基于多条件随机场模型的图像3D空间布局理解

刘 威[1,2],周 婷[2],袁 淮[1],赵 宏[1]
(1.东北大学研究院,辽宁沈阳 110819;2.东软集团股份有限公司,辽宁沈阳 110179)

摘 要: 图像3D空间布局理解在自动驾驶系统以及目标识别中扮演着重要的角色.本文提出一种基于多条件随机场模型集成的图像3D空间布局理解算法.首先,基于多次图像分割生成多个不同尺度的超像素图像;然后,结合LBP表面纹理特征、LM滤波器组获得的方向纹理特征、颜色特征以及图像中超像素的位置和形状特征,建立各尺度的超像素图像中超像素的特征表达;最后,为各尺度的超像素图像分别构建相应的条件随机场模型,并应用D-S证据合成理论对多个条件随机场模型的推断结果进行集成,实现对图像3D空间布局的理解.在公共数据集GC和KITTI Layout上的实验结果表明,同已有算法相比,本文提出的算法提高了图像3D空间布局理解的准确率.

关键词: 3D空间布局;多次图像分割;超像素特征表达;条件随机场模型;证据合成
中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112 (2017)02-0328-09
电子学报URL: http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2017.02.010

3D Spatial Layout Understanding from Image Based on Multiple CRFs

LIU Wei[1,2], ZHOU Ting[2], YUAN Huai[1], ZHAO Hong[1]
(1. *Research Academy, Northeastern University, Shenyang, Liaoning* 110819, *China*;
2. *Neusoft Corporation, Shenyang, Liaoning* 110179, *China*)

Abstract: 3D spatial layout understanding from images plays an important role in the autonomous driving and object recognition. This paper proposes a 3D spatial layout understanding algorithm based on multiple CRFs. Firstly, multiple different scales super-pixel image are generated based on the multiple segmentation. Then, the feature of super-pixel are constructed based on LBP surface texture feature, orientation texture feature from LM filters, color feature, and location and shape feature of super-pixels in the image. Finally, the CRF model on every scale super-pixel image is built, the Dempster-Shafer theory of evidence is used to integrate the inference result of multiple CRF models and the 3D spatial layout understanding from an image is realized. The experiments on the public database Geometric Context and KITTI Layout demonstrate that the algorithm proposed in this paper improves the average accuracy of 3D spatial layout understanding comparing to the existing state-of-art.

Key words: 3D spatial layout; multiple image segmentation; feature of super-pixel; conditional random field models; evidence combine

1 引言

图像3D空间布局理解是根据多种图像特征信息,将图像中每个像素点标记为不同几何类别的过程,天空(sky)、立体物(vertical)以及地面(support)是三类常见的几何类别.由于图像3D空间布局信息能够模拟人的感知机理,反映从相机角度呈现在图像中的表面方向信息以及真实场景的空间上下文关系,因此,图像3D空间布局理解得到了研究者的广泛关注,常被应用在自动驾驶系统中道路、自由驾驶空间检测[1,2]、目标识别[3]等领域.

从模型上看,现有图像3D空间布局理解算法可分为两大类:(1)以超像素为处理单元建立决策树模型对超像素分类;(2)以单次分割获得的超像素为处理单元建立单一关系模型(如条件随机场模型、贝叶斯模型)推断超像素所属标签类别.第一类算法一般只考虑单个超像素的局部信息以及位置等几何特征信息,忽略了图像中的上下文关系,如文献[4~6]中使用决策树推断超像素的几何类别标签,只考虑了超像素本身的

收稿日期:2015-06-12;修回日期:2015-12-09;责任编辑:孙瑶
基金项目:国家自然科学基金(No.61273239);中央高校基本业务费(No.N151802001)

特征对其类别标签的影响,而忽略了相邻超像素之间的上下文关系. 文献[7,8]在上述算法的基础上利用大量的未标记数据来提高图像标记的平均准确率,但仍没有利用图像上下文关系信息. 同第一类算法相比,第二类算法考虑了图像上下文关系信息,但它们的关系模型仅建立在单次图像分割的基础上,图像标记结果难以避免图像分割错误的影响. 如文献[9]在单次超像素分割的基础上建立条件随机场模型对图像中的 3D 空间布局进行理解,文献[10]基于单次分割结果直接从测试图像中获得经典贝叶斯模型,求取最大后验概率来标注图像. 上述方法都没有考虑超像素分割的不准确性对模型推断的影响.

针对上述图像 3D 空间布局理解算法存在的问题,以及单次图像分割方法很难保证对任意图像进行分割所得超像素都只包含单一对象或者与对象边缘精确吻合的情况,本文采用多次分割框架,生成多个不同尺度的超像素图像,结合 LBP 表面纹理特征、LM 滤波器组获得的方向纹理特征、颜色特征以及图像中的位置和形状特征,建立各尺度超像素图像中各超像素的特征表达,并为各尺度的超像素图像构建相应的条件随机场模型. 最后应用 D-S 证据合成理论对多个条件随机场模型的推断结果进行集成,实现对图像的 3D 空间布局的推断和解释. 本文结构首先对研究背景进行简单介绍;其次介绍本文提出的 3D 空间布局理解算法;最后,通过实验证明本文所提算法的有效性,并给出结论.

2　研究背景

2.1　条件随机场

条件随机场(Conditional Random Field,CRF)是由 Lafferty 等提出的一种概率图模型[11]. CRF 模型可用于描述节点之间的关系,近年来,被广泛应用到场景理解[9]中,并取得了较好的效果. 在场景理解中,一般将像素或超像素定义为图中的节点,邻域内相邻像素或超像素之间的邻接关系作为图中节点间的无向连接,节点的特征对应于观测随机变量序列 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n$,节点的类别标号对应于待观测向量 $\boldsymbol{y}_1,\boldsymbol{y}_2,\cdots,\boldsymbol{y}_n$. 为了对图中各节点类别进行自动标记从而实现场景理解,建立的条件随机场模型通常如下所示:

$$E=\sum_i \boldsymbol{\psi}_i(\boldsymbol{\alpha},y_i,\boldsymbol{x}_i)+\sum_i\sum_j \boldsymbol{\psi}_{ij}(\boldsymbol{\beta},y_i,y_j,\boldsymbol{x}_{i,j}) \tag{1}$$

其中,$\boldsymbol{\Psi}_i$ 和 $\boldsymbol{\Psi}_{ij}$ 分别为一元和二元势函数,$\boldsymbol{\alpha}$ 和 $\boldsymbol{\beta}$ 分别为一元势函数、二元势函数中的权重系数矩阵,$\boldsymbol{x}_i$ 表示第 i 个节点的观测向量,$\boldsymbol{x}_{i,j}$ 表示相邻两节点间的二元差异性表达向量,y_i 表示第 i 个节点的类别标号.

2.2　集成学习

集成学习是使用一系列分类器进行学习,并使用某种规则把各个分类器的分类结果进行整合从而获得比单个分类器更好分类效果的一种机器学习方法[12]. 近年来,机器学习领域的研究者提出了多种集成学习方法,如基于 boosting、bagging 的自适应集成学习[13,14]、规则集成学习[15]等,在目标检测、识别领域取得了很好的效果. 如文献[16]利用贝叶斯规则集成多个 CRF 模型实现遥感图像中城市区域检测,同单个 CRF 模型相比,CRF 集成模型显著地提高了检测准确率. 上述集成学习方法的一个共同点在于,参与集成的每个分类器都是使用相同的样本集合,以更新样本权重的方式进行训练,对于利用相互独立的样本集合训练获得的分类器则不能有效集成. 为此,本文采用 D-S 证据理论来集成多个相互独立的不同尺度超像素图像 CRF 推断模型,见 3.2.2 节.

2.3　颜色矩特征

颜色矩是 stricker 和 orengo 提出的一种简单有效的颜色特征[17]. 这种特征的数学基础是图像中的任何颜色分布均可以用它的矩表示. 由于颜色分布信息主要集中在低阶矩中,因此,颜色的一阶矩(Mean)、二阶矩(Variance)和三阶矩(Skewness)常被用于表达图像的颜色分布特征:

$$\begin{aligned} m_i &= \frac{1}{M}\sum_{j=1}^{M} p_{ij} \\ \sigma_i &= \sqrt{\left(\frac{1}{M}\sum_{j=1}^{M}(p_{ij}-m_i)^2\right)} \\ s_i &= \sqrt[3]{\left(\frac{1}{M}\sum_{j=1}^{M}(p_{ij}-m_i)^3\right)} \end{aligned} \tag{2}$$

其中,m_i 表示均值,反映图像的整体明暗程度;σ_i 为方差,反映图像颜色分布的范围;s_i 为斜度,描述图像颜色分布的对称性;p_{ij} 表示第 j 个像素点在第 i 个颜色通道的像素值,M 为计算颜色矩的图像区域中包含的像素点个数.

2.4　LM 滤波器组

LM 滤波器组是由 Thomas 和 Jitendra 提出的一个 48 维的滤波器组[18],它由 36 个方向滤波器、8 个高斯拉普拉斯滤波器以及 4 个高斯滤波器组成,如图 1 所示. 基于 LM 滤波器组的响应所构建的特征向量,能够很好的描述图像纹理特征,因此在场景理解中被广泛应用[5].

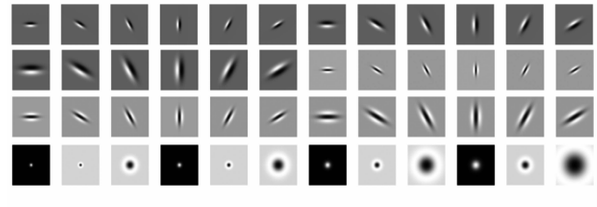


图1　LM滤波器组

3 基于多条件随机场模型的图像 3D 空间布局理解算法研究

3.1 基于单次图像分割的图像 3D 空间布局理解模型

3.1.1 超像素特征表达

图像 3D 空间布局理解的一大难题在于找到具有强区分能力的特征表达,基于像素点的图像特征表达存在以下问题:像素点只是图像的一系列离散表达,本身携带的信息量是有限的,其自身的颜色、亮度等特征不足以决定该像素点属于何种语义类型. 而通过无监督分割获得的超像素能够提供图像的先验分割信息,以分割后的超像素作为基本处理单元,比单个像素点具有更丰富的特征用以决定其所属类别. 由于图像 3D 布局中天空、立体物及地面在颜色、纹理以及位置形状等几何特征方面具有可区分性,因此,本文使用文献[5]提供的初始图像分割结果生成超像素图像,建立超像素特征表达 $f_s=\{f_{\text{color}},f_{a\_\text{texture}},f_{o\_\text{texture}},f_{l\_s}\}$,其中 f_{color}、f_{a_texture}、f_{o_texture}、f_{l_s} 分别为基于超像素的颜色特征、表面纹理特征、方向纹理特征以及位置和形状特征.

颜色特征对图像本身的尺寸、方向、视角的依赖性较小,被广泛用于图像检索、目标跟踪、目标识别等领域. 由于天空、立体物以及地面在颜色分布上通常具有可区分性,为此,本文利用颜色直方图和颜色矩[17]两种特征,建立超像素 S_p 的颜色特征表达:

$$f_{\text{color}}=\{\boldsymbol{m}_{r,g,b},\boldsymbol{m}_{h,s,v},\boldsymbol{\sigma}_{r,g,b},\boldsymbol{\sigma}_{h,s,v},\boldsymbol{s}_{r,g,b},\boldsymbol{s}_{h,s,v},\boldsymbol{H}_h,\boldsymbol{H}_s\} \tag{3}$$

其中,$\boldsymbol{H}_h$、$\boldsymbol{H}_s$ 分别表示在 hsv 颜色空间中 h 和 s 颜色通道直方图特征. 本文将 h 和 s 通道分别等分为 5 个和 3 个 bin. $\boldsymbol{m}_{r,g,b}=\{m_r,m_g,m_b\}$、$\boldsymbol{\sigma}_{r,g,b}=\{\sigma_r,\sigma_g,\sigma_b\}$、$\boldsymbol{s}_{r,g,b}=\{s_r,s_g,s_b\}$、$\boldsymbol{m}_{h,s,v}=\{m_h,m_s,m_v\}$、$\boldsymbol{\sigma}_{h,s,v}=\{\sigma_h,\sigma_s,\sigma_v\}$ 和 $\boldsymbol{s}_{h,s,v}=\{s_h,s_s,s_v\}$ 分别表示超像素 S_p 在 rgb 和 hsv 颜色空间中各颜色通道的一阶矩、二阶矩及三阶矩特征.

天空、立体物和地面通常具有不同的表面纹理,而 LBP 特征能够很好的描述纹理特性,因此,本文结合 LBP 特征[19]的思想,统计超像素 S_p 的表面纹理特征. 首先,使用 3×3 窗口对图像中的每个像素计算 LBP 编码. 在本文中采用 256 种模式的编码方式. 然后,在超像素 S_p 内统计各个 LBP 码对应的像素点个数直方图,得超像素 S_p 的表面纹理特征为:

$$f_{a\_\text{texture}}=\boldsymbol{H}_{\text{LBP}}=\{h_0,\cdots,h_i,\cdots,h_{255}\} \tag{4}$$

其中,h_i 表示超像素 S_p 中 LBP 编码值为 i 的像素个数归一化后的值:

$$h_i=\frac{1}{M}\sum_{j=1}^{M}1\{i=\text{LBP}_j\} \tag{5}$$

M 为超像素 S_p 中包含的像素点个数,LBP_j 为超像素 S_p 中第 j 个像素点的 LBP 编码值,$1\{\cdot\}$ 为符号函数,当括号中的条件成立时函数取值 1,否则取值 0.

考虑到立体物在图像中通常具有更多垂直方向或沿道路方向的纹理特性. 为此,本文采用前述 LM 滤波器组[18]对图像中每个像素进行滤波,然后,统计每个超像素 S_p 的方向纹理特征,与文献[5]只使用 LM 滤波器组中 15 个滤波器不同的是,本文使用了整个 LM 滤波器组. 在本文中,LM 滤波器组本文建立的超像素 S_p 的方向纹理特征表达为:

$$f_{o\_\text{texture}}=\{\boldsymbol{m}_{\text{LM}},\boldsymbol{H}_{\text{LM}}\} \tag{6}$$

其中,$\boldsymbol{m}_{\text{LM}}$ 为超像素 S_p 中像素点在 48 个滤波响应上的均值 $\boldsymbol{m}_{\text{LM}}=\{m_1,\cdots,m_i,\cdots,m_{48}\}$,$\boldsymbol{H}_{\text{LM}}$ 为超像素 S_p 中所有像素点在 48 个滤波响应上最大响应值统计直方图 $\boldsymbol{H}_{\text{LM}}=\{q_1,\cdots,q_i,\cdots,q_{48}\}$. 本文中,$m_i$ 以及 q_i 的计算公式如下所示:

$$m_i=\frac{1}{M}\sum_{j=1}^{M}I_j^i \tag{7}$$

$$q_i=\frac{1}{M}\sum_{j=1}^{M}1\{i=\arg\max_k(I_j^k)\} \tag{8}$$

M 为超像素 S_p 包含的像素点个数,I_j^i 为超像素 S_p 中的第 j 个像素点在第 i 个滤波器上的滤波响应值. I_j^i 为超像素 S_p 中的第 j 个像素点在第 k 个滤波器上的滤波响应值,$1\{\cdot\}$ 的取值与式(5)相同. 通过建立以上响应均值和最大响应值直方图的特征表达方式,使得本文很容易获得超像素 S_p 的方向纹理特性.

此外,分析图像的 3D 空间布局可知天空、立体物及地面一般分别具有位于消失线以上、穿过消失线、在消失线以下等特性,为此,本文引入消失线位置这一关键特征. 则通过动态消失线的估计[20],计算超像素与消失线之间的位置关系,结合超像素在图像中的位置以及形状,建立超像素的位置和形状特征如下:

$$\begin{aligned}&f_{l\_s}=\{f_{ar},f_{lc},f_{vl}\};f_{ar}=l_x/l_y\\&f_{lc}=\{s_{x1},s_{x2},s_{y1},s_{y2}\};\\&f_{vl}=\{s_{y1}-h_{vl},s_{y2}-h_{vl},1\{(s_{y1}-h_{vl})>0\}\\&\quad+1\{(s_{y2}-h_{vl})>0\}+1\}\end{aligned} \tag{9}$$

其中,f_{ar} 为超像素区域外接矩形纵横比,s_{x1} 和 s_{x2} 为超像素区域中像素点列坐标从小到大排序后排在前 10% 和前 90% 的坐标值,s_{y1} 和 s_{y2} 分别为行坐标从小到大排序后排在前 10% 和前 90% 的坐标值,h_{vl} 为图像消失线行坐标位置,$1\{\cdot\}$ 的取值与式(5)相同.

3.1.2 基于条件随机场模型的图像 3D 空间布局理解算法

为实现图像 3D 空间布局的理解,本文结合上述建立的超像素特征表达,采用条件随机场模型[11]对超像素图像进行关系建模,并将超像素定义为图中的节点,邻域内相邻超像素之间的邻接关系作为图中节点间的无向连接,超像素的特征对应于观测随机变量序列 $\boldsymbol{X}$,

超像素的类别标号对应于待观测向量 $\boldsymbol{Y}$,则图像中所有超像素的特征向量 $\boldsymbol{X}$ 与其类别标号 $\boldsymbol{Y}$ 可以视为一个条件随机场($\boldsymbol{X},\boldsymbol{Y}$). 本文针对 3.1.1 节中获取的超像素图像,建立式(1)所描述的条件随机场模型,其中,本文一元和二元势函数的表达如下所示:

$$\begin{aligned}\varphi_i(\boldsymbol{\alpha},y_i,\boldsymbol{x}_i)&=\boldsymbol{\alpha}\boldsymbol{x}_i\\ \varphi_{i,j}(\boldsymbol{\beta},y_i,y_j,\boldsymbol{x}_{i,j})&=\boldsymbol{\beta}\boldsymbol{x}_{i,j}\end{aligned}\tag{10}$$

其中,y_i 为第 i 个超像素的类别标号,$\boldsymbol{x}_i$ 为第 i 个超像素的一元特征表达向量,$\boldsymbol{x}_{i,j}$ 为相邻两个超像素的二元差异性特征表达向量. 采用 3.1.1 中提出的 f_s 描述上述模型中超像素的一元特征表达向量,则 $\boldsymbol{x}_i=\{f_s,1\}^{\mathrm{T}}$. 二元特征表达向量 $\boldsymbol{x}_{i,j}=\{\boldsymbol{x}_{i,j},1\}^{\mathrm{T}}$,其中 $\boldsymbol{x}_{i,j}$ 为相邻两超像素相似程度的置信度值. 在判别两个特征向量相似程度时,一些文献将相似性问题转换为二者特征向量之间差值向量的二类分类问题,在实际应用中显示出比直接计算两个特征向量最近距离方法更好的判别性能,如文献[21]将两张人脸相似性判别转换为分类问题. 为此,在计算两个相邻超像素 S_i 和 S_j 的相似度时,本文借鉴同样的思想,将相邻两个超像素的相似性问题转换为二者特征向量之间差值向量的二类分类问题,并使用 SVM 分类器进行分类. 具体地,在 SVM 分类器的训练样本中,若 S_i 和 S_j 具有相同标签,则将它们特征向量的差值向量 $f_{i,j}$ 作为一个正样本的特征向量,反之,若具有不同标签,则将 $f_{i,j}$ 作为一个负样本的特征向量:

$$\begin{aligned}f_{i,j}=\{&\boldsymbol{m}^i_{r,g,b}-\boldsymbol{m}^j_{r,g,b},\boldsymbol{m}^i_{h,s,v}-\boldsymbol{m}^j_{h,s,v},|\boldsymbol{H}^i_h-\boldsymbol{H}^j_h|,\\ &|\boldsymbol{H}^i_s-\boldsymbol{H}^j_s|,|\boldsymbol{H}^i_{\mathrm{LM}}-\boldsymbol{H}^j_{\mathrm{LM}}|,|\boldsymbol{H}^i_{\mathrm{LBP}}-\boldsymbol{H}^j_{\mathrm{LBP}}|\}\end{aligned}\tag{11}$$

其中,$\boldsymbol{m}^i_{r,g,b}-\boldsymbol{m}^j_{r,g,b}$、$\boldsymbol{m}^i_{h,s,v}-\boldsymbol{m}^j_{h,s,v}$ 分别为 S_i 与 S_j 的 rgb、hsv 颜色通道均值之差,$|\boldsymbol{H}^i_h-\boldsymbol{H}^j_h|$、$|\boldsymbol{H}^i_s-\boldsymbol{H}^j_s|$、$|\boldsymbol{H}^i_{\mathrm{LM}}-\boldsymbol{H}^j_{\mathrm{LM}}|$ 以及 $|\boldsymbol{H}^i_{\mathrm{LBP}}-\boldsymbol{H}^j_{\mathrm{LBP}}|$ 分别为 S_i 与 S_j 的 h、s 通道直方图距离、方向纹理特征中的直方图距离及表面纹理特征直方图距离. 本文使用 chi-square 距离计算两直方图之间的距离,计算公式如下:

$$\chi^2(\boldsymbol{H}_1,\boldsymbol{H}_2)=\sum_d^D\frac{(\boldsymbol{H}^d_1-\boldsymbol{H}^d_2)^2}{\boldsymbol{H}^d_1+\boldsymbol{H}^d_2}\tag{12}$$

其中,$\boldsymbol{H}_1$、$\boldsymbol{H}_2$ 分别为两个 D 维的直方图. 记 SVM 分类器的输出结果为 f_{svm},为了将 SVM 的输出结果作为相似程度置信度值,本文利用全局归一化方法(Global Normalization)[22]对测试样本的分类结果进行归一化,然后使用 sigmoid 函数将归一化的结果转换为 S_i 和 S_j 相似程度的置信度值. 具体方法如下:

$$x_{i,j}=s(g(f_{\mathrm{svm}}))=\begin{cases}s\left(\dfrac{f_{\mathrm{svm}}-\mu^+}{\sigma^+}\right), & \text{if}\quad f_{\mathrm{svm}}>0\\ -s\left(-\dfrac{f_{\mathrm{svm}}-\mu^-}{\sigma^-}\right), & \text{if}\quad f_{\mathrm{svm}}\leqslant 0\end{cases}\tag{13}$$

其中 μ^+、μ^- 分别表示输出为正的和输出为负的所有值的平均值,σ^+、σ^- 为对应的标准差,$s\{\cdot\}$ 为 sigmoid 函数.

在对上述建立的条件随机场模型进行训练时,本文将 soft-max 多类分类器[23]的参数训练结果,作为极大似然估计(Maximum Likelihood Estimation,MLE)中一元势函数参数矩阵 $\boldsymbol{\alpha}$ 的初始化值. soft-max 优化目标函数如式(14)所示.

$$\begin{aligned}J(\boldsymbol{\alpha})=&-\frac{1}{N}\left[\sum_{n=1}^N\sum_{l=1}^L 1\{y^{(n)}=l\}\cdot\log\frac{\mathrm{e}^{\boldsymbol{\alpha}_l^{\mathrm{T}}\boldsymbol{x}^{(n)}}}{\sum_{l=1}^L\mathrm{e}^{\boldsymbol{\alpha}_l^{\mathrm{T}}\boldsymbol{x}^{(n)}}}\right]\\ &+\frac{\lambda}{2}\sum_{m=1}^M\sum_{l=1}^L\alpha^2_{m,l}\end{aligned}\tag{14}$$

其中,N 为训练数据中所有超像素个数,$L=3$ 为本文标签类别个数(天空、立体物、地面),M 为超像素特征向量的维数,$y^{(n)}$ 表示第 n 个超像素的实际标签,$\boldsymbol{x}^{(n)}$ 表示第 n 个超像素的特征向量,$\boldsymbol{\alpha}_l$ 表示特征向量在 l 第个类别标签上的影响参数,$\alpha_{m,l}$ 为第 m 维特征在第 l 个类别标签上的影响参数,λ 为正则化项权重值.

为了对上述极大似然估计算法中两个权重系数矩阵进行训练,本文采用分段训练和联合训练相结合的方式,亦即在一元势函数参数矩阵 $\boldsymbol{\alpha}$ 固定的情况下对二元势函数参数矩阵 $\boldsymbol{\beta}$ 进行训练,然后对二者同时训练. 本文采用的极大似然估计代价函数如式(15)所示. 其中,R 为相邻两超像素之间的差异性特征向量维数,$\boldsymbol{x}^{(n,k)}$ 为第 n 个与第 k 个超像素之间的差异性特征向量,A_n 为第 n 个超像素的邻接超像素集合,$u^{(n)}$ 为第 n 个超像素在当前参数下由 mean-field 推断[24]得到的第 n 个超像素的类别标签分布,$\boldsymbol{\beta}$ 为二元势函数参数矩阵,λ 和 δ 分别为一元和二元参数矩阵正则化项权重值.

$$\begin{cases}J(\boldsymbol{\alpha},\boldsymbol{\beta})=\dfrac{1}{N}\displaystyle\sum_{n=1,k\in A_n}^N\{C-\sum_{l=1}^L u^{(n)}\cdot\log(u^{(n)})\}\\ \qquad\qquad+\dfrac{\lambda}{2}\displaystyle\sum_{m=1}^M\sum_{l=1}^L\alpha^2_{m,l}+\dfrac{\delta}{2}\sum_{r=1}^R\sum_{l=1}^L\beta^2_{r,l}\\ C=\displaystyle\sum_{l=1}^L\Big[(u^{(n)}-y^{(n)})\cdot\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{x}^{(n)}+(u^{(n)}-y^{(n)})\\ \qquad\cdot\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{x}^{(n,k)}(u^{(k)}-y^{(k)})\Big]\end{cases}\tag{15}$$

3.2　基于多次图像分割的图像 3D 空间布局理解模型

为了避免单次超像素分割的不准确性对模型推断准确性的影响,本文提出基于多次图像分割的图像 3D 空间布局理解模型,通过多次图像分割算法生成多个不同尺度的超像素图像,并对各尺度的超像素图像分别建立条件随机场模型,然后通过 D-S 证据合成理论集

成多个模型推断结果,提高 3D 空间布局理解的准确性.

3.2.1 基于多次分割的多尺度超像素图像生成

本文采用文献[5]提出的多次分割框架生成多个不同尺度的超像素图像,每次图像分割结果都对应一个尺度的超像素图像. 多次分割算法的原理是假设初始分割获得的图像分割结果为 R_0,其包含 N_0 个超像素. 事先设定目标超像素为 N_i 个,在 N_0 个超像素中随机选择 N_i 个超像素并赋予超像素标号 $1,\cdots,N_i$,对 N_0 没有被标号的超像素 s 遍历与其相邻的超像素集合 A,通过超像素 s 与集合 A 中每个超像素的相似度以及集合 A 中每个超像素的标号情况,计算超像素 s 被标记为各个标号的概率值,取最大概率对应的标号作为超像素 s 的标号,由此则获得第 i 次分割图像 R_i. 类似地,通过设定不同的目标超像素个数 N_i,即可获得多个新的分割图像 $R_i(i=1,2,\cdots)$. 在上述确定超像素 s 标号的过程中,本文同样使用 3.1.2 节中由 SVM 分类器输出结果经归一化获得的相邻两超像素相似程度的置信度值 $x_{i,j}$ 描述超像素 s 与集合 A 中每个超像素之间的相似性程度.

本文设定对单次分割图像 R_0 合并后的目标超像素个数分别为[60,70,80,90,100,120,140,150,170,180,200,240,260,280,300]共 15 种方式对超像素进行合并,获得 15 个新的分割图像 $R_i\{i=1,\cdots,15\}$. 加上初始分割图像 R_0 本文共生成 16 个尺度的超像素图像.

3.2.2 基于多个条件随机场模型集成的图像 3D 空间布局理解算法

对应上述每个尺度的超像素图像,可分别建立一个 3.1.2 节中所述的条件随机场模型,从而获得同一场景的多个独立推断结果. 本文借鉴集成学习[12]的思想,采用 D-S 证据理论[25]来集成多个不同条件随机场模型的推断结果. 值得注意的是本文提出的多条件随机场模型集成方法中,每个条件随机场模型由其对应尺度的超像素图像提供训练样本进行训练,由此可通过训练样本的不同来确保各个条件随机场模型之间的差异性.

记 CRF_0 为对应初始分割图像 R_0 的条件随机场模型,$\mathrm{CRF}_n\{n=1,\cdots,15\}$ 分别为对应 15 个新的分割图像 $R_i\{i=1,\cdots,15\}$ 的条件随机场模型. 由于 15 种不同合并方式得到的分割图像彼此之间是相互独立的,因此,其对应的条件随机场模型之间是相互独立的,则它们的推断结果满足证据之间相互独立的条件,可将它们的推断结果应用 D-S 证据合成理论进行集成. 令超像素内包含的像素点的标签分布与超像素的标签分布相同,记 N 个条件随机场推断结果融合后,像素点 pix_j 属于标签 l 的概率为 $\boldsymbol{m}(l)$,则其集成公式如式(16)所示. 其中,$S_i,\cdots,S_n,\cdots,S_N$ 表示在 N 次图像分割结果中,像素点 pix_j 所在的超像素,L 为标签个数,$p_{sn}^1,\cdots,p_{sn}^l,\cdots,p_{sn}^L$ 为基于第 n 次分割建立的条件随机场推断结果中像素点 pix_j 所在超像素属于各个标签的概率分布.

$$\begin{cases}\boldsymbol{m}(l)=\dfrac{\prod_{n=1}^{N}p_{S_n}^{l}}{1-K}\\ K=\sum_{o_1=1}^{L}\cdots\sum_{o_n=1}^{L}\cdots\sum_{o_N=1}^{L}p_{S_1}^{o_1}p_{S_2}^{o_2}\cdots p_{S_n}^{o_n}\cdots p_{S_N}^{o_N}-\sum_{o=1}^{L}\prod_{n=1}^{N}p_{S_n}^{o}\end{cases}\tag{16}$$

令式(16)中 $N=15$,将上述 15 个条件随机场模型 $\mathrm{CRF}_n\{n=1,\cdots,15\}$ 推断结果进行集成,对集成后像素点 pix_j 属于各类标签的概率分布进行归一化:

$$\boldsymbol{m}(l)_{\mathrm{norm}}=\frac{\boldsymbol{m}(l)}{\sum_{q=1}^{L}\boldsymbol{m}(q)}\tag{17}$$

在得到归一化结果 $\boldsymbol{m}(l)_{\mathrm{norm}}$ 后,将其与 CRF_0 的推断结果再次应用式(16)进行二次集成,最终得到像素点 pix_j 属于标签 l 的概率.

4 实验结果

为验证所提出算法有效性,本文在 GC[5] 和 KITTI Layout 数据集[26]上对所提出算法进行了评价,并与其它算法进行了对比. 其中,GC 数据集共包含 300 张自然场景图像,被 3D 布局理解算法研究者广泛使用. 为与其它方法比较,本文按照文献[5]的实验数据集划分方式进行训练和交叉验证实验. 表 1 给出以像素为评估单位的天空、立体物及地面的查全率(Recall)、平均准确率(Average Accuracy)、错误率(Error rate)、查准率(Precision)[27]以及查错率(1-Precision)[28]指标的评价结果. 各指标计算公式如下:

$$\mathrm{Recall}_l=\frac{TP_l}{P_l}\tag{18}$$

$$\mathrm{Accuracy}=\frac{\sum_{l=1}^{L}TP_l}{\sum_{l=1}^{L}P_l}\tag{19}$$

$$\mathrm{Error\_rate}=1-\mathrm{Accuracy}\tag{20}$$

$$\mathrm{Precision}_l=\frac{TP_l}{TP_l+FP_l}\tag{21}$$

其中,TP_l、P_l 分别表示正确识别的和实际的第 l 类样本个数,$L=3$ 为标签类别数.

表 1　不同算法在 GC 数据集上各类别查准率、查错率、查全率、平均准确率及错误率(单位:%)

算法	查准率			查错率			查全率			平均准确率	错误率
	天空	立体物	地面	天空	立体物	地面	天空	立体物	地面		
文献[4]算法	94.6	81.6	86.5	5.4	18.4	13.5	90.0	89.0	78.0	86.0	14.0
文献[5]算法	94.6	85.8	87.3	5.4	14.3	12.7	90.0	90.0	84.0	88.1	11.9
文献[6]算法(vlGT)	--	--	--	--	--	--	--	--	--	89.0	11.0
文献[6]算法(vlAuto)	--	--	--	--	--	--	--	--	--	87.0	13.0
文献[7]算法	--	--	--	--	--	--	--	--	--	87.9	12.1
文献[8]算法	93.0	86.3	84.6	7.1	13.7	15.4	93.0	89.0	84.0	88.4	11.6
文献[9]算法	--	--	--	--	--	--	--	--	--	86.9	13.1
文献[10]算法	86.3	86.3	84.6	13.7	13.7	15.4	93.0	84.0	78.0	84.0	16.0
Base line	83.9	89.2	87.3	16.1	10.8	12.7	97.9	85.3	84.7	87.4	12.6
BL + C	87.3	89.0	87.4	12.7	11.0	12.6	97.7	86.7	85.1	88.1	11.9
BL + C + LBP	93.9	87.2	88.9	6.1	12.8	11.1	93.0	89.9	85.6	88.9	11.1
BL + C + LBP + ALM	92.4	88.5	88.3	7.6	11.5	11.7	95.0	89.3	83.1	89.2	10.8
SSCRF - SVM	82.0	90.5	88.8	18.0	9.5	11.2	98.9	84.8	86.5	88.0	12.0
SSCRF	**94.6**	**87.7**	**89.0**	**5.4**	**12.3**	**11.0**	**93.1**	**90.4**	**86.3**	**89.4**	**10.6**
MSCRF	**96.0**	**89.3**	**89.2**	**4.0**	**10.7**	**10.8**	**95.3**	**90.7**	**87.6**	**90.6**	**9.4**

注"--"表示对应算法没有给出相应的评估值. Base line:文献[5]特征 + CRF + SVM 边缘判别;BL + C:Base line 基础上采用 chi 方度量特征向量之间的距离;BL + C + LBP:BL + C 追加 LBP 特征;BL + C + LBP + ALM:BL + C + LBP 追加完整的 LM 滤波器组;SSCRF:Single Segment CRF,本文基于单次分割的算法,BL + C + LBP + ALM 追加颜色矩特征;SSCRF - SVM:SSCRF 中不使用 SVM 分类器进行边缘判别;MSCRF:Multiple Segment CRF,本文基于多次分割的算法. 文献[6]算法(vlGT)为使用人工标记消失线所得结果,文献[6]算法(vlAuto)为使用动态计算消失线所得结果. 文献[8]算法为技术报告结果.

本文也给出了文献[4~10]算法的实验结果. 从表 1 可以看出,本文提出的基于多次图像分割的图像 3D 空间布局理解算法(记为 MSCRF,见第 3 节描述)在各类别查准率、查全率以及平均准确率较以往算法都有明显提升,并且错误率也显著降低. 表 1 中也给出了基于单次图像分割的图像 3D 空间布局理解模型(记为 SSCRF,见第 2 节描述),结果显示其效果也明显好于其它算法. 同时,为了说明本文提出的 chi 方距离度量、各种超像素特征以及采用 SVM 度量相邻超像素相似性的有效性,表 1 中也分别给出了不同的实验结果. 从结果也可以看出,这些方法和特征对提升 3D 空间布局理解平均准确率也具有明显作用. 值得注意的是,这里,为公平比较,本文方法和对比文献算法均采用了相同的动态消失线计算方法.

表 2 给出了本文提出算法在 KITTI Layout 数据集[26]上的评估结果. 受篇幅所限,这里仅将与本文同样采用多次分割框架且结果表现最好的文献[5]作为对比. 从结果可以看出,本文提出的算法较文献[5]算法提升了 3D 空间布局理解各项指标. 值得注意的是,由于文献[8]没有提供其算法中用于获取上下文先验信息的未标记图像数据集,所以这里无法与其进行比较.

表 3 给出了本文提出的 3D 空间布局理解算法各类别查全率和平均准确率的统计性评估结果. 限于篇幅,本文只在 GC 数据集上重复 5 次 5 组交叉验证实验,从实验结果可以看出,本文提出的方法同样优于对比方法.

表 2　不同算法在 KITTI Layout 数据集上各类别查准率、查错率、查全率、平均准确率及错误率(单位:%)

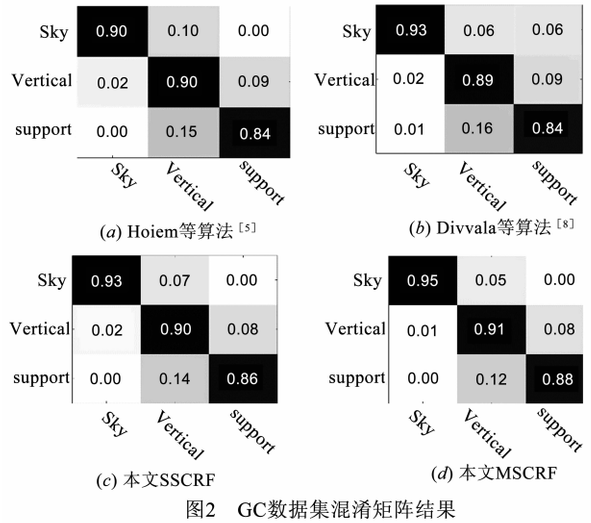
算法	查准率			查错率			查全率			平均准确率	错误率
	天空	立体物	地面	天空	立体物	地面	天空	立体物	地面		
文献[5]算法	92.3	96.1	52.5	7.7	3.9	47.5	89.0	78.4	93.4	82.1	17.9
SSCRF	**93.1**	**97.8**	**56.8**	**6.9**	**2.2**	**43.2**	**94.9**	**79.3**	**96.2**	**84.2**	**15.8**
MSCRF	**93.5**	**98.6**	**56.8**	**6.5**	**1.4**	**43.2**	**96.8**	**84.8**	**97.2**	**85.6**	**14.4**

表 3 不同算法在 GC 数据集上各类别的查全率和平均准确率重复实验统计性和算法显著性评估(单位:%,除显著性检验值)

算法	查全率均值			查全率标准差			平均准确率均值	平均准确率标准差	显著性检验值			
	天空	立体物	地面	天空	立体物	地面			z_{sky}	$z_{vertical}$	$z_{support}$	$z_{average}$
文献[5]算法	90.2	89.9	84.0	0.9	0.7	0.6	88.1	0.4				
SSCRF	**93.9**	**90.5**	**85.4**	**1.8**	**0.5**	**2.3**	**89.3**	**0.4**	**8.9**	**1.9**	**5.2**	**6.4**
MSCRF	**95.4**	**90.7**	**87.7**	**0.7**	**0.3**	**0.2**	**90.6**	**0.2**	**12.9**	**2.6**	**13.8**	**13.3**

进一步地,为了验证提出算法与对比算法相比是否有显著提升,本文参照文献[29]对提出的基于单次分割和基于多次分割的算法推断所得各类别的查全率以及平均准确率统计结果(见表 3)分别进行了显著性检验. 其中,本文显著性水平取值为 $\alpha=0.05$,由表 3 计算所得 SSCRF、MSCRF 分别相对于文献[5]算法的显著性检验值都大于 $z_{\alpha=0.05}=1.645$,这证明本文提出的算法与文献[5]提出的算法相比有显著提升. 采用相同方式同样可以证明本文提出的基于多次分割的算法相比基于单次分割的算法也有显著提高.

图 2 分别给出了本文提出的 MSCRF,SSCRF 算法以及参考文献[5,8]算法的混淆矩阵评估结果. 从比较结果可以看出,本文提出的算法有效减小了天空、立体物以及地面被误识为其他类别的比例.



(a) Hoiem等算法[5] (b) Divvala等算法[8]

(c) 本文SSCRF (d) 本文MSCRF

图2 GC数据集混淆矩阵结果

图 3 给出了本文提的出基于多条件随机场模型的图像 3D 空间布局理解算法在 GC 数据集上的代表图像实验结果,图中同样给出了 Hoiem 等提出的算法结果作为对比. 从图中可以看出,本文提出的算法对天空、立体物以及地面的正确识别都有很大提升,识别结果更接近 Hoiem 等人提供的 Ground Truth 数据.

表 4 给出了本文算法处理图像的平均时间性能以及根据图像复杂程度不同的处理时间范围,作为对比,也给出了文献[5]的处理时间. 所有算法都采用相同的

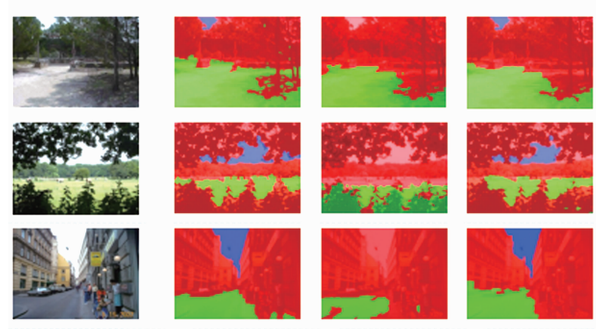


图3 文献[5]算法及本文算法在GC上的实验结果,图中从左到右依次为原图、Ground truth、Hoiem.et.al[5]、MSCRF. 蓝色为天空,红色为立体物,绿色为地面

硬件 i7-3770 CPU @3.40GHz 和 MATLAB 处理语言. 从实验结果可以看出,本文提出的 SSCRF 算法与文献[5]的算法在计算效率上具有可比性. 本文提出的 MSCRF 算法虽然处理时间大幅增加,但在引入并行操作后(将基于多次分割分别建立条件随机场模型推断的过程实现并行)算法效率显著提高.

表 4 GC 数据集上不同算法的效率评估(640×480)

算法	平均处理时间	处理时间范围
文献[5]算法	10s	9s~11s
SSCRF	14s	13s~18s
MSCRF	570s	552s~582s
MSCRF(并行)	114s	110s~116s

5 结束语

本文提出一种图像 3D 空间布局理解算法. 该算法利用多次图像分割生成多个不同尺度的超像素图像,并结合纹理、颜色、位置和形状特征,建立超像素特征表达. 在此基础上,构建各尺度超像素图像对应的条件随机场模型,并应用 D-S 证据合成理论对多个模型推断结果进行集成,实现对图像 3D 空间布局的理解. 不同数据集上的实验结果表明,同已有算法相比,本文提出的算法明显提高了图像 3D 空间布局理解中各类别的查准率、查全率和平均准确率.

参考文献

[1] Alvarez J M, Gevers T, Lopez A M. 3D scene priors for

road detection[A]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. San Francisco,CA:IEEE Press,2010. 57 - 64.

[2] Alvarez J M,Lopez AM,Gevers T. Combining priors,appearance,and context for road detection[J]. IEEE Transactions on Intelligent Transportation Systems,2014,15(3):1168 - 1178.

[3] Hai W,Chao-chun Y,Ying-feng C. Smart road vehicle sensing system based on monocular vision[J]. Journal for Light and Electron Optics,2015,126(4):386 - 390.

[4] Hoiem D,Efros A A,Hebert M. Geometric context from a single image[A]. Proceedings of Tenth IEEE International Conference on Computer Vision (ICCV)[C]. USA:IEEE,2005. 654 - 661.

[5] Hoiem D,Efros A A,Hebert M. Recovering surface layout from an image[J]. International Journal of Computer Vision,2007,75(1):151 - 172.

[6] Hoiem D,Efros A A,Hebert M. Closing the loop in scene interpretation[A]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition[C]. Anchorage,AK:IEEE Press,2008. 1 - 8.

[7] Divvala S K,Efros A A,Hebert M. Can similar scenes help surface layout estimation[A]. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops[C]. USA:IEEE Press,2008. 1 - 8.

[8] Divvala S K,Efros A A,Hebert M. Unsupervised Patch-Based Context From Millions of Images[R]. Technology Report CMU - RI-TR-11-38,Robotics Institute:Carnegie Mellon University,2011.

[9] Gould S,Fulton R,Koller D. Decomposing a scene into geometric and semantically consistent regions[A]. Proceedings of IEEE 12th International Conference on Computer Vision[C]. Kyoto:IEEE Press,2009. 1 - 8.

[10] Lazebnik S,Raginsky M. An empirical Bayes approach to contextual region classification[A]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition[C]. Miami,FL:IEEE Press,2009. 2380 - 2387.

[11] Lafferty J,McCallum A,Pereira F. Conditional random fields:Probabilistic models for segmentation and labeling sequence data[A]. Proceedings of International Conference on Machine Learning[C]. USA:Morgan Kaufmann,2001. 282 - 289.

[12] Zhi-Hua Zhou. Ensemble Methods:Foundations and Algorithms[M]. USA:CRC Press,2012.

[13] 姚旭,王晓丹,张玉玺. 基于随机子空间和 Adaboost 的自适应集成方法[J]. 电子学报,2013,41(4):810 - 814.
YAO Xu,WANG Xiao dan,ZHANG Yu xi. A self-adaptation ensemble algorithm based on random subspace and Adaboost[J]. Acta Electronica Sinica,2013,41(4):810 - 814. (in Chinese)

[14] 毕凯,王晓丹,姚旭. 一种基于 Bagging 和混淆矩阵的自适应选择性集成[J]. 电子学报,2014,42(4):711 - 716.
BI Kai,WANG Xiaodan,YAO Xu. Adaptively selective ensemble algorithm based on bagging and confusion matrix[J]. Acta Electronica Sinica,2014,42(4):711 - 716. (in Chinese)

[15] 朱鹏飞,等. 基于随机化属性选择和邻域覆盖约简的集成学习[J]. 电子学报,2012,40(2):273 - 279.
ZHU Peng-fei,et al. Ensemble learning based on randomized attribute selection and neighborhood covering reduction[J]. Acta Electronica Sinica,2012,40(2):273 - 279. (in Chinese)

[16] Ping Zhong,Runsheng Wang. A multiple conditional random fields ensemble model for urban area detection in remote sensing optical images[J]. IEEE Transactions on Geoscience and Remote Sensing,2007,12(45):3978 - 3988.

[17] Stricker M,Orengo M. Similarity of color images[J]. Proceedings of SPIE Storage and Retrieval for Image and Video Databases,1995,2420:381 - 392.

[18] Thomas L,Jitendra M. Representing and recognizing the visual appearance of materials using three-dimensional textons[J]. International Journal of Computer Vision,2001,43(1):29 - 44.

[19] Ojala T,Pietik? inen M. Multi-resolution gray scale and rotation invariant texture classification with local binary patterns[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2002,24(7):971 - 987.

[20] Kosecka J,Zhang W. Video compass[A]. Proceedings of ECCV[C]. Berlin:Springer - Verlag,2002. 1 - 15.

[21] Guo G,Li S Z,Chan K L. Face recognition by support vector machines[A]. Proceedings of International Conference on Automatic Face and Gesture Recognition[C]. Grenoble:IEEE Press,2000. 196 - 201.

[22] Liu Cheng - Lin. Classifier combination based on confidence transformation[J]. Pattern Recgonition,2005,38(1):11 - 28.

[23] Bishop C M. Neural Networks for Pattern Recognition[M]. USA:Oxford University Press,2004.

[24] Jordan M I,Ghahramani Z,Jaakkola T S,et al. An introduction to variational methods for graphical models[J]. Machine Learning,1999,37(2):183 - 233.

[25] Shafer G. A Mathematical Theory of Evidence[M]. Princeton,NJ:Princeton University Press,1976.

[26] José M álvarez,Yann LeCun,Theo Gevers,Antonio M

López. Road scene segmentation from a single image [A]. Proceedings of European Conference on Computer Vision (ECCV) [C]. Florence: Springer - Verlag, 2012. 1 - 14.

[27] 秦锋,杨波,程泽凯. 分类器性能评价标准研究[J]. 计算机技术与发展,2006,16(10):85 - 88.
Qin Feng, Yang Bo, Cheng Ze - kai. Research on measure criteria in evaluating classification performance[J]. Computer Technology and Development, 2006, 16(10): 85 - 88. (in Chinese)

[28] 何周灿,王庆. 一种用于图像匹配的快速有效的二分哈希搜索算法[J]. 西北工业大学学报,2010,28(4):609 - 615.
He Zhoucan, Wang Qing. A fast and effective dichotomy - based hash (DBH) search algorithm for image matching[J]. Journal of Northwestern Polytechnical Universty, 2010, 28(4): 609 - 615. (in Chinese)

[29] 盛聚,谢千式. 概率论与数理统计[M]. 北京:高等教育出版社,2008.

作者简介

刘 威 男,1975 年 6 月出生,辽宁沈阳人,博士,东北大学副教授,教授级高级工程师. 主要研究领域为汽车辅助驾驶、智能交通、智能安防,研究方向为计算机视觉、图像处理、模式识别.
E-mail: lwei@ neusoft. com

周 婷 女,1990 年 2 月出生,四川乐山人,东软汽车电子研究员. 主要研究领域为汽车辅助驾驶、自动驾驶,研究方向机器学习.
E-mail: zhou _ ting@ neusoft. com