

基于 SIFT 和 Hu 特征融合的单目视觉识别算法研究*

李中生¹, 赵彬², 刘振宇³, 赵雪⁴, 田立敏¹

- (1.空军驻沈阳地区军事代表室, 辽宁 沈阳 110016;
2.沈阳新松机器人自动化股份有限公司, 辽宁 沈阳 110168;
3.沈阳工业大学 信息科学与工程学院, 辽宁 沈阳 110870;
4.沈阳防锈包装材料有限公司, 辽宁 沈阳 110084)

摘要: 研究了机器视觉技术在三维物体识别定位问题中的应用。利用 Visual C++ 编程软件建立了摄像机标定界面, 实现对摄像机进行快速标定的功能。提出了一种 SIFT 特征和 Hu 不变矩融合算法, 该算法是一种融合了局部特征和全局特征的算法。其中全局特征反映了对三维物体图像的整体信息进行粗略匹配和定位, 局部特征可以在全局特征中进行更准确特征匹配, 该算法对伸缩、旋转和平移等有很好的抵抗能力。实验结果表明, 该视觉算法可以有效解决三维物体匹配问题, 并且有效提高了系统的识别速率和效率, 满足物体识别的目的。

关键词: 机器视觉; 三维物体识别; 模式识别; 图像处理

中图分类号: TP391

文献标识码: A

文章编号: 1674-7720(2013)13-0041-05

Research of monocular vision recognition algorithm based on SIFT and Hu feature fusion

Li Zhongsheng¹, Zhao Bin², Liu Zhenyu³, Zhao Xue⁴, Tian Limin¹

- (1. Airforce Military Representative Office Resident in Shenyang Region, Shenyang 110016, China;
2. Siasun Robot & Automation Co., Ltd., Shenyang 110168, China;
3. Information Science and Engineering School, Shenyang University of Technology, Shenyang 110870, China;
4. Shenyang Rustproof Packaging Material Co., Ltd., Shenyang 110084, China)

Abstract: This paper researched machine vision technology on the application of the 3D object recognition and positioning problem. Visual C++ programming software was set up for camera calibration interface, which realized function that camera was quickly calibrated. This paper put forward a kind of SIFT features and Hu invariant moment fusion algorithm, which was a kind of combined with local characteristics and global characteristics. The global features roughly matched and fixed position to reflect the overall information of 3D object image, and local characteristics accurately matched feature based on global characteristics. The algorithm has a good resistant ability for expansion, rotation and translation. The experimental results show that visual algorithm can effectively solve the matching problem of 3D objects and efficiency improves of system identification rate and efficiency to achieve the purpose of object recognition.

Key words: machine vision; 3D object recognition; pattern match; image processing

在工业自动化的大背景下, 机器视觉检测技术受到越来越多的重视, 其在识别和定位目标有着较好的应用前景^[1]。一般来说, 目标识别定位大多依赖于图像匹配

* 基金项目: 国家高技术研究发展 863 计划 (2012AA041405); 沈阳市工业科技攻关项目 (F12-010-2-00)

技术。按照匹配基元的不同, 图像匹配方法主要分为区域匹配、相位匹配、特征匹配, 其中特征匹配因不直接依赖于灰度而得到广泛应用。

目前, 基于机器视觉技术识别算法研究已经从最初

的实验室逐渐走向实际应用阶段。例如,Zhu Junchao、Zhou Jie 和 Zhang Baofeng 研究了相关视觉处理算法,提出了一种区分行人和车辆的识别算法^[2];Wen Ying、Lu Yue 和 Yan Jingqi 等人提出了一种提出了一个新颖的阴影去除技术和字符识别算法,该算法应用于智能交通系统的车牌识别^[3];Xie F、Xu G、Cheng Y 和 TIAN Y 等人提出了一种改进型细化算法,该算法应用于人体姿势识别系统^[4];NOOR A、MANIVANAN N 和 BALACHANDRAN W 介绍了一种对指纹模板形成和匹配的自动识别算法,该算法保持了很高的精度,误误差率不到 3.5%^[5]。

本文提出一种基于尺度不变特征转换(SIFT)和 Hu 特征融合的单目视觉识别算法。首先通过单目摄像机将三维目标采集到计算机内,然后采用 Hu 不变矩提取全局特征进行粗略识别定位,最后采用 SIFT 算法进行更准确的局部特征匹配。该融合算法不仅具有尺度、旋转、仿射、视角、光照不变性,对目标的运动、遮挡、噪声等因素影响也能保持较好的匹配效果,而且有效提高了 SIFT 的算法速率。

1 算法总体设计

算法流程是三维目标识别系统的核心,决定该系统的工作方式。基于 SIFT 和 Hu 特征融合的单目视觉识别系统的图像处理流程算法流程如图 1 所示,其步骤如下:

(1)单目摄像机标定:摄像机标定的目的就是建立图像坐标系与世界坐标系之间的关系,用于对识别出的三维目标进行后续的抓取处理。

(2)图像采集:通过单目摄像机采集三维物体的图像序列,并通过以太网传入到计算机内,待视觉软件进行模式识别。

(3)Hu 不变矩:其中全局特征反映了三维物体图像的整体信息,用于进行粗略的特征匹配和定位。

(4)SIFT 特征提取:首先对原始图像进行高斯滤波,建立高斯金字塔,获取高斯差分来找到感兴趣的关键点;然后经过平稳筛选获得真正关键点;最后根据每一个关键点邻域,为关键点分配梯度的模和方向。

(5)SIFT 特征匹配:局部特征可以在全局特征中更准确地进行特征匹配,采用优先 K-D 树近似搜索算法进行搜索来查找每个特征点的最近邻和次近邻特征点。计算特征点到最近邻距离与次近邻距离的比值,若小于阈值 T ,则该特征点与最近邻特征点匹配成功。利用 RANSAC 随机抽样一致性算法去除错误匹配,提高鲁棒性。

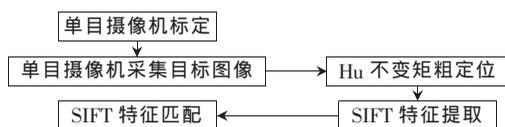


图 1 三维目标识别流程

2 摄像机标定

摄像机标定是三维视觉检测领域中的基本问题,作

为信息获取单元的单目摄像机,其标定参数的精度将直接影响到测量结果,因而是视觉测量中的关键和前提^[6]。一般来说,对于摄像机的标定建立摄像机成像的几何模型和确定模型的几何参数分为两步。在摄像机标定中,大多使用张正友标定法,张正友标定法使用棋盘格为标定模板,需要利用摄像机以不同视点角度拍摄 3 幅模板图像,利用图像特征点标定出摄像机参数^[7]。图 2 为摄像机标定的角点识别示意图,根据图像点之间对应关系即可标定出摄像机内参数与外参数。针孔模型用来解释真实场景到投影平面上的成像。



图 2 摄像机标定的角点识别示意图

有了这两种坐标系,就可以对应地在真实物理尺寸和图像像素值之间进行坐标转换,关系为:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} dx & 0 & u_0 \\ 0 & dy & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (1)$$

世界坐标系 (X_w, Y_w, Z_w) 就是现实的三维坐标系,它是在空间内一点建立起来的直角三维坐标系。它与摄像机坐标系的关系为:

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} R_{3 \times 3} & T_{3 \times 1} \\ 0^T & 0 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (2)$$

其中, $R_{3 \times 3} = (R_x, R_y, R_z)$ 为旋转矩阵; $T_{3 \times 1} = (R_x, R_y, R_z)T$ 为平移向量。

将摄像机坐标系 (O_c, X_c, Y_c, Z_c) 在针孔摄像机模型中透视投影,得到理想的成像坐标系中物理坐标 (X_u, Y_u) 。

$$\begin{cases} X_u = f \frac{X_c}{Z_c} \\ Y_u = f \frac{Y_c}{Z_c} \end{cases} \quad (3)$$

其中, f 为摄像机的焦距。

理想成像平面坐标 (X_u, Y_u) 与实际的成像坐标系 (X_d, Y_d) 的转换关系为:

$$\begin{bmatrix} X_d \\ Y_d \end{bmatrix} = (1+k_1 r^2 + k_2 r^4) \begin{bmatrix} X_u \\ Y_u \end{bmatrix} + \begin{bmatrix} 2p_1 X_u Y_u + p_2 (r^2 + 2X_u^2) \\ p_1 (r^2 + 2Y_u^2) + 2p_2 X_u Y_u \end{bmatrix} \quad (4)$$

$$r = X_u^2 + Y_u^2$$

其中, k_1, k_2 为径向畸变系数, p_1, p_2 为切向畸变系数。

实际的成像平面坐标 (O_d, X_d, Y_d) 与计算机图像坐

标 (O, u, v) 的线性转换为:

$$\begin{cases} u=f_x X_d+c_x \\ v=f_y Y_d+c_y \end{cases} \quad (5)$$

$$\begin{cases} f_x=f/\Delta x \\ f_y=f/\Delta y \end{cases} \quad (6)$$

其中, (c_x, c_y) 是基准点, Δx (Δy) 为水平 (垂直) 方向上相邻像素之间的距离 (单位为 mm/pixel), f_x 和 f_y 是以像素为单位的焦距。

根据式 (2)~式 (6), 得到:

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} [R_{3 \times 3} \quad T_{3 \times 1}] \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (7)$$

根据式 (7) 可以得到三维世界坐标系与二维坐标系之间的一种对应关系。

3 Hu 不变矩

矩的定义源于概率理论, 是一种重要的数字特征, 同时也具有直观的物体意义^[8]。Hu 不变矩是 1962 年 Hu 最早提出的二维不变矩理论, 是提取图像平移、大小、旋转情况等形态特征的方法^[9]。图像匹配实质是根据已知图像特征从目标待测图像中寻找相同图像特征的过程^[10]。本文使用 Hu 不变矩来提取三维物体的特征信息。

对于一幅大小为 $M \times N$ 的图像 $I(i, j)$, 其 $(p+q)$ 阶矩为:

$$M_{pq} = \sum_{i=1}^M \sum_{j=1}^N i^p j^q I(i, j) \quad (p, q=0, 1, 2, \dots) \quad (8)$$

为了保证在图像平移、缩放、旋转时保持不变性, 定义中心矩为:

$$u_{pq} = \sum_{i=1}^M \sum_{j=1}^N (i-\bar{i})^p (j-\bar{j})^q I(i, j) \quad (p, q=0, 1, 2, \dots) \quad (9)$$

用一阶矩 M_{10} 和 M_{01} 分别除以零阶矩 M_{00} , 得到的是物体的质心坐标:

$$(\bar{i}, \bar{j}) = (M_{10}/M_{00}, M_{01}/M_{00}) \quad (10)$$

对于上述中心距, 可得到平移与缩放不变矩, 定义为:

$$v_{pq} = \frac{u_{pq}}{u_{00}^{(p+q)/2+1}} \quad (11)$$

利用式 (11), 可得图像的 Hu 不变矩为:

$$\varphi_1 = v_{20} + v_{02} \quad (12)$$

$$\varphi_2 = (v_{20} - v_{02})^2 + 4v_{11}^2 \quad (13)$$

$$\varphi_3 = (v_{30} - 3v_{12})^2 + (3v_{21} - v_{03})^2 \quad (14)$$

$$\varphi_4 = (v_{30} + v_{12})^2 + (v_{03} + v_{21})^2 \quad (15)$$

$$\varphi_5 = (v_{30} - 3v_{12})(v_{30} + v_{12})[(v_{30} + v_{12})^2 - 3(v_{03} + v_{21})^2] + (3v_{21} - v_{03})(v_{21} + v_{03})[3(v_{30} + v_{12})^2 - (v_{03} + v_{21})^2] \quad (16)$$

$$\varphi_6 = (v_{20} - v_{02})[(v_{30} + v_{12})^2 - (v_{03} + v_{21})^2] + 4v_{11}(v_{30} + v_{12})(v_{03} + v_{21}) \quad (17)$$

$$\varphi_7 = (3v_{21} - v_{03})(v_{30} + v_{12})[(v_{30} + v_{12})^2 - 3(v_{03} + v_{21})^2] + (3v_{21} - v_{03})(v_{21} + v_{03})[3(v_{30} + v_{12})^2 - (v_{03} + v_{21})^2] \quad (18)$$

这 7 个不变矩称为 Hu 不变矩的 7 阶不变矩, 在图

像平移、约束缩放、旋转下具有不变性。该全局特征反映了三维物体图像的整体信息, 其中还包括目标物体大致的位置信息, SIFT 就可以根据三维目标的大致信息进行局部的特征提取匹配, 有效加快了 SIFT 识别算法。

4 SIFT 目标识别

虽然全局特征反映了三维物体图像的整体信息, 其中还包括三维目标的位置信息, 但是还需要在全局特征中更准确地进行特征匹配。SIFT 从尺度不变关键点中获取图像的特征^[11-13]。SIFT 算法是一种提取局部特征的算法, 在尺度空间寻找极值点, 提取位置、尺度、旋转不变量^[14-15]。本文在 SIFT 算法基础上实现目标识别过程, 并在搜索策略和剔除错误匹配上进行了改进, 大大提高了实时性和鲁棒性。SIFT 特征匹配算法主要包括两个阶段: SIFT 特征的生成以及 SIFT 特征向量的匹配。

4.1 特征提取

SIFT 可以根据三维目标的大致信息进行局部的特征提取匹配, 有效提高了 SIFT 识别算法的时间。SIFT 特征的生成一般包括以下几个步骤:

(1) 构建尺度空间, 检测极值点, 获得尺度不变性。高斯卷积核是实现尺度变换的唯一变换核, 所以一幅二维图形尺度空间定义为:

$$\begin{cases} G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x+y)^2}{2\sigma^2}} \\ L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \end{cases} \quad (19)$$

其中, $I(x, y)$ 为二维图像, $G(x, y, \sigma)$ 为二维高斯函数, (x, y) 为空间坐标。σ 是尺度空间因子, 值越小表示图像被平滑得越少, 相应的尺度也就越小。大尺度对应于图像的概率特征, 小尺度对应于图像的细节特征。为了检测尺度空间的稳定的关键点, 采用不同尺度的高斯差分核与图像卷积来检测那些在局部位置的极值点。LoG 算子的尺度归一化近似 DoG 算子为:

$$\begin{aligned} D(x, y, \sigma) &= [G(x, y, k\sigma) - G(x, y, \sigma)] * I(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma) \end{aligned} \quad (20)$$

DoG 算子与尺度归一化的高斯拉普拉斯函数非常近似。构造 $D(x, y, \sigma)$ 可以采用高斯金字塔的方法。图 3 所示为构造的不同尺度的高斯图像。

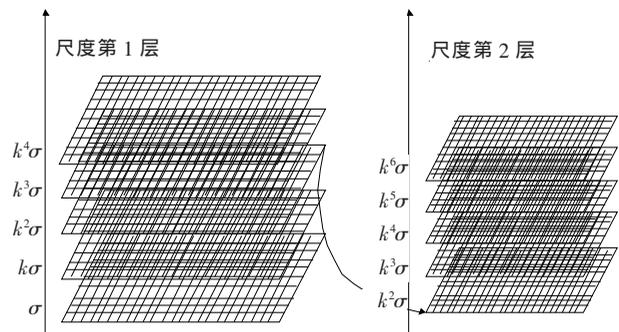


图 3 不同层尺度的高斯图像

首先采用不同尺度因子的高斯核对图像进行卷积,以得到图像的不同尺度空间,将这一组图像作为金字塔图像的第1层;接着对第1层图像进行2倍尺度采样来得到金字塔图像的第2层中的第一幅图像,对该图像采用不同尺度因子的高斯核进行卷积,以获得金字塔图像中第2层的一组图像;再以金字塔图像中第2层中的图像进行2倍尺度采样来得到金字塔图像的第3层中的第一幅图像,对该图像采用不同尺度因子的高斯核进行卷积,以获得金字塔图像中第3层的一组图像。依次类推,从而获得了图3所示的高斯图像。

将图3所示的不同尺度的高斯图像每一层相邻的高斯图像相减,就得到了图4所示的高斯差分图像。因为高斯差分函数是归一化的高斯拉普拉斯函数的近似,所以可以从高斯差分结构中提取极值点作为候选特征点。对DoG尺度空间每个点与相邻尺度和相邻位置的点逐个进行比较,得到的局部极值位置即为特征点所处的位置和对应的尺度。

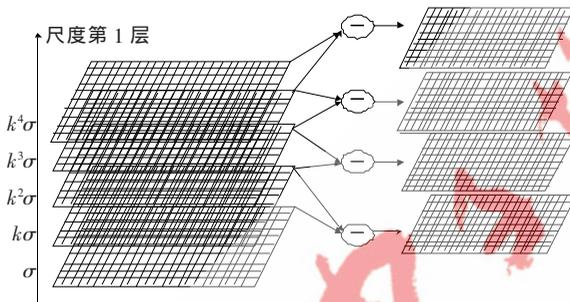


图4 高斯差分图像

(2)极值点的精确定位,删除边缘特征点得到关键点,并且确定关键点的位置和所处的尺度。如图5所示,计算图像相关点在每一尺度下DoG的响应值,与同尺度的8个相邻点和上下相邻尺度对应的 9×2 个点共26个点比较,如果该点值最大或者最小,则该点将作为一个局部极值点,记下它的位置和对应尺度。

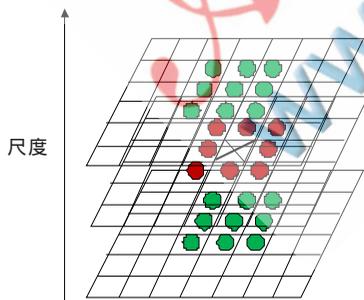


图5 DoG尺度空间局部极值检测

利用Hessian矩阵 H 过滤掉边缘处的低对比度特征点,也就是通过计算高斯差分算子的主曲率来滤除不稳定的边缘响应点,从而确立特征点,公式如下:

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix} \quad (21)$$

矩阵 H 的特征值与 D 的主曲率是成正比例的,假设 α 是矩阵 H 较大的特征值, β 是矩阵 H 较小的特征值,可以通过式(21)和式(22)计算出 α 、 β 的比例 γ 。

$$\text{Tr}(H) = D_{xx} + D_{yy} = \alpha + \beta \quad (22)$$

$$\text{Det}(H) = D_{xx}D_{yy} - (D_{xy})^2 = \alpha\beta \quad (23)$$

其中, $\text{Tr}(H)$ 表示矩阵 H 的迹, $\text{Det}(H)$ 表示矩阵 H 行列式的值矩阵。令 $\alpha = \gamma\beta$,则有:

$$\frac{\text{Tr}(H)^2}{\text{Det}(H)} = \frac{(\alpha + \beta)^2}{\alpha\beta} = \frac{(\gamma\beta + \beta)^2}{\gamma\beta^2} = \frac{(\gamma + 1)^2}{\gamma} \quad (24)$$

对 γ 取一阈值,当 γ 大于阈值时即作为边缘特征点删除。

(3)为特征点方向分配方向值,使用极值点邻域梯度的主方向作为关键点的方向特征,以实现算子对尺度和方向的无关性。利用特征点邻域像素的梯度方向分布特性为每个关键点指定方向参数,使算子具备旋转不变性。利用关键点邻域像素的梯度及方向分布的特性,可以得到梯度模值和方向为:

$$m(x, y) = \sqrt{[L(x+1, y) - L(x-1, y)]^2 + [L(x, y+1) - L(x, y-1)]^2} \quad (25)$$

$$\theta(x, y) = \tan^{-1} \frac{L(x, y+1) - L(x, y-1)}{L(x+1, y) - L(x-1, y)} \quad (26)$$

在以关键点为中心的邻域窗口内采样,并用直方图统计邻域像素的梯度方向。图6所示为8个方向的直方图计算结果。梯度直方图的范围是 $0 \sim 360^\circ$,其中每 10° 一个方向,总共36个方向。

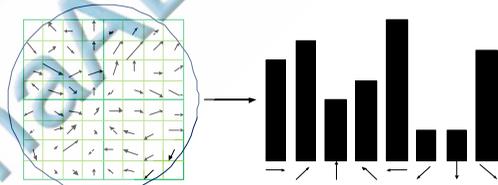
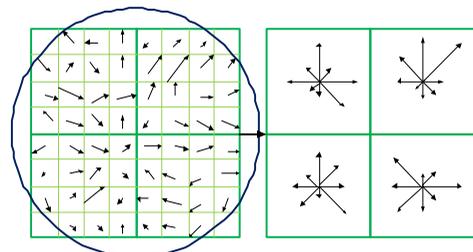


图6 8个方向的直方图计算结果

至此,图像的关键点已检测完毕,每个关键点有3个信息,分别为位置、尺度和方向。

(4)关键点描述子生成。图7所示为由关键点邻域梯度信息生成特征向量。每个特征点使用44邻域内共16个种子点来描述,这样对于一个关键点就可以产生128个数据,最终形成128维的SIFT特征向量。此时SIFT特征向量已经去除了尺度变化、旋转等几何变形因素的影响,再将特征向量的长度归一化,则可以进一步去除光照变化的影响,从而精确描述特征点。



(a)邻域梯度方向 (b)关键点特征向量

图7 由关键点邻域梯度信息生成特征向量

4.2 图像匹配

本文采用最近邻 NN(Nearest Neighbor)算法配准,即采用样本特征点的最近邻特征点距离与次近邻特征点距离的比值来对特征点进行匹配。最近邻特征点是指目标图像中与模板图像特征点具有最短欧几里德距离的特征点。次近邻特征点是指具有比最近邻距离稍长的欧几里德距离的特征点。用最近邻与次近邻比值来进行特征点的匹配可以取得很好的效果。其内容如下:

(1)采用欧式距离作为特征点间的相似性度量。

(2)采用优先 K-D 树近似 BBF(Best-Bin-First)搜索算法行搜索来查找每个特征点的最近邻和次近邻特征点。K-D 树搜索算法是二叉检索树的扩展,K-D 树的每一层将空间分成两个,树的顶点节点按一维进行划分,下一层节点按照另一维进行划分,以此类推各个维循环往复。K-D 树的实质就是一种数据结构,用于搜索高维最邻近点,同时也是一种二叉树,每个节点是一个高维向量。

(3)计算特征点到最近邻距离与次近邻距离的比值,若小于阈值 T ,则该特征点与最近邻特征点匹配成功。

(4)利用 RANSAC 随机抽样一致性算法去除错误匹配,提高鲁棒性。

本文详细介绍了单目摄像机的标定过程,并且阐述了机器视觉技术中的 SIFT 和 Hu 特征融合算法在三维物体识别领域中的应用,该算法可以准确地完成匹配任务。同时,使用 OpenCV 视觉算法库在 VC++ 环境下完成了仿真软件的编制,图 8 和图 9 是基于 SIFT 和 Hu 特征融合的单目视觉识别牛奶盒的效果图。



图 8 牛奶盒匹配的效果图 1

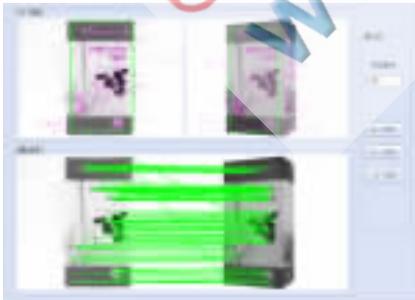


图 9 牛奶盒匹配的效果图 2

仿真结果表明,相比于单纯利用 SIFT 算法进行三维目标匹配,本文所论述的算法准确、快速,能够有效地进行目标的识别和粗定位。

参考文献

- [1] 赵宇峰,高超,王建国.基于机器视觉的工业产品表面缺陷检测算法研究[J].计算机应用与软件,2012,29(2):152-154.
- [2] Zhu Junchao, Zhou Jie, Zhang Baofeng. Research on moving human and vehicle's recognition algorithm [C]. Qinhuangdao: ICCDA,2010.
- [3] Wen Ying, Lu Yue, Yan Jingqi, et al. An algorithm for license plate recognition applied to intelligent transportation system [J]. IEEE Intelligent Transportation Systems Society, 2011, 13(3): 830-845.
- [4] Xie F, Xu G, Cheng Y, et al. Human body and posture recognition system based on an improved thinning algorithm[J]. Institution of Engineering and Technology,2011, 5(5): 420-428.
- [5] NOOR A, MANIVANAN N, BALACHANDRAN W. Transformation invariant algorithm for automatic fingerprint recognition [J]. Institution of Engineering and Technology, 2012, 48(14): 834-835.
- [6] 陈天飞,马孜,李鹏,等.一种基于非量测畸变校正的摄像机标定方法[J].控制与决策,2012,27(2):243-251.
- [7] 陈爱华,高诚辉,何炳蔚.基于正交消失点对的摄像机标定方法[J].仪器仪表学报,2012,33(1):161-166.
- [8] 张伟,何金国.Hu 不变矩的构造与推广[J].计算机应用,2010,30(9):2449-2452.
- [9] 田秋红,孙政荣.基于 Hu 不变矩和 BP 网络的条形码图像识别方法[J].计算机工程与设计,2012,33(4):1563-1568.
- [10] 王洪涛,丁国清.基于不变矩图像匹配的工件种类判别应用研究[J].制造业自动化,2012,34(1):141-144.
- [11] 熊英,马惠敏.3 维物体 SIFT 特征的提取与应用[J].中国图象图形学报,2010,15(5):814-819.
- [12] 刘立,万亚平,刘朝晖,等.基于 SIFT 匹配算法的移动机器人单目视觉定位研究[J].系统仿真学报,2012,24(9):1-4.
- [13] 朱飞,王兴起.基于 SIFT 算法的体育类图像分类与应用研究[J].计算机应用与软件,2011,28(10):231-234.
- [14] 董蓉,李勃,陈启美.基于 SIFT 特征的目标多自由度 mean-shift 跟踪算法[J].控制与决策,2012,27(3):399-407.
- [15] 息朝健,郭三学.基于简化 Forstner 算子改进的 SIFT 无人机图像识别算法[J].计算机应用与软件,2012,29(5):254-300.

(收稿日期:2013-04-16)

作者简介:

李中生,男,1987年生,硕士,主要研究方向:视觉伺服信息处理。

赵彬,男,1987年生,软件工程师,主要研究方向:视觉伺服信息处理。