

文章编号:1000-582X(2012)01-059-06

应用人脸姿态估计实现人机信息交互

毛玉星,于燕来,韩倍涛,徐少志,成华安

(重庆大学 输配电装备及系统安全与新技术国家重点实验室,重庆 400044)

摘要:针对残疾人士不用手实现计算机操作的问题,提出了一种通过面部姿态的视觉分析手段控制鼠标操作的新方法。首先,由 CCD 摄像头获得人脸视频图像,采用 Haar 特征和基于 AdaBoost 算法的目标检测手段确定人脸区域,再通过图像分析手段提取 2 个眼角、2 个嘴角和鼻尖共 5 个特征点;然后,以一幅正面图像的特征点位置为参考,对实时图像中人脸的空间姿态进行估计;最后,根据人脸的 3 个偏转角定义并生成人机交互信息,用于计算机人机界面中鼠标位置的移动和操作控制。实验表明,在良好的光照条件下,该方法可以快速准确完成人脸姿态估计并获得鼠标控制信息,帮助伤残人士实现人机信息交互。

关键词:人机交互;图像处理;人脸检测;姿态估计;视觉鼠标

中图分类号:TP391.41

文献标志码:A

Human computer interaction with face poses estimation

MAO Yu-xing, YU Yan-lai, HAN Bei-tao, XU Shao-zhi, CHENG Hua-an

(State Key Laboratory of Power Transmission Equipment & System Security and New Technology, Chongqing University, Chongqing 400044, P. R. China)

Abstract: Focusing on the problem of computer operation without hand for the disabled people, a new method of mouse control based on visual analysis of the face poses is proposed. First, the face images from CCD camera are obtained and the face areas are located based on AdaBoost algorithms with Haar features. Then, 5 feature points including 2 eye corners, 2 mouth corners and nose tip are detected with image analysis. After that, with help of the feature points in a predefined frontal face image, the face poses are estimated in the real-time face images. Finally, the 3 deflection angles are adopted to define human computer interactive information, which can be used to determine positions and operations of the mouse in the human computer interactive interface. The experimental results demonstrate the efficiency of our method that the face poses can be estimated accurately as well as the operation information to the mouse be achieved rapidly. It is valuable in human computer interaction for the disabled people.

Key words: human computer interaction; image processing; face detection; pose estimation; visual mouse

人机交互技术(Human Computer Interaction Techniques)^[1-4]是指通过计算机输入输出设备,以有效的方式实现人与计算机对话的技术。它包括机

器通过输出或显示设备给人提供大量有关信息及提示,人通过输入设备给机器输入有关信息,回答问题等。人机交互技术是计算机用户界面设计中的重要

收稿日期:2011-08-12

基金项目:国家自然科学基金资助项目(50877071);重庆市科技公关资助项目(CSTC2010AA5049);“输配电装备及系统安全与新技术”国家重点实验室自主研究项目(2007DA10512709213)

作者简介:毛玉星(1967-),男,重庆大学副教授,博士,主要从事图像处理与计算机视觉,计算机测量与仪器研究。
(E-mail)myx@cqu.edu.cn。

内容之一,它与认知学、人机工程学、心理学等学科领域有密切的联系。在人机交互领域,现在广泛使用的有键盘、鼠标、光笔等传统装置,然而在一些特殊情况下传统装置有其局限性,例如某些多媒体游戏的复杂界面操作以及残疾人使用计算机,迫切需要设计不依赖肢体动作的输入装置作为现有人机交互方式的补充。

为了适应不同的交互环境,适用于不同的交互人群,国内外学者已经广泛研究基于视觉(图像)、听觉(语音)、触觉(压力、温度)等交互方法,有些方法已经初步投入使用。在视觉交互方法中,目前主要有视线追踪技术^[2-4]、手语识别技术^[5]等研究热点。视线追踪技术是通过摄像头和瞳孔定位技术感知人的关注视点,从而驱动鼠标定位及相关操作,然而有两个重要缺点:一是瞳孔移动是不连续的,不能与连续移动的视点很好吻合,降低了跟踪的准确性;二是由于生理与心理原因,人眼的移动具有随意性,某些情况下人眼视线所在位置并不反应自身所关注的主观意愿。

为克服视线跟踪等人机交互方法的缺陷,笔者采用人脸姿态估计方法,利用视频摄像头获得人脸图像,通过图像与视频处理技术实现人脸检测与关键点定位,根据多帧序列图像的特征点位置分布估计人脸姿态,并由姿态信息实现人机交互,适用于一些特殊的交互人群(如残疾人)和交互环境(如多媒体游戏)。

1 系统概述

利用人脸姿态变化定义鼠标移动和操作信息,既可以克服视线跟踪人机交互方法的一些固有缺陷,又可以不依赖于人的肢体运动拓宽应用领域,其依赖性仅在于准确估计人脸姿态。笔者采用 5 个人脸特征点位置计算空间偏转角,具有实现方便、精度高、速度快的优势。在特征点定位方面,已有许多行之有效的方法提出^[6-8]。鉴于系统用于人机交互环境,可以在客观上保证良好的光照条件。此外,为加快人脸特征点提取速度并提高准确性,可以先进行人脸区域定位。目前人脸检测方法已经有大量文献可供参考^[9-14],技术已非常成熟,其中 Viola 等提出的基于 Haar 特征的 AdaBoost 目标检测方法^[12,13]通过积分图加快了特征提取速度,实现了人脸的实时检测。在姿态判别方面,目前也有多种算法提出^[15-16],其中 Chen 等提出了一种视频序列的人脸姿态估计方法^[15],采用人脸区域的椭圆模型和嘴相对人脸区域的位置和方向估计人脸姿态,其缺点是

当人脸出现头发遮挡等干扰时将影响模型建立。Gong 等提出了一种利用人脸三维模型与二维人脸图像特征点相结合实现姿态估计的新方法^[16],具有较好的效果,但是计算复杂,同时估计精度依赖于基于单幅正面图像的人脸三维模型建立的准确性,这本身是一个具有挑战性的问题。笔者所用系统中直接采用两幅图像的 5 个特征点估计人脸姿态,只要特征点定位准确,就能保证估计精度,同时计算速度快,可以满足实时性要求。系统组成见图 1。

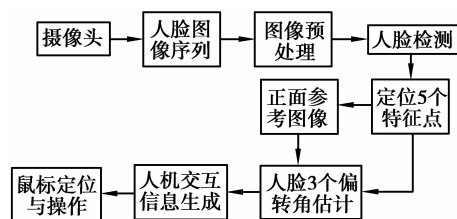


图 1 系统结构框图

对从摄像头采集的人脸序列图像进行消噪和增强等预处理后,首先进行人脸检测确定初步位置,再定位 2 个眼角、2 个嘴角和鼻尖 5 个特征点,然后以一幅正面图像的特征点位置为参考,结合任意图像的特征点位置计算人脸的 3 个偏转角度,最后依据这些角度定义鼠标的位置和操作方法,从而实现人机交互。

2 算法实现

2.1 人脸特征点定位

系统涉及的人脸空间姿态估计依赖于文中定义的 5 个特征点位置,因此人脸特征点的精确定位是准确估计人脸空间偏转角度的关键。系统的具体应用环境可以满足良好的光照条件,为特征点定位奠定了基础。为减少计算复杂度,提高定位准确性,先通过人脸检测方法确定搜索范围,然后在指定范围内定位特征点。

2.1.1 人脸检测

在人脸检测方面,目前广泛使用的是基于 Haar 特征的 AdaBoost 目标检测算法^[12,13]。该方法最初由 Viola 等学者于 2001 年提出。他们采用待检测目标区域的多种 Haar 特征值作为目标的特征描述,采用 AdaBoost 方法,根据特征对训练样本的分类能力进行特征选择,并对被选特征进行加权组合得到分类器。考虑到 Haar 特征主要表现为图像的矩形区域灰度值求和,他们创造性地提出了积分图方法加快计算速度。也正是他们在这一领域的特殊贡

献,使实时人脸检测成为可能。最初他们提出的 Haar 特征为 4 种,随后为加强目标检测中对偏转目标的适应性,一些学者对 Haar 的旋转特征进行了扩展,大大提高了目标检测的准确度。

基于 Haar 特征的 AdaBoost 目标检测算法已成为 OpenCV 的功能部件。OpenCV 是 Intel 公司开发的开源图像处理和计算机视觉函数库,它由一系列 C 函数和少量 C++ 类构成,实现了图像处理和计算机视觉方面的很多通用算法。笔者采用 OpenCV 实现人脸检测,在较好的光照条件下几乎没有误检,并具有良好的实时性,为特征点定位创造了条件。图 2 是不同姿态的人脸检测结果。



图 2 应用 OpenCV 的人脸检测结果

2.1.2 人脸特征点定位方法

用于人脸姿态估计的特征点定义为 2 个眼角、2 个嘴角和鼻尖,见图 3 中的 A、B、C、D、E 点。图中的 F 和 G 是虚拟点,分别由 A、B 和 D、E 点求中点得到,用于角度计算。此外,为便于上述特征点检测,对人脸检测结果划分成 5 个区域,其中第 4 区的高度是其他各区域的 2 倍,见图 4。经过划分,在人脸偏转角不太大的情况下,左右眼睛会位于第 2、3 区,鼻子和嘴则分别位于第 4、5 区,为特征点定位创造了条件。文中对特征点的定位流程如图 5 所示。

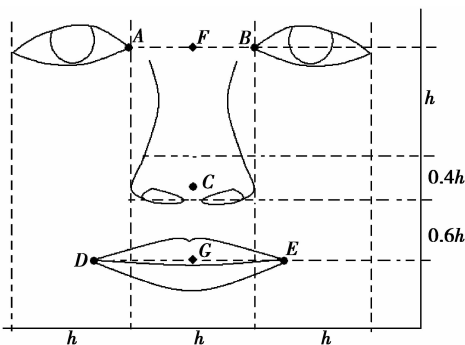


图 3 人脸特征点定义

首先对人脸的眼和鼻子区域利用 SUSAN 算子进行角点检测^[17]。SUSAN 算子只对候选点周边像素的灰度值进行运算,获得的变换图像对角部敏感,具有较大的灰度值,非常适合用于角点检测,同时其



图 4 人脸区域划分

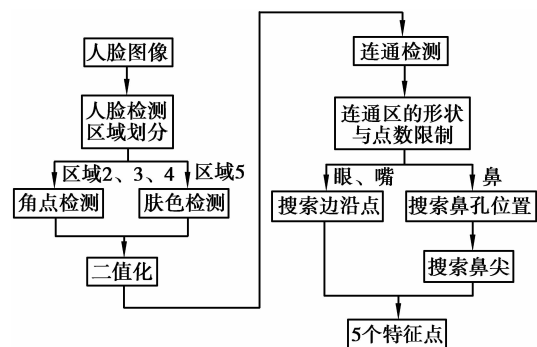


图 5 人脸特征点定位流程

抗噪声能力很强,运算量小;对于嘴部,角点检测的效果不理想,虽然唇部与面部颜色反差较大,但转换为灰度后却很接近,而 SUSAN 检测方法是针对灰度图像的,因此嘴的检测效果远不如眼睛和鼻孔。笔者采用广泛使用的基于高斯混合模型(GMMs)的肤色检测方法实现嘴和面部皮肤的区分。高斯混合模型可以在 Cr、Cb 颜色空间采用多个二维高斯模型估计皮肤的颜色分布,对光线的强弱具有鲁棒性。系统中用两个高斯模型,从网络上搜集了大量的不同人种和光照条件的皮肤样本,采用最大期望(Expectation-Maximization, EM)算法训练模型参数,得到皮肤模型后,将面部各点的色彩值代入,依据皮肤的似然概率决定变换值。最后,分别将上面两种方法得到的图像做二值化处理。为增强二值化过程对光照强弱的适应能力,文中采用 Otsu 算法计算阈值。图 6 展示了图 2 的人脸图像的角点检测和肤色检测结果,其中角点检测结果用于眼、鼻检测,而肤色检测结果用于确定嘴的位置。

上述方法得到的人脸器官图像,会存在较多的离散点,影响特征点定位,因此需要消除干扰。考虑到人的眼、鼻、嘴的几何特征,文中采用连通检测方

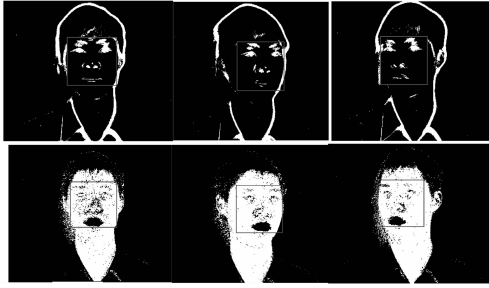


图 6 人脸图像的角点检测和肤色检测结果

法,提取各区域内的连通信息。对所有连通区得到的连通点数量和外接矩形的宽高信息,并分别针对不同器官制定约束条件,将不符合条件的点去掉,为特征点的搜索创造了条件:

1) 搜索眼角点,区域 2 为右眼,从区域 2 最右边一列像素开始从下往上搜索,搜索到的第一个为白色的像素点即为右眼角点;同理搜索左眼角点则从区域 3 最左边一列像素开始从下往上搜索,第一个为白色的像素点即为左眼角点。

2) 搜索嘴角点,在区域 5 中搜索嘴的左下和右下角点作为两个嘴角点。

3) 搜索鼻尖,对区域 4 做水平方向的灰度统计,经平滑处理后以极大值的位置作为鼻孔的纵坐标,再在鼻孔上方搜索高光点即为鼻尖的大致位置。但是如果人脸姿势有角度偏转,高光点往往并不在鼻尖处。由于角度的偏转,会使高光部分沿着鼻梁向鼻尖处延伸,因此为了调整鼻尖的定位,可以在原来搜索到的高光点以下 5 个像素的位置再次左右搜索高光点,以此高光点的横坐标和原高光点的纵坐标确定鼻尖的位置。

2.2 人机交互信息生成

2.2.1 人脸空间姿态估计

系统中的人机交互依赖于人脸的空间姿态,需要通过前面得到的特征点位置进行估计。定义人脸的 3 个空间偏转角 α, β, γ , 见图 7。

首先对特征点的坐标进行预处理,参照图 3,对 A、B 两点坐标求平均得到中心点 F。指定 F 为坐标原点,水平向右为横轴 u , 竖直向上为纵轴 v , 对后面所有图像点的二维坐标均以此为参考进行变换。对 D、E 两点坐标求平均得到中心点 G, 连同两个眼角、眼角中心点和鼻尖共 5 个特征点 A、B、C、F、G。

确定一幅正面图像, 设定其 3 个偏转角 α, β, γ 为 0, 定位 5 个特征点 A~E, 按照上面方法计算 A、B、C、F、G 点的坐标并记录。作为参考图像, 以下标 1 作为标记。于是这些坐标点描述为: A 点 $(uA1, vA1)$, B 点 $(uB1, vB1)$, C 点 $(uC1, vC1)$, F 点 $(uF1,$

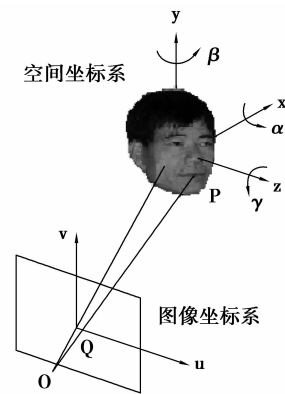


图 7 人脸 3 个偏转角定义

$vF1)$, G 点 $(uG1, vG1)$ 。对任意时刻采集的人脸图像, 同理得到 5 个点的坐标, 以下标 2 为标记: A 点 $(uA2, vA2)$, B 点 $(uB2, vB2)$, C 点 $(uC2, vC2)$, F 点 $(uF2, vF2)$, G 点 $(uG2, vG2)$ 。

对于两幅图像中 5 个点的坐标, 利用针孔摄像机模型和对极几何原理进行数学推导, 计算出图像 2 中人脸的姿态, 得到 3 个偏转角度 α, β, γ 。针孔摄像机模型:

$$\hat{sm} = K[R | T]\hat{M}, \quad (1)$$

其中 s 为尺度因子, $\hat{m} = [u, v, 1]^T$ 图像点的齐次坐标, K 为像机的内部参数矩阵, 可以定义成 $\text{diag}(f, f, 1)$ 。 $T = [t_x, t_y, t_z]^T$ 是平移矩阵, $\hat{M} = [x, y, z, 1]^T$ 是空间人脸三维点的齐次坐标。 R 是以 α, β 和 γ 为基础的旋转矩阵。可以根据两幅已知图像的 5 个特征点的 \hat{m} 信息, 通过一些合理的假设, 在不知道 \hat{M} 的情况下估算 R 中的 3 个角度值。

由于只涉及角度估计, 算法中的运算式全部以差-商方式出现, 可以对针孔相机模型的参数作如下假设: $s=1, K=\text{diag}(1, 1, 1), T=[0, 0, 0]^T$ 。设:

$$M = \frac{v_{B2} - v_{A2}}{u_{B2} - u_{A2}}, O = \frac{v_{G2} - v_{F2}}{u_{G2} - u_{F2}}, P = \frac{u_{B1} - u_{A1}}{u_{B2} - u_{A2}}. \quad (2)$$

$$c = \cos\gamma, d = \sin\gamma, h = \sin\beta. \quad (3)$$

可以推导出:

$$\gamma = \arctan M, \quad (4)$$

$$\beta = \arccos \frac{1}{P_c}, \quad (5)$$

$$\alpha = \arctan \left(\frac{c + Od}{Ohc - hd} \right). \quad (6)$$

由于在计算过程中, β 存在一个正负符号问题, 不能由 A、B、F、G 4 点唯一确定, 因此需要引入 C 点, 根据 C 与 F、G 连线的相对位置决定其符号。C 点是鼻尖位置, 由于只用于决定 β 的符号, 即确定人

脸的左右偏转问题, C 点的定位误差通常不会对角度计算产生影响。

2.2.2 人机交互信息生成

计算出 α 、 β 、 γ 3 个角度值后, 可用于定义鼠标位置和操作方式。 α 、 β 为 0 定义为屏幕中心, α 变化时鼠标上下移动, β 变化时鼠标左右移动, 角度大于或等于 20° 时鼠标定位到屏幕边沿。 γ 角度的突变定义鼠标操作方式, γ 为正且连续两帧之间角度变化介于 $3^\circ \sim 8^\circ$ 为单击左键, 超过 8° 为双击左键, γ 为负且连续两帧之间角度变化介于 $-3^\circ \sim -8^\circ$ 为单击右键, 超过 -8° 为双击右键。由于 γ 用于定义鼠标操作, 规定在引起鼠标操作时忽略 α 和 β 的变化, 即鼠标位置固定, 指定所选目标。

3 实验结果

根据文中算法, 采用 VC++ 编写了相关的软件, 其中人脸检测部分调用 OpenCV 函数库实现。采用一个 CCD 摄像头和视频采集卡实现人脸图像实时采集。软件实现了人脸特征点定位和鼠标控制信息生成, 并能按照设计方案驱动鼠标正常操作。图 8 是特征点检测和姿态估计结果。实验表明, 当光照良好、各个待检特征点无遮挡时, 检测效果好, 并对人脸表情具有一定鲁棒性。由于人脸皮肤纹理和光照影响, 容易引起高光点的误检, 造成鼻尖定位不准确, 但由于鼻尖定位受鼻孔位置约束, 通常偏差不大, 根据 2.2.1 节的分析, 对角度计算不会产生大的影响。软件的处理速度达到每秒 15 帧, 分析表明, 时间的消耗主要在于人脸区域定位和连通检测算法, 需要进一步完善与优化。



$\alpha=0^\circ$	$\alpha=-2.5^\circ$	$\alpha=-3.7^\circ$
$\beta=0^\circ$	$\beta=14.6^\circ$	$\beta=-16.8^\circ$
$\gamma=0^\circ$	$\gamma=-0.9^\circ$	$\gamma=-1.8^\circ$

图 8 人脸特征点检测和姿态估计结果

4 结语

笔者提出了一种基于人脸姿态估计的人机交互方法, 通过人脸检测, 5 个特征点的提取实现人脸的姿态估计, 计算出人脸的 3 个空间偏转角度并由此定义鼠标的位置和操作信息, 成为一种附加的人机

交互方法, 可以服务于肢体残疾人士进行简单的人机操作或多媒体环境。该方法可以弥补传统的视线跟踪方法的一些缺陷, 具有一定的应用价值。实验表明, 在良好的应用环境下, 特征点定位和姿态估计效果良好, 但需要进一步提高处理速度, 并改进交互操作信息的定义方法, 从而提高系统效率和实用性。

参考文献:

- [1] DIX A, FINLAY J, GREGORY D A. Human-computer interaction [M]. 3rd edition. Prentice Hall, 2003: 25-40.
- [2] VENERI G, FEDERIGHI P, ROSINI F, et al. Influences of data filtering on human-computer interaction by gaze-contingent display and eye-tracking applications [J]. Computers in Human Behavior, 2010, 26 (6): 1555-1563.
- [3] HAFEGEE A, BARROW R. Eye tracking and gaze based interaction within immersive virtual environments [J]. Lecture Notes in Computer Science, 2009, 5545: 729-736.
- [4] STRANDVALL T. Eye tracking in human-computer interaction and usability research [J]. Lecture Notes in Computer Science, 2009, 5727: 936-937.
- [5] 郭康德, 张明敏, 孙超, 等. 基于视觉技术的三维指尖跟踪算法 [J]. 计算机研究与发展, 2010, 47 (6): 1013-1019.
GUO KANG -DE, ZHANG MING -MIN, SUN CHAO, et al. 3D fingertip tracking algorithm based on computer vision [J]. Journal of Computer Research and Development, 2010, 47(6): 1013-1019.
- [6] ZHOU H L, XIE L J, FANG X L. Visual mouse: SIFT detection and PCA recognition [C] // Proceedings of 2007 International Conference on Computational Intelligence and Security Workshops, Dec. 15-19, 2007, Harbin, China. [S. l.]: IEEE Press, 2007: 263-266.
- [7] HUANG C, HUANG Y. Facial expression recognition using model-based feature extraction and action parameters classification [J]. Journal of Visual Communication and Image Representation, 1997, 8(3): 278-290.
- [8] VUKADINOVIC D, PANTIC M. Fully automatic facial feature point detection using Gabor feature based boosted classifiers [C] // Proceedings of 2005 IEEE International Conference on Systems, Man and Cybernetics, October 10-12, 2005, Waikoloa, Hawaii, USA. [S. l.]: IEEE Press, 2005 (2): 1692-1698.
- [9] ROWLEY H A, et. al. Neural network -based face detection [J]. IEEE Transactions on Pattern Analysis

- and Machine Intelligence, 1998, 20(1):23-38.
- [10] SUNG K K, POGGIO T. Example-based learning for view-based human face detection [J]. IEEE Transactions Pattern analysis and Machine Intelligence, 1998, 20(1):39-51.
- [11] SOCHMAN J, MATAS J. AdaBoost with totally corrective updates for fast face detection [C] // Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition, May 17-19, 2004, Seoul, Korea. [S.l.]: IEEE Press, 2004: 445-450.
- [12] JONES M, VIOLA P. Fast multi-view face detection. Technical report Mitsubishi Electric Research Laboratories[EB/OL]. (2003-08-01). <http://faculty.ucmerced.edu/mhyang/.../face-detection-chapter.pdf>.
- [13] VIOLA P, JONES M. Rapid object detection using a boosted cascade of simple features [J]. Computer Vision and Pattern Recognition, 2001(1): 8-14.
- [14] SCHNEIDERMAN H. A statistical approach to 3D object detection applied to faces and cars [D]. Pittsburgh: Robotics Institute, Carnegie Mellon University, 2000.
- [15] CHEN X, YANG Q, LIAO H, et. al. Real-time face pose estimation in video sequence [C] // Proceedings of the 2010 2nd International Workshop on Education Technology and Computer Science, March 6-7, 2010, Wuhan, China. [S.l.]: IEEE Press, 2010, 2: 24-27.
- [16] GONG X, WANG G, XIONG L. Single 2D image-based 3D Face reconstruction and its application in pose estimation [J]. Fundamenta Informaticae, 2009, 94(2):179-195.
- [17] STEPHEN M, MICHAEL J, SMITH B. SUSAN-a new approach to low level image processing [J]. International Journal of Computer Version, 1997, 23(1):45-78.

(编辑 王维朗)

~~~~~

(上接第 51 页)

- [3] 惠希增. 变频电磁场处理油田水防垢技术研究[D]. 北京:中国石油大学,2007.
- [4] BAKER J S, JUDD S J. Magnetic amelioration of scale formation[J]. Water Research, 1996,3(2):247-260.
- [5] 曲久辉,刘会娟. 水处理电化学原理与技术[M]. 北京:科学出版社,2007.
- [6] 冯慈璋,马西奎. 工程电磁场导论[M]. 北京:高等教育出版社,1999.
- [7] 保罗. 电磁兼容导论[M]. 温映红,译. 北京:机械工业出版社,2006.
- [8] OPPENHEIM A V, SCHAFER R W. 数字信号处理[M]. 董士嘉,译. 北京:科学出版社,1981.
- [9] 王博,何为. 高频电子阻垢仪的研制与实验研究[J]. 化工自动化及仪表,2009,36(6):49-52.  
WANG BO, HE WEI. High frequency electronic scale apparatus and experimental study [J]. Chemical Engineering, 2009, 36 (6):49-52.
- [10] 杨萍,兀旦晖,杨良煜. DDS 技术在正弦信号发生器中的应用 [J]. 计算机测量与控制, 2008, 16 (11): 1738-1740.  
YANG PING, WU DAN-HUI, YANG LIANG-YU. DDS sine wave generator in the application [J]. Computer Measurement and Control, 2008, 16 (11): 1738-1740.
- [11] KOBE S, MCGUINNESS P J. The influence of the magnetic field on the crystallisation form of calcium carbonate and the testing of a magnetic water-treatment device [J]. Journal of Magnetism and Magnetic Materials, 2001, 236(1-2):71-76.
- [12] KOBE S, DRAZIC G, CEFALAS A C, et al. Nucleation and crystallization of CaCO<sub>3</sub> in applied magnetic fields[J]. Crystal Engineering, 2002, 5 (3-4): 243-253.
- [13] COEY J M D, CASS S. Magnetic water treatment[J]. Journal of Magnetism and Magnetic Materials, 2000, 209:71-74.
- [14] ALIMI F. Influence of magnetic field on calcium carbonate precipitation in the presence of foreign ions[J]. Surface Engineering and Applied Electrochemistry, 2009, 45(1): 56-62.
- [15] XING. X K, Research on the electromagnetic anti-fouling technology for heat transfer enhancement[J]. Applied Thermal Engineering, 2008, 28 (8-9): 889-894.

(编辑 王维朗)