.....	32
公式 5-14.....	32
公式 5-15.....	32
公式 5-16.....	33
公式 5-17.....	33
公式 5-18.....	33

公式 5-19.....	33
公式 5-20.....	33
公式 5-21.....	33
公式 5-22.....	33
公式 5-23.....	34
公式 5-24.....	34
公式 5-25.....	34
公式 5-26.....	34
公式 5-27.....	34
公式 5-28.....	34
公式 5-29.....	34
公式 5-30.....	34
公式 5-31.....	35
公式 5-32.....	35
公式 5-33.....	35
公式 5-34.....	35
公式 5-35.....	35
公式 5-36.....	35
公式 5-37.....	35
公式 5-38.....	36
公式 5-39.....	36
公式 5-40.....	37
公式 5-41.....	37
公式 5-42.....	38
公式 5-43.....	39
公式 5-44.....	39
公式 5-45.....	40
公式 5-46.....	40
公式 5-47.....	40
公式 5-48.....	40
公式 5-49.....	40

公式 5-50.....	40
公式 5-51.....	40
公式 5-52.....	40
公式 5-53.....	41
公式 5-54.....	41
公式 5-55.....	41
公式 5-56.....	41
公式 5-57.....	41
公式 5-58.....	41
公式 5-59.....	41
公式 5-60.....	41
公式 5-61.....	41
公式 5-62.....	41
公式 5-63.....	42
公式 5-64.....	42
公式 5-65.....	42
公式 5-66.....	42
公式 5-67.....	42
公式 5-68.....	42

参考文献

- [1] Broggi A, Cerri P, Medici P, et al. Real Time Road Signs Recognition. Proceedings of IEEE Intelligent Vehicles Symposium (2007), Istanbul, Turkey, 2007. 981–986
- [2] Escalera A, Luis E Moreno M, Salichs M A, et al. Road Traffic Sign Detection and Classification. Proceedings of IEEE Transactions on Industrial Electronics, volume 44, 1997. 848–859
- [3] Escalera A, Armingol J M, Mata M. Traffic sign recognition and analysis for intelligent vehicles. Image and Vision Computing, 2003, 21(3):247 – 258
- [4] Xiaohong G, Kunbin H, Passmore P, et al. Colour vision model-based approach for segmentation of traffic signs. J. Image Video Process., 2008, 2008:1–7
- [5] X G, N S, K H, et al. Vision models based identification of traffic signs. Proceedings of Proceedings of the First European Conference on Color in Graphics, Image and Vision, Poitiers, France, 2002. 47–51
- [6] Fleyeh H. Road and Traffic Sign Color Detection and Segmentation - A Fuzzy Approach. Proceedings of MVA, 2005. 124–127
- [7] Viola P, Jones M J. Robust Real-Time Face Detection. International Journal of Computer Vision, 2004, 57(2):137–154
- [8] Papageorgiou C, Poggio T. A trainable system for object detection. International Journal of Computer Vision, 2000, 38(1):15–33
- [9] Bahlmann C, Zhu Y, Ramesh V, et al. A System for Traffic Sign Detection, Tracking, and Recognition Using Color, Shape, and Motion Information. Proceedings of IEEE Intelligent Vehicles Symposium (IV 2005), 2005
- [10] Lienhart R, Maydt J. An Extended Set of Haar-like Features for Rapid Object Detection. Proceedings of IEEE ICIP 2002, volume 1, 2002. 900–903
- [11] Soetedjo A, Yamada K. Fast and Robust Traffic Sign Detection. Proceedings of 2005 IEEE International Conference on Systems, Man and Cybernetics, Waikoloa, Hawaii, 2005. 1341–1346
- [12] Barczak A L C. Toward an Efficient Implementation of a Rotation Invariant Detector using Haar-Like Features. Proceedings of IVCNZ05, IDunedin, NZ, 2005. 31–36
- [13] Paclík P, Novovicova J, Pudil P, et al. Road sign classification using Laplace kernel classifier. Pattern Recognition Letters, 2000, 21(13-14):1165–1173
- [14] Hoiem D, Efros A, Hebert M. Putting Objects in Perspective. International Journal of Computer Vision, 2008, 80(1):3–15

- [15] Dalai N, Triggs B, Rhone-Alps I, et al. Histograms of oriented gradients for human detection. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005, volume 1, 2005
- [16] Heitz G, Koller D. Learning Spatial Context: Using Stuff to Find Things. Proceedings of Proc. 10th European Conference on Computer Vision, 2008
- [17] Piccioli G, Micheli E D, Parodi P, et al. Robust method for road sign detection and recognition. Proceedings of Image and Vision Computing (1996), volume 14. 209–223
- [18] Ying Z, Comaniciu D, Ramesh V, et al. An Integrated Framework of Vision-based Vehicle Detection with Knowledge Fusion. Proceedings of IEEE Intelligent Vehicles Symposium, 2005, 2005. 199–204
- [19] WOHLFART B, EDMAN K A P. Rectangular Hyperbola Fitted To Muscle Force-Velocity Data Using Three Dimensional Regression Analysis. Proceedings of Experimental Physiology (1994), volume 79. 235–239
- [20] Bishop C, et al. Pattern recognition and machine learning. Springer New York:, 2006: 461–466

致 谢

衷心感谢导师张钹院士和李建民老师对我的精心指导。感谢清华大学计算机系智能与系统国家实验室智能车项目组提供的设备、数据和实验平台。

声 明

本人郑重声明：所提交的学位论文，是本人在导师指导下，独立进行研究工作所取得的成果。尽我所知，除文中已经注明引用的内容外，本学位论文的研究成果不包含任何他人享有著作权的内容。对本论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明。

签 名：_____ 日 期：_____

附录 A 外文资料的调研阅读报告（英文）

Traffic/Road sign detection and recognition has long as a subtask been explored and studied in the areas of automation and robotics. It is also studied as a specific object-detecting task in computer vision and image processing. The existing means to this task can be categorized into two major classes: color-based algorithms^[1-6] and shape-based algorithms^[7-11].

[1-3] all use color-thresholding to distinguish pixels belonging to the target object(s), and several following processing steps to group pixels into bounding boxes, though the versions of color-thresholding and the processing steps they exploit are different. Among shape-based methods, those developed in human-face detection tasks are most popular and successful, many of which use Haar-wavelet features^[8]. A Haar-wavelet consists of two rectangular pulses, usually one positive and the other negative. A feature value is the convolution of the wavelet and the stimulus. Specific to image, a Haar-wavelet is two adjacent rectangles sharing an edge, one of which is denoted +1 and the other -1. The feature value is calculated from the sum of pixels covered by the +1 rectangle minus the sum of pixels covered by the -1 rectangle. Haar-wavelet has a fast and efficient computing algorithm, called Integral Image Method, which comes from the simple observation that the sum of a rectangular region can be calculated from its four corners, as long as we know the sums of the rectangles with each of these four corner as its bottom-right and the whole image's top-left corner as its top-left. Besides basic Haar-wavelet features, [10] introduces diagonally tilted Haar features and its fast calculating algorithm. This extension of the basic Haar-wavelets can more precisely capture edges and lines at 45° than two diagonal rectangles. [9] introduces color-Haar wavelets, by splitting image into 7 color channels and treating wavelets on each channel all as independent features. The choice of channels or color space can be in fact arbitrary, and thus this extension can capture characteristics only obvious in a specific color channel. [12] discussed the idea that a Haar liked feature tilted at any degree can be expressed equivalently by a linear combination of two basic Haar-wavelets, which is named by the author as PEF (Pair of Equivalent Features).

Though common in the usage of Haar-wavelet features, researchers can employ various kinds of classifiers to make a detector. [7] introduces the classic Cascade Classifier. A cascade classifier filters the samples stage by stage, where each stage is itself a classifier. The classifying capability of the classifiers at later stages are stronger than the former ones. Any time a sample is classified as negative at a stage, it is thrown and doesn't get to the next stage. Only the samples passing through all stages are treated as positive. The goal of each stage classifier is to rule out as many negative samples as possible, as long as it can maintain a considerably high hit rate, that is to say letting almost all positive samples pass. Since most of negative samples get excluded in former stages, very few negative samples can make it to later and more complicated stage classifiers; and because former stage classifiers are fast and cheap, the expected computation complexity for every sample can be much lower than directly using the most complicated and strong classifier. It can also be seen that this cascade structure is most suitable in settings with much more negative samples than positives, which is exactly the case in our road sign detection task, thanks to the sliding window method utilized by the detector.

For the cascade classifier developed in [7], each stage classifier is a linear combination of several Haar-wavelet features, which are selected from all possible Haar-wavelets (which can be millions) by Adaboost algorithm in training. The training samples for every stage consist of all the positive samples and the negative samples having passed former stages. Adaboost is used to choose and combine more and more features until the combination can successfully classify those training samples, by "successfully" we mean achieving a given false alarm rate and hit rate.

Other than Haar-wavelet features and cascade classifier, [11] proposes an ellipse detector consists of six local component detectors. [13] uses Laplace kernel as classifier. [9] uses LDA (Latent Dirichlet Allocation) probability model as classifier.

Having chosen and constructed a local road sign detector for static image frame, one can transfer attention to the context and spatial information from the whole image, such as horizon, sky, ground, and the dynamic information buried in video sequences, such as the motion pattern of the target object. Such information can be exploited to predict the probability that a target object might appear at a given position in next

frame, and hence improving the accuracy of detector or enhancing speed by focusing the sliding in high probability areas.

[14] discussed means of utilizing geometric and perspective information, introducing a probability graph model including three kinds of information: the outputs of local detector, the roughly segmented regions (an image is segmented into three kinds of regions: sky, ground and vertical planes), and perspective parameters comprising the location of horizon in the image and the real-world height of the camera. All the sources of these kinds of information are partial and inaccurate, but they can be iterated and enhanced mutually within the probability model. The local detector it uses is a SVM with HOG^[15] features, typical in pedestrian and car detection. [16] proposes another probability graph model, using more finely segmented regions (over-segmented or "superpixel" image) with labels for each region and capturing geometric relationship (above, below, left, right, within, etc.) between pairs of regions. Addition to probability graph model, [17] designed a Kalman filter based on the motion dynamics of the road signs. [18] uses a method call "Knowledge-Fusion" to construct motion model.