

基于透视 N 点模型的实时单目深度估计方法

郭克友^{1*}, 杨民¹, 张沫², 郭晓丽¹, 李雪¹

¹北京工商大学人工智能学院, 北京 100048;

²交通运输部公路科学研究院, 北京 100088

摘要 行车间距检测是汽车主动安全辅助驾驶系统的关键技术之一, 为了提高车辆行驶过程中行车间距检测的精度与实时性, 提出了一种实时单目深度估计方法。首先, 构建畸变模型并用相机标定算法进行单目相机标定。然后, 以车牌作为前车目标定位基准, 采用颜色、轮廓的车牌筛选算法快速提取前车车牌信息。最后, 基于方向梯度直方图特征和支持向量机实现车牌的精准定位。实验结果表明, 相比其他方法, 融合已知车牌的透视 N 点深度估计模型精度高、实时性好。本方法对前车车牌定位的识别率为 99.326%, 行车间距的检测误差小于 10%, 处理一张图像所需的时间约为 170 ms, 满足车辆行驶过程中对车间距检测的应用需求。

关键词 机器视觉; 深度模型; 机器学习; 支持向量机

中图分类号 TP391.4

文献标志码 A

doi: 10.3788/LOP202158.0615005

Real-Time Monocular Depth Estimation Method Based on Perspective N -Point model

Guo Keyou^{1*}, Yang Min¹, Zhang Mo², Guo Xiaoli¹, Li Xue¹

¹School of Artificial Intelligence, Beijing Technology and Business University, Beijing 100048, China;

²Research Institute of Highway Ministry of Transport, Beijing 100088, China

Abstract Driving distance detection is one of the key technologies of the car's active safety driving assistance system. This paper proposes a method of real-time monocular depth estimation, to improve the accuracy and real-time performance of the distance detection during the vehicle driving process. First, the distortion model is constructed, and the camera calibration algorithm is used to calibrate the monocular camera. Then, the license plate is used as the target location benchmark of the front vehicle, and the license plate information of the front vehicle is extracted quickly by using the license plate filtering algorithm of color and contour. Finally, based on directional gradient histogram feature and support vector machine, the license plate is located accurately. Experimental results show that, compared with other methods, the perspective N -point depth estimation model fused with known license plates has high accuracy and good real-time performance. This method has a recognition rate of 99.326% for the location of the license plate of the front vehicle, and the detection error of the driving distance is less than 10%, and the time required to process an image is about 170 ms, which meets the application requirements for the detection of the distance between vehicles during the vehicle driving process.

Key words machine vision; deep model; machine learning; support vector machine

OCIS codes 150.0155; 150.5670; 100.4999; 110.2960

收稿日期: 2020-07-24; 修回日期: 2020-08-27; 录用日期: 2020-09-08

基金项目: 交通运输行业重点科技项目清单(2018-ZD1-010)、北京工商大学 2020 年研究生科研能力提升计划

* E-mail: guoky@th.btbu.edu.cn

1 引言

世界卫生组织发布的《2018 年全球道路安全报告》指出,全球每年因道路交通事故死亡的人数多达 135 万,道路交通事故成为 5~29 岁儿童和年轻人死亡的首要原因^[1]。严重的道路交通事故已成为全球各国必须共同面对的问题,而传统的汽车安全装置如安全带、安全气囊属于被动安全装置,只能降低事故发生时人员和车辆的损伤程度,不能有效预防事故的发生。因此,将传统的被动安全技术发展成以事故预防为核心的主动安全技术具有重要意义。主动安全技术的关键是及时探测汽车行驶时前方的车辆、障碍物等,向驾驶员反馈其准确位置信息的同时做出相应处理,以避免事故发生。

常见的前方车辆测距方法有激光雷达测距、机器视觉测距等。激光雷达测距应用范围广、抗干扰强,但存在价格高、体积大等不足^[2]。机器视觉测距具有获取外界信息丰富、造价低以及可适应于前方同车道内车辆位置近距离测距等优点,成为前车近距离探测领域的研究热点。机器视觉测距分为双目(多目)视觉测距和单目视觉测距,双目视觉测距的原理与人眼类似,固定好两个摄像头的位置后,利用双目视差原理进行测距。丁萌等^[3]在无监督学习框架下,将深度估计问题转化为图像重建问题,根据双目图像的三角测量原理实现深度估计,但该方法得到的最终图像对应的是像素级深度。李阳等^[4]指出双目视觉测距在纹理稀疏的场景中难以找到足够多的匹配特征,从而出现较大误差甚至失效。双目视觉的局限性,限制了双目视觉深度估计的应用范围,因此人们逐渐开始转向对单目视觉测距的研究。

对于单目视觉测距的研究中,张法全等^[5-6]先用目标图像模糊度检测方法获取边缘宽度,然后通过样条插值算法求取目标物体的深度信息,但该方法只能测量近距离深度信息。王剑等^[7]通过改变相机焦距获取多张图像,在固定尺寸和焦距两种情况下成像,利用目标物体尺寸和焦距之间的关系,定量求取深度信息,但该方法要求对相机进行在线标定,不满足实时性的要求。黄风山等^[8]采用与相机主动标定类似的原理,控制相机进行多组平移或旋转运动,从而测定目标物体的深度,但该方法不适用于相机固定的场所。

近年来,人们开始用机器学习、深度学习等方法测量单目图像的深度。Liu 等^[9]提出了一种基于深度卷积神经网络(CNN)的等效模型深度估计方法,

利用深度值连续的特性,将深度估计问题转化为条件随机场(CRF)的学习问题并求解,从而预测图像的深度。但该方法不能利用任何的几何条件,求解的深度不够精确。Cao 等^[10]将深度估计问题当作像素分类问题,首先将深度值进行离散化,然后通过卷积深度残差网络对每个像素的深度类别进行预测,最后得到一个深度估计信息的概率分布图。袁建中等^[11]提出了一种基于残差网络(ResNet)的深度 CNN 模型,模型的所有层用密集连接卷积网络(DenseNet)直接连接在一起,最后通过神经网络生成道路深度信息的预测图。但该方法最后得到的深度信息只是一个深度范围,并不是确定值。

上述方法不能满足车辆在行驶过程中对于实时测距的需求,因此,本文提出了一种基于透视 N 点投影(PNP)模型的实时单目深度估计方法,对前车车距进行实时检测,并以实际道路图像与仿真模型对该方法进行验证。实验结果表明,PNP 单目深度估计模型的建立,可为今后智能车辆实现无人驾驶奠定理论基础。

2 本方法的整体框架

图 1 为本方法的总体框架,主要包括单目相机标定、车牌粗定位、车牌精确定位以及深度估计模型四部分。首先,对单目相机进行标定,获取相机的内参属性;然后,引入浅层机器学习对车牌进行快速筛选并提取车牌的方向梯度直方图(HOG)特征;其次,用支持向量机(SVM)选参策略对提取的车牌 HOG 特征进行精确定位;最后,提出三种深度估计模型。对比发现,融入已知车牌信息的 PNP 模型结合相机位姿可实时检测前方车距,且检测精度高、实时性好。

3 车牌定位

3.1 单目相机标定

实验以惠普 F515 行车记录仪相机作为单目相机,该相机属于鱼眼相机,具有视角大、成本低、性能稳等特点。但鱼眼相机的结构特点导致得到的物体图像存在畸变,因此需要对畸变图像进行校正后再标定。首先对单目相机进行初次标定,获取相机内参数和畸变系数;然后构建畸变模型,并用双线性插值法对畸变图像进行校正;最后按照马颂德等^[12]提出的移动平面模板标定算法得到标定误差,并将校正后的图像再次按照普通相机模型进行标定,得到校正后的相机参数。单目相机的标定流程如图 2 所示。

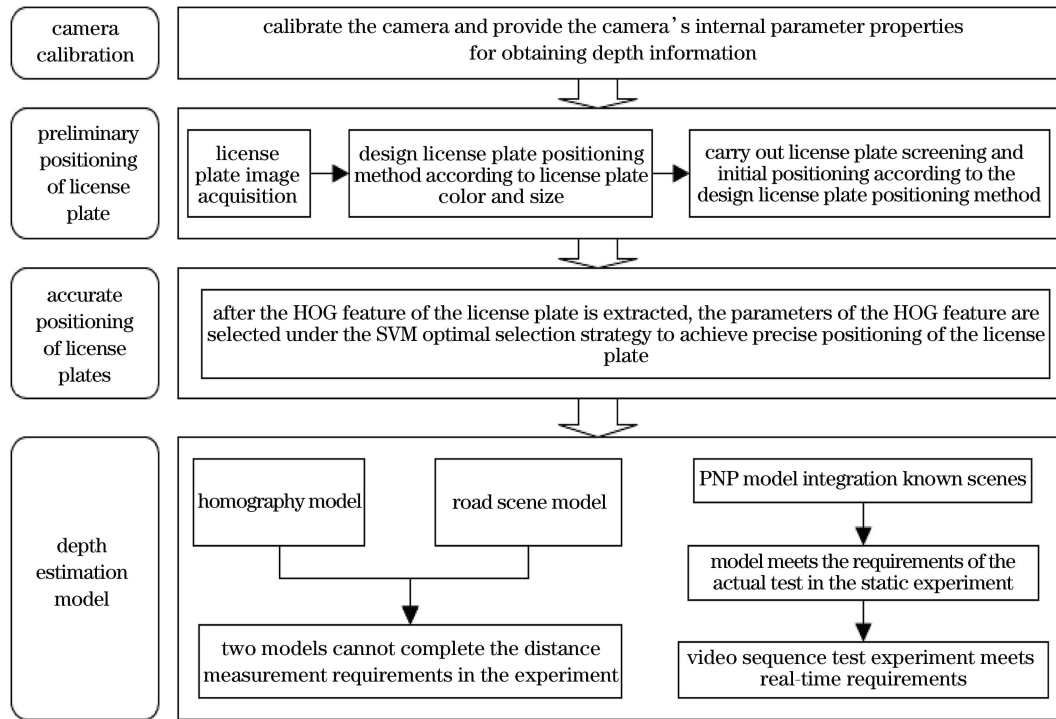


图 1 本方法的具体实现流程

Fig. 1 Specific implementation process of our method

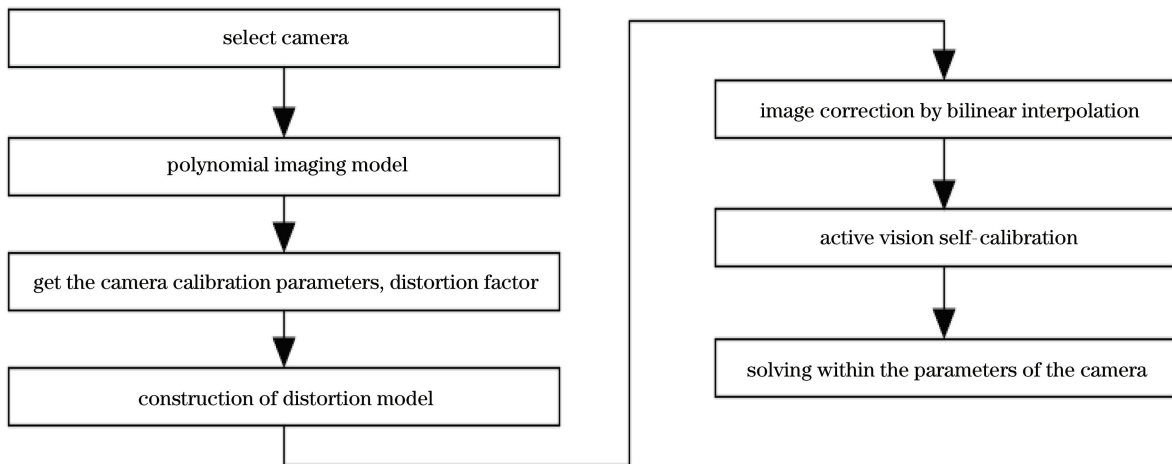


图 2 相机标定流程图

Fig. 2 Flow chart of the camera calibration

3.2 车牌粗定位

在车牌粗定位中,为提高实时检测的识别率,提出了一种综合颜色、垂直纹理、车牌大小、边缘等信息的车牌定位算法,具体流程如图 3 所示。

针对车牌具有的颜色特点,提出利用颜色空间和 Sobel 综合搜索方法定位车牌图像。首先,通过颜色匹配,将处在黄色和蓝色阈值空间内的像素点

设置为 255,其他像素点设置为 0;然后,将得到的二值图像按照特定矩形模板进行开运算处理;最后,将原来分散的区域聚合,得到完整的二值化连通区域。为判断区域的合理性,对该区域进行轮廓查找并计算最小外接矩形及其尺寸、长宽比、倾斜角度。经过上述步骤,在给定的原始目标图像上可获得初步筛选的图像块,结果如图 4 所示。

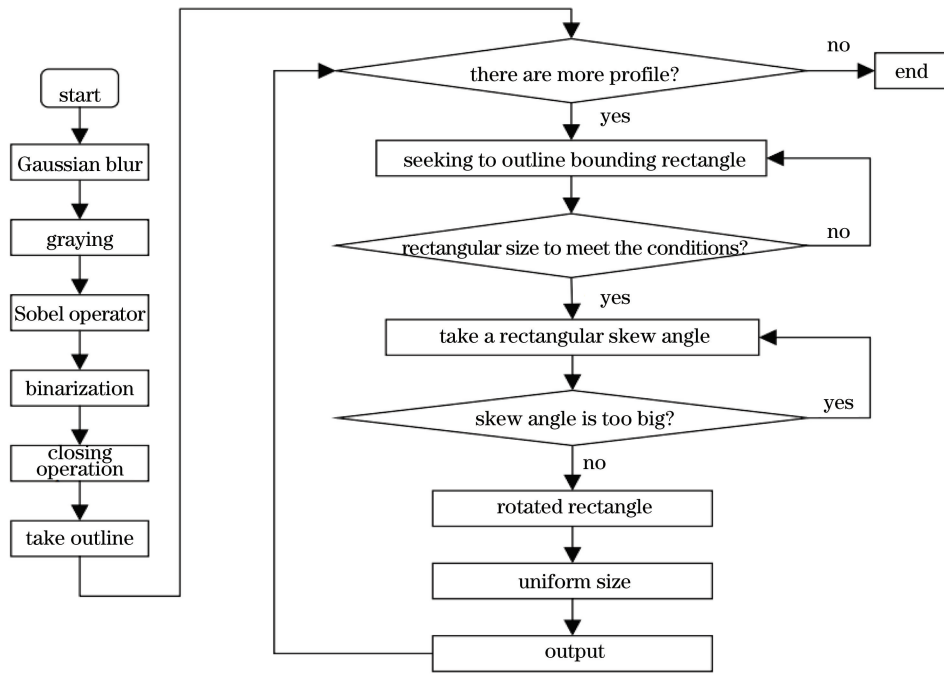


图 3 车牌定位算法的流程图

Fig. 3 Flow chart of the license plate location algorithm



图 4 初步筛选的车牌图像

Fig. 4 License plate images obtained through preliminary screening

3.3 车牌的精确定位

1) 车牌的特征提取

初步筛选出车牌图像后,需对车牌特征进行提取,以保证在不同背景下均能完成对车牌的识别。提取车牌图像的颜色直方图特征时发现,车牌图像的颜色具有高度相似性,仅利用颜色直方图特征很难完成对车牌图像的判断。文献[13]的研究结果表明,单独使用 Haar 特征时漏检率、正检率较低,单独使用 HOG 特征可在两者中得到较好的结果。文献[14]分析 HOG 特征时发现,HOG 特征着重使用向量描述待检测物体的外形,利用梯度检测时有较好的抗几何和光学变化能力。因此,实验选用 HOG 特征提取车牌特征,根据 Dalal 等^[15]提出的 HOG 特征计算流程提取车牌的 HOG 特征。

2) 车牌的精确定位

对获取的车牌 HOG 特征进行归一化处理,用基于 SVM 的方法对车牌进行精确定位,获取车牌角点的位置信息。文献[16]中的 SVM 在实际使用中,先用粒子群优化(PSO)算法和遗传算法(GA)

确定初步最优解范围,再用网格法进行高精度运算,求解精确解。基于该策略可以充分利用 SVM 的分类性能、加快 SVM 的收敛速度、提高最优解的精度。

4 行车间距深度估计模型

4.1 单应性模型

单应性模型主要根据单应性原理测量位姿,实验通过构建单目相机和车牌之间的单应性矩阵 H 确定实际的位置和姿态。单应性模型的原理如图 5 所示,车牌在实际物体上方建立世界坐标系 $O_w-X_wY_wZ_w$,此时世界坐标系与相机坐标系 $O_c-X_cY_cZ_c$ 之间的关系为 (R_1, T_1) ,其中, R, T 分别为旋转与平移矩阵。同理,该车牌在实际物体下方建立的世界坐标系 $O'_w-X'_wY'_wZ'_w$ 与相机坐标系之间的关系为 (R_2, T_2) 。根据刚体平移旋转,得到两车牌平面之间存在的单应性矩阵 H 满足

$$[x_w \ y_w \ z_w]^T = R_2^{-1}R_1 [x'_w \ y'_w \ z'_w]^T + R_2^{-1}(T_1 - T_2), \quad (1)$$

式中, $[x_w \ y_w \ z_w]^T$ 为坐标系 (X_w, Y_w, Z_w) 中

车牌上点的坐标, $[x'_w \ y'_w \ z'_w]^T$ 为同一个点在坐标系 $O'_w-X'_wY'_wZ'_w$ 中的坐标。

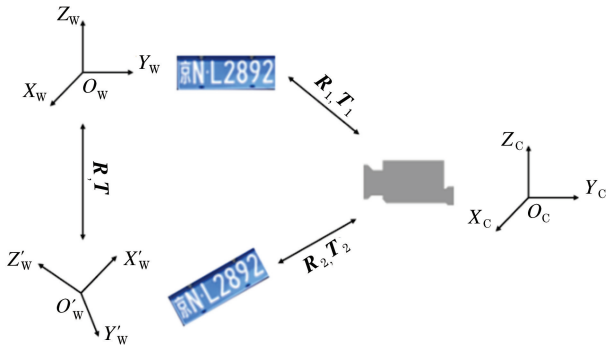


图 5 单应性模型的原理图

Fig. 5 Schematic diagram of the homography model

实际测距中, 车牌的尺寸信息是已知的, 假定世界坐标系中的模板车牌图像平面(I_1)与单目相机拍摄的车牌图像平面(I_1^*)已知, 可以根据两个图像平

面的单应性矩阵确定二者的位置信息。由(1)式可知, 两个平面的单应性矩阵可以由同一个点对应的两个图像平面确定。实际车牌平面内的点 U 在两个不同视角下, 得到 I_1, I_1^* 上对应的投影点分别为 $U_1 = [u_1 \ v_1 \ 1]^T$ 和 $U_1^* = [u_1^* \ v_1^* \ 1]^T$, 此时图像平面对 (I_1, I_1^*) 关于车牌平面与单应性矩阵 H 的关系可表示为

$$U_1^* = HU_1 \quad (2)$$

$$\text{设定 } H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}, \text{ 由 } U_1, U_1^* \text{ 可知,}$$

$h_{33}=1$, 则单应性矩阵 H 有 8 个未知量 $h_{11}, h_{12}, h_{13}, h_{21}, h_{23}, h_{23}, h_{31}, h_{32}$ 。根据车牌模板已知的四个顶点位置信息及其对应的图像位置信息 U_1, U_2, U_3, U_4 , 结合(2)式可得到更新后的顶点坐标 $U_1^*, U_2^*, U_3^*, U_4^*$, 从而得到线性方程组

$$\begin{bmatrix} u_1^* \\ v_1^* \\ u_2^* \\ v_2^* \\ u_3^* \\ v_3^* \\ u_4^* \\ v_4^* \end{bmatrix} = \begin{bmatrix} u_1 & v_1 & 1 & 0 & 0 & 0 & -u_1 u_1^* & -u_1^* v_1 \\ 0 & 0 & 0 & u_1 & v_1 & 1 & -u_1 v_1^* & -v_1^* v_1 \\ u_2 & v_2 & 1 & 0 & 0 & 0 & -u_2 u_2^* & -u_2^* v_2 \\ 0 & 0 & 0 & u_2 & v_2 & 1 & -u_2 v_2^* & -v_2^* v_2 \\ u_3 & v_3 & 1 & 0 & 0 & 0 & -u_3 u_3^* & -u_3^* v_3 \\ 0 & 0 & 0 & u_3 & v_3 & 1 & -u_3 v_3^* & -v_3^* v_3 \\ u_4 & v_4 & 1 & 0 & 0 & 0 & -u_4 u_4^* & -u_4^* v_4 \\ 0 & 0 & 0 & u_4 & v_4 & 1 & -u_4 v_4^* & -v_4^* v_4 \end{bmatrix} \begin{bmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32} \end{bmatrix} \quad (3)$$

参照文献[17], 利用数据归一化的直接线性变换(DLT)算法, 求解(3)式中的 8 个未知量, 得到车牌平面的单应性矩阵 H 。根据文献[18]分解单应性矩阵, 求解对应的旋转和平移信息。实际中发现, 构建单应性模型时, 需要精确测定相机标定模板的位姿信息, 且单应性矩阵的计算量大, 难以实时测量位姿。

4.2 简化道路场景几何模型

在几何模型中, 实际拍摄场景的过程与相机标定过程相似, 利用实际拍摄过程中物距(图像的实际深度信息)总是远大于像距的性质, 假设相机拍摄过程中物体的像在相机焦平面上^[19]。基于该假设建立成像过程中的几何模型, 如图 6(a)所示, 其中, A, B 为与地面垂直的物点, G 为地面点, a, b, g 为物点 A, B 以及地面点 G 经透镜作用后在成像平面上对应的成像点, p 为地平线在成像平面上的成像点, Cc 为相机主轴。

将几何模型扩展到高速公路的场景中, 建立路面坐标系, 利用小孔成像模型将车载相机工况进行简化, 得到简化的道路场景模型, 如图 6(b)所示。其中, O 为相机安装位置, 平面 $ABCDU$ 为路平面, Y 轴为汽车前进方向, 梯形 $ABCD$ 在图像中的像为矩形 $abcd$ (尺寸为 $W \times H$), OG 为相机中轴, 与 Y 轴垂直的方向为 X 轴。平面 $ABCDU$ 上的点 $P(X_P, Y_P)$ 在图像中对应像点 $p(x_p, y_p)$ 。

设定的场景参数: 相机的安装高度 $L_{OI} = h$, 俯仰角 $\angle IOG = \gamma_0$, 相机镜头的垂直、水平视场角分别为 $\angle FOE = 2\alpha_0, \angle JOK = 2\beta_0$ 。为建立路平面和像平面上点的对应关系, 对相机中轴垂直面和平行面进行分析。

沿 Y 轴方向, 对 $P(X_P, Y_P)$ 点进行投影, 得到 P_Y 点, 取相机中轴垂直面进行分析, 如图 6(c)所示。在路平面 $ABCDU$ 中, G 为坐标系原点, $\angle GOP_Y = \alpha, \angle GOJ = \alpha_0$, 路面坐标系下 P 点的纵

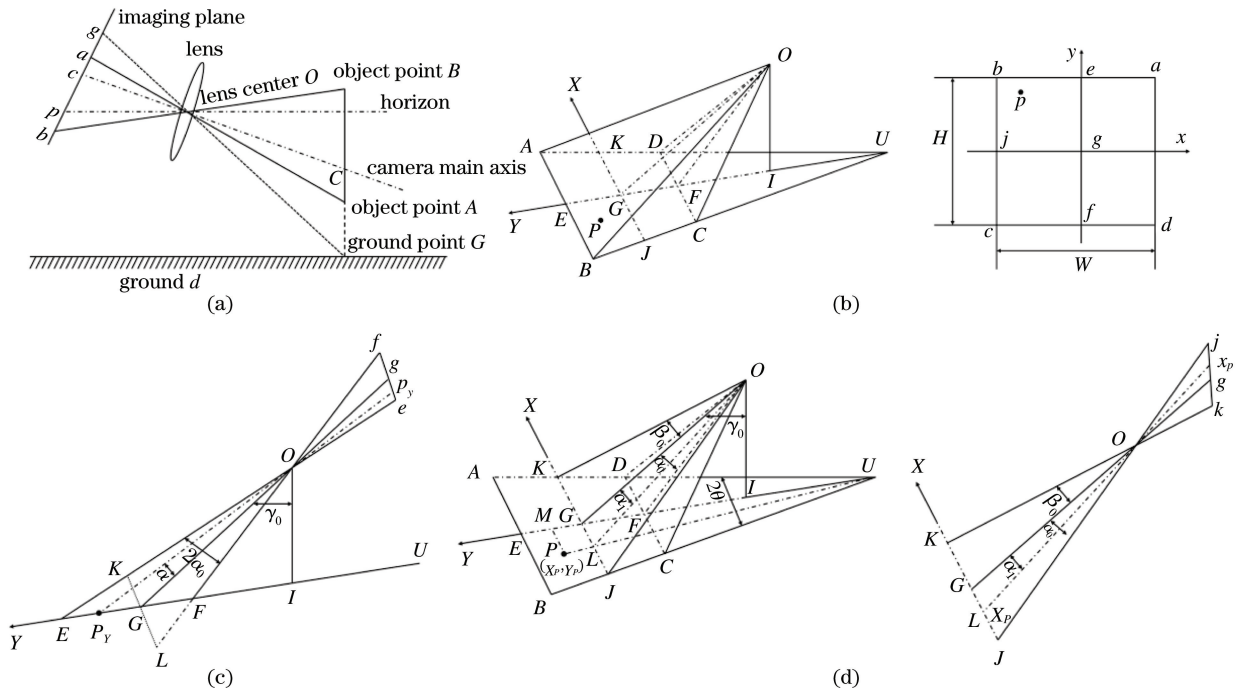


图 6 简化的道路场景几何模型。(a)深度几何模型;(b)车载相机路面模型与成像示意图;(c)垂直方向的成像原理;(d)水平方向的成像原理

Fig. 6 Simplified geometry model of road scene. (a) Depth geometry model; (b) road model of the vehicle-mounted camera and imaging diagram; (c) principle of the vertical imaging ; (d) principle of the horizontal imaging

坐标为 Y_p , $\tan \alpha = \frac{2y_p \cdot \tan \alpha_0}{H}$, 令 $k_1 = \frac{2 \tan \alpha_0}{H}$, $k_2 = \tan \gamma_0$, 得到

$$Y_p = L_{GP_Y} = L_{IP_Y} - L_{IG} = h \cdot k_1 \cdot y_p \left(\frac{1 + k_2^2}{1 - k_2 \cdot k_1 \cdot y_p} \right). \quad (4)$$

同理, 建立相机中轴平行面, 其投影成像原理如图 6(d) 所示。根据垂直视场角和俯仰角的位置, 在垂直面 OGI 中 $L_{GF} = L_{GI} - L_{IF} = h [\tan \gamma_0 - \tan(\gamma_0 - \alpha)]$, (5) 式中, L_{GF}, L_{GI}, L_{IF} 分别为垂直面 OGI 中 GF, GI, IF 的长度。

在 $\triangle OFI, \triangle OGI, \triangle OFC$ 中可以得到 L_{OF}, L_{GI}, L_{FC} ; 在路平面 $ABCDU$ 中, 根据三角形相似原理, 可以得到 L_{UG} ; 在 $\triangle OGL$ 可以得到 L_{LG} 。在图 6(d) 中, 有 $\angle KOG = \beta_0$, 则存在

$$\angle GOL = \alpha_1 = \arctan \left(\frac{2 \cdot x_p \cdot \tan \beta_0}{W} \right), \quad (6)$$

由 $L_{LG}, \angle GOL$ 及 $\triangle UGL \sim \triangle MPU$, 得到

$$X_p = \frac{(L_{UG} + Y_p)}{L_{UG}} \cdot \frac{h}{\cos \gamma_0} \cdot \tan \alpha_1, \quad (7)$$

令 $k_3 = \frac{h}{\cos \gamma_0}, k_4 = \frac{2 \tan \beta_0}{W}$, 可以得到路面和图像

对应点的转换关系, 所求的深度信息可表示为 $L_{IG} + L_p$ 。

4.3 融合车牌信息的 PNP 模型

考虑到实际应用场景中的车牌信息已知, 提出了一种融合车牌信息的 PNP 模型。相比其他行车深度估计方法, 本方法的优点: 1) 车牌图像具有特征明显、易于获取等特点, 便于建立机器学习图像库; 2) 现有 PNP 模型中, 投影学的反向工程不具备唯一确定性, 不能作为单目视觉理论模型。在现行车牌标准下, 车牌固定且具有颜色、物理尺寸已知等特点, 可将精确定位后的车牌尺度信息与 PNP 模型融合进行深度理论求解, 得到车牌与相机之间的深度信息, 进而确定行车间距。

根据文献[20]可知, 在 PNP 模型中, 当 $N = 4$ 时, 如果四个三维点在同一平面且为非光心所在的面, 则 PNP 模型的位姿存在唯一解。将 PNP 模型融合已定位的车牌信息, 对相机的坐标系进行标定, 建立模型并求解。利用空间点和图像点的对应关系, 得到假定的空间物点位置模型如图 7 所示。在平行四边形 $A'B'C'D'$ 中, 以 D' 点为原点 O' 。将 $D'C'$ 作为 x 轴正方向, 建立空间平面坐标系, z 坐标为 0。平行四边形的边长 $L_{A'B'} = Q, L_{A'D'} = T$ 。其中 $\angle A'D'C' = \theta$ 。在 PNP 中求解旋转矩阵 $R =$

$[\mathbf{r}_1 \ \mathbf{r}_2 \ \mathbf{r}_3]^T$ 与平移向量 \mathbf{t} 。已知 $\mathbf{r}_3 = \mathbf{r}_1 \times \mathbf{r}_2$, 摄像机内参标定后, 深度信息的求解转化为对 $\mathbf{r}_1, \mathbf{r}_2, \mathbf{t}$ 以及比例因子 S 的求解, 可表示为

$$S_i \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \mathbf{K}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{t}) \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix}, \quad (8)$$

式中, \mathbf{K} 为相机的内参矩阵, S_i 为第 i 个图像点的比例因子。

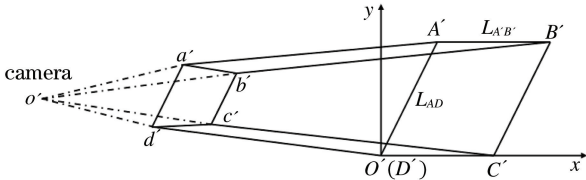


图 7 四边形空间物点坐标系的模型

Fig. 7 Model of the quadrilateral space object point coordinate system

根据平行四边形 $A'B'C'D'$ 四个点建立的世界坐标齐次形式, 依次为 $(Q \cos \theta, Q \sin \theta, 1)^T$, $(T + Q \cos \theta, Q \sin \theta, 1)^T$, $(T, 0, 1)^T$, $(0, 0, 1)^T$, 对应的像素点分别记为 $\mathbf{p}_3, \mathbf{p}_4, \mathbf{p}_2, \mathbf{p}_1$ 。将 $\mathbf{p}_3, \mathbf{p}_4, \mathbf{p}_2, \mathbf{p}_1$ 代入校正后的图像, 标定得到内参矩阵 \mathbf{K}

$$\begin{cases} s_4 \mathbf{p}_4 = T\mathbf{K}\mathbf{r}_1 + Q \cos \theta \mathbf{K}\mathbf{r}_1 + Q \sin \theta \mathbf{K}\mathbf{r}_2 + \mathbf{K}\mathbf{t} \\ s_3 \mathbf{p}_3 = Q \cos \theta \mathbf{K}\mathbf{r}_1 + Q \sin \theta \mathbf{K}\mathbf{r}_2 + \mathbf{K}\mathbf{t} \\ s_2 \mathbf{p}_2 = T\mathbf{K}\mathbf{r}_1 + \mathbf{K}\mathbf{t} \\ s_1 \mathbf{p}_1 = \mathbf{K}\mathbf{t} \end{cases}, \quad (9)$$

式中, $s_i = 1/S_i$, 物点 i 所在位置的深度值 d_i 可表示为

$$\begin{cases} s_1 = d_1 s_4 \\ s_2 = d_2 s_4, \\ s_3 = d_3 s_4 \end{cases}, \quad (10)$$

$$\mathbf{r}_1 = \frac{\mathbf{K}^{-1} s_4}{T} (d_2 \mathbf{p}_2 - d_1 \mathbf{p}_1), \quad (11)$$

$$\mathbf{t} = \mathbf{K}^{-1} s_1 \mathbf{p}_1 = \mathbf{K}^{-1} d_1 s_4 \mathbf{p}_1, \quad (12)$$

根据正交向量的性质 $\mathbf{r}_1^T \mathbf{r}_1 = 1$, 得到

$$s_4 = \frac{T}{|\mathbf{K}^{-1} (d_2 \mathbf{p}_2 - d_1 \mathbf{p}_1)|}, \quad (13)$$

根据旋转矩阵列向量的性质 $\mathbf{r}_1^T \mathbf{r}_2 = 0$, 得到

$$\mathbf{r}_2 = \frac{\mathbf{K}^{-1} s_3 \mathbf{p}_3 - Q \cos \theta \mathbf{r}_1 - \mathbf{t}}{|\mathbf{K}^{-1} s_3 \mathbf{p}_3 - Q \cos \theta \mathbf{r}_1 - \mathbf{t}|}. \quad (14)$$

由 $\mathbf{r}_3 = \mathbf{r}_1 \times \mathbf{r}_2$ 可确定 \mathbf{r}_3 , 最终得到唯一正确的 $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{t}$ 以及比例因子 S 。

5 实验验证与分析

5.1 车牌精确定位结果及分析

在车牌精确定位中, 先将筛选出的车牌图像样本尺寸 $W \times H$ 统一为 $128 \text{ pixel} \times 32 \text{ pixel}$, 然后提取车牌的 HOG 特征 (1620 维), 最后进行归一化处理。将归一化后的 HOG 特征按照 SVM 选参策略提取车牌的位置信息。先使用 PSO 算法进行初步寻参, 寻优过程采用三折交叉验证, 初步得到参数组合: 惩罚因子 $C_p = 1.3204$ 、惩罚参数 $g = 0.01$ 。然后用 GA 优化, 得到最佳参数组合: $C_p = 1.0053, g = 0.004673$ 。通过网格法用两种算法得到的最佳参数范围 (C_p, g) 与提取的 HOG 特征进行精确寻参, 得到精确参数组合: $C_p = 1.16, g = 0.0025$ 。上述训练中的数据如表 1 所示, 在实际检测中用训练得到的参数对 1039 张汽车图像进行检测, 错误识别 7 张, 识别率达到 99.326%。

表 1 SVM 选取的组合特征车牌数据

Table 1 Combination characteristic license plate data selected by SVM

Experiment method	Best parameter set (C_p, g)	Experience risk/%	Confidence risk/%	Structured risk/%	Training time/s
PSO	(1.3204, 0.01)	0	0.4757	0.4757	12364.9101
GA	(1.0053, 0.04673)	0	0	0	4521.0216

5.2 简化道路场景模型验证及分析

本模型采用静态实验方案, 在 Matlab 平台上对算法进行模拟实验。为降低误差, 手动选取角点像素位置。用已经标定的相机对图像进行校正后, 手动选取路平面坐标系中同一条直线上的三个物点, 基本信息如表 2 所示。

表 2 物点的基本信息

Table 2 Basic information of the objects

Object point	True depth information/m	Image coordinates
Object point1	5.5	(1220, 751)
Object point2	6.0	(1260, 753)
Object point3	6.5	(1289, 759)

相机的垂直视场角为 58.1092° , 水平视场角为 150° 。相机的安装高度固定为 150 mm , 俯仰角约为 80° 。考虑到测量误差, 设定俯仰角的变化范围为 $79^\circ \pm 0.1^\circ$, 计算出距离的大小。当固定俯仰角为 78.4° 时, 安装高度的范围为 $(150 \pm 10)\text{ mm}$, 计算出距离的大小。通过静态实验分析两种情况的简化道路场景模型时发现, 当相机安装高度的变化量为 20 mm 时, 实际测量结果的最大变化量为 1000 mm ; 当俯仰角的变化量为 0.2° 时, 实际测量结果的最大变化量为 3000 mm 。原因是实际安装中, 难以保证相机高度及俯仰角的安装精度。

5.3 融入车牌信息 PNP 模型的验证及分析

为验证本方法的准确性与可靠性, 根据标定的相机内参对成像过程进行模拟, 具体步骤如下。

1) 随机产生 OA 、 OC 及其夹角, 得到平行四边形 $ABCD$ 的四个顶点; 设 D 点为世界坐标系原点, 四个点的齐次坐标分别为 P_1 、 P_2 、 P_3 、 P_4 。

2) 参照相机成像模型及内参标定结果, 随机产生旋转矩阵 $R = [r_1 \ r_2 \ r_3]^T$ 和平移向量 t , 模拟四边形与相机之间的相对位置关系; 并对 P_1 、 P_2 、 P_3 、 P_4 进行模拟相机成像, 得到图像坐标点 p_1 、 p_2 、 p_3 、 p_4 。

3) 将图像坐标点带入上述模型中, 求解平移向量 t 与旋转向量 r_1 、 r_2 、 r_3 。

4) 将计算的矩阵与步骤 2) 中随机产生的结果进行对比。

实验过程中, 产生一组旋转矩阵 R 和平移向量 t , 像素点加上随机噪声后, 对每组点进行 10 次实验, 取其均值。随机选取 18 组数据, 将模型计算得到的平移向量 t 与步骤 2) 中随机产生的平移向量进行对比。实验结果表明, 最大误差为 0.7% , 这表明在噪声影响下, 本方法的计算结果比较精确。

利用融合已定位车牌的 PNP 模型, 在 VS2013 环境下进行静态图像实验。实验中手动标注车牌的

四个顶点, 利用 Eigen3.3.3 库校正图像后, 带入模型计算目标物体的深度信息, 在可视化界面上提取距离信息。在 VS2013 平台上用 C 语言进行开发实验, 用行车记录仪采集的视频进行验证。实验环境: 相机的安装高度为 680 mm , 相机的俯视角为 90° 。软件运行环境为 Intel(R) Core(TM) i3-7100 CPU@3.9 GHz; 软件为 VS2013+Opencv3.0+Eigen3.3.3。

以视野内的车牌为识别目标, 设定测量距离为 $750, 1500, 2250, 3000, 3750, 4500, 5250, 6000, 6750, 7500, 8250, 9000, 9750, 10500, 11250, 12000, 12750, 13500, 14250, 15000, 15750\text{ mm}$ 。每种距离下拍摄 10 张静态图像, 最终测量结果的平均值如图 8 所示。可以发现, 实验过程中, 当目标处在 10 m 范围内时, 误差可控制在 $2.1\% \sim 10\%$ 。随着目标点距离的增加, 误差也会随之加大。原因主要是相机像素质量低, 当目标距离较远时, 图像中汽车尾部像素面积较小, 难以识别车牌进行准确定位。可通过提高相机像素或实现目标物体的精确定位进行改善。

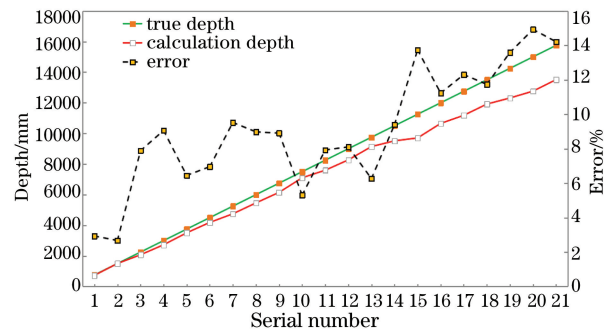


图 8 深度测量结果

Fig. 8 Depth measurement result

将本方法应用在前方车辆实时测距系统上, 在视频序列上进行测试, 汽车行驶路段为城市道路。用本方法进行前方车辆的自动定位, 融合车牌的几何尺寸进行前方车距检测, 效果如图 9 所示。实验中处理一张图像所需的时间约为 170 ms , 满足实时性的要求, 且能进行安全距离报警。

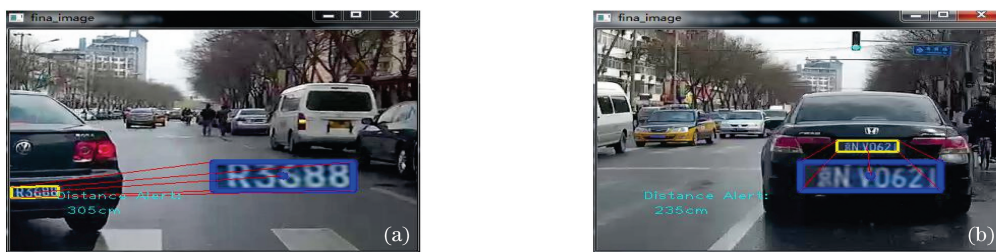


图 9 本方法在视频序列上的应用效果图。(a) 视频序列效果图 1; (b) 视频序列效果图 2

Fig. 9 Application effect diagram of our method on video sequence. (a) Video sequence effect chart1; (b) video sequence effect chart2

6 结 论

提出了一种基于 PNP 模型的实时单目深度估计方法。首先,在现有方法的基础上引入了浅层机器学习,综合车牌的颜色和轮廓信息,对车牌进行初步定位;然后,基于 SVM 对车牌进行精确定位,并将精确定位的车牌信息融入 PNP 模型中;最后,在静态图像与视频序列上进行测试,均得到了较好的测距效果。理论分析及实验结果表明,本方法可以有效解决车辆行驶过程中行车间距的实时检测问题,提高了测距的精度以及单目视觉测距的适应性,且实时性较好,满足车载行车间距检测的实际应用需求,可为行车图像测距和智能车辆实现无人驾驶提供一定的理论基础。

参 考 文 献

- [1] Zhang Y L. World Health Organization released the report on global road safety in 2018 [J]. Chinese Journal of Disaster Medicine, 2019, 7(2): 100.
张亚丽. 世界卫生组织发布《2018 年全球道路安全现状报告》[J]. 中华灾害救援医学, 2019, 7(2): 100.
- [2] Xiong Y C, Li X O. Review on obstacle avoidance methods based on ranging technology[J]. Progress in Biomedical Engineering, 2018, 39(4): 204-208.
熊毅超, 李晓欧. 基于测距技术的避障方法综述[J]. 生物医学工程学进展, 2018, 39(4): 204-208.
- [3] Ding M, Jiang X Y. Scene depth estimation based on monocular vision in advanced driving assistance system[J]. Acta Optica Sinica, 2020, 40(17): 1715001.
丁萌, 姜欣言. 先进驾驶辅助系统中基于单目视觉的场景深度估计方法 [J]. 光学学报, 2020, 40(17): 1715001.
- [4] Li Y, Chen X W, Wang Y, et al. Progress in deep learning based monocular image depth estimation[J]. Laser & Optoelectronics Progress, 2019, 56(19): 190001.
李阳, 陈秀万, 王媛, 等. 基于深度学习的单目图像深度估计的研究进展 [J]. 激光与光电子学进展, 2019, 56(19): 190001.
- [5] Zhang F Q, Lu L P, Shen M D, et al. Measurement method to object distances by monocular vision [J]. Acta Photonica Sinica, 2009, 38(2): 453-456.
张法全, 路立平, 沈满德, 等. 单目视觉目标距离测量方法研究 [J]. 光子学报, 2009, 38(2): 453-456.
- [6] Xu C, Gao M, Cao H. The attitude angle estimation-based distance measurement of tank target in monocular image [J]. Acta Photonica Sinica, 2015, 44(5): 140-147.
徐超, 高敏, 曹欢. 单目图像中姿态角估计的坦克目标测距方法 [J]. 光子学报, 2015, 44(5): 140-147.
- [7] Wang J, Wang Y Q. A monocular stereo vision algorithm based on bifocal imaging [J]. Robot, 2007, 29(1): 41-44, 50.
王剑, 王元庆. 基于双焦成像的单眼立体视觉算法 [J]. 机器人, 2007, 29(1): 41-44, 50.
- [8] Huang F S, Liu E F, Fang Y X, et al. Single camera stereo vision recognition for parts pose based on intelligent three coordinate measuring machine [J]. Optics and Precision Engineering, 2013, 21(5): 1326-1332.
黄风山, 刘恩福, 方忆湘, 等. 基于智能三坐标测量机的零件位姿单目立体视觉识别 [J]. 光学精密工程, 2013, 21(5): 1326-1332.
- [9] Liu F Y, Shen C H, Lin G S, et al. Learning depth from single monocular images using deep convolutional neural fields [EB/OL]. [2020-07-05]. <https://arxiv.org/abs/1502.07411>.
- [10] Cao Y, Wu Z F, Shen C H. Estimating depth from monocular images as classification using deep fully convolutional residual networks [EB/OL]. [2020-07-02]. <http://arxiv.org/abs/1605.02305>.
- [11] Yuan J Z, Zhou W J, Pan T, et al. Road scene depth estimation based on deep convolutional neural networks [J]. Laser & Optoelectronics Progress, 2019, 56(8): 081501.
袁建中, 周武杰, 潘婷, 等. 基于深度卷积神经网络的道路场景深度估计 [J]. 激光与光电子学进展, 2019, 56(8): 081501.
- [12] Ma S D, Zhang Z Y. Computer vision: computing theory and algorithmic basis [M]. Beijing: Science Press, 1998.
马颂德, 张正友. 计算机视觉: 计算理论与算法基础 [M]. 北京: 科学出版社, 1998.
- [13] Guo K Y, Guo X L, Wang Y W. Multi-attribute cascading identify variable contour moving targets [J]. Computer Measurement & Control, 2016, 24(3): 263-266.
郭克友, 郭晓丽, 王艺伟. 多特征级联式检测变轮廓运动目标 [J]. 计算机测量与控制, 2016, 24(3): 263-266.
- [14] Guo J X, Chen W. Face recognition based on HOG multi-feature fusion and random forest [J]. Computer Science, 2013, 40(10): 279-282, 317.
郭金鑫, 陈玮. 基于 HOG 多特征融合与随机森林的

- 人脸识别 [J]. 计算机科学, 2013, 40(10): 279-282, 317.
- [15] Dalal N, Triggs B. Histograms of oriented gradients for human detection [C]//2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), June 20-25, 2005, San Diego, CA, USA. New York: IEEE Press, 2005: 886-893.
- [16] Guo K Y, Guo X L, Wang Y W. Analysis and strategy for parameter optimization methods of SVM [J]. Computer Measurement & Control, 2016, 24(6): 255-259.
郭克友, 郭晓丽, 王艺伟. SVM 参数优化方法分析与决策 [J]. 计算机测量与控制, 2016, 24(6): 255-259.
- [17] Liu S, Ye Y T, Sun Q, et al. Research of chessboard's corner detection based on homography[J]. Electronic Design Engineering, 2012, 20(6): 114-117.
刘嵩, 叶玉堂, 孙强, 等. 基于单应性矩阵的棋盘格角点检测研究 [J]. 电子设计工程, 2012, 20(6): 114-117.
- [18] Faugeras O D, Lustman F. Motion and structure from motion in a piecewise planar environment [J]. International Journal of Pattern Recognition and Artificial Intelligence, 1988, 2(3): 485-508.
- [19] Li L, Zhang M J, Xiong Z H, et al. Depth estimation from a single still image of street scene based on content understanding[J]. Robot, 2011, 33(2): 174-180.
李乐, 张茂军, 熊志辉, 等. 基于内容理解的单幅静态街景图像深度估计 [J]. 机器人, 2011, 33(2): 174-180.
- [20] Fakhir M M, Woo W L, Chambers J A, et al. Perspective projection for variance pose face recognition from camera calibration [J]. Proceedings of SPIE, 2016, 9896: 98961L.