

文章编号:1007-2780(2019)04-0417-06

基于深度卷积神经网络的多特征融合的手势识别

负卫国,史其琦*,王 民

(西安建筑科技大学 信息与控制工程学院, 陕西 西安 710055)

摘要:针对传统的分类方法由于提取的特征比较单一或者分类器结构过于简单,导致手语识别率较低的问题,本文将深度卷积神经网络架构作为分类器与多特征融合算法进行结合,通过使用纹理特征结合形状特征做到有效识别。首先纹理特征通过 LBP、卷积神经网络和灰度共生矩阵方法得到,其中形状特征向量由 Hu 氏不变量和傅里叶级数组成。为了避免过拟合现象,使用“dropout”方法训练深度卷积神经网络。这种基于深度卷积神经网络的多特征融合的手语识别方法,在“hand”数据库中,对 32 种势的识别率为 97.73%。相比一般的手语识别方法,此方法鲁棒性更强,并且识别率更高。

关键词:手势识别;手势提取;多特征融合;深度卷积神经网络;鲁棒性

中图分类号:TP391.41 **文献标识码:**A **doi:**10.3788/YJYXS20193404.0417

Multi-feature fusion gesture recognition based on deep convolutional neural network

YUN Wei-guo, SHI Qi-qi*, WANG Min

(School of Information and Control Engineering,
Xi'an University of Architecture and Technology, Xi'an 710055, China)

Abstract: To solve the problem that the traditional classification method has a relatively single extraction feature or the structure of the classifier is too simple, leading to a low recognition rate of the sign language, in this paper, the deep convolutional neural network architecture was combined as a classifier and a multi-feature fusion algorithm, and it was effectively identified by using texture features and shape features. Firstly, the texture features were obtained by LBP, convolutional neural network and gray-level co-occurrence matrix methods, in which the shape feature vector consisted of Hu's invariant and Fourier descriptors. To avoid overfitting, the "dropout" method was used to train deep convolutional neural networks. This kind of multi-feature fusion sign language recognition method based on deep convolutional neural network has a recognition rate of 97.73% for 32 gestures in the "hand" data-

收稿日期:2018-10-16;修订日期:2019-03-07.

基金项目:国家自然科学基金(No.61373112);住房城乡建设部科学技术项目计划(2016-R2-045);陕西省自然科学基金基础研究资金(No.2014JM8348)

Supported by National Natural Science Foundation of China (No.61373112); Ministry of Housing and Urban-Rural Development Science and Technology Project plan(2016-R2-045); Fundamental Research Fund of Natural Science of Shaanxi Province (No.2014JM8348).

* 通信联系人, E-mail: sqq187000978170@163.com

base. Compared with the general sign language recognition method, this method is more robust and has a higher recognition rate.

Key words: gesture recognition; gesture extraction; multi-feature fusion; deep convolutional neural network; robustness

1 引言

由于手语特征具有多样性,因此在以计算机作为辅助的手语识别其研究对象通常都是手势图像识别。手势识别的关键是手势提取,即从手势图像中挖掘出能够部分或完全表达手势的信息的如形状上下文等基于形状的^[1];局部二值模式(Local Binary Patterns, LBP)等基于纹理的^[2];傅里叶描述子等基于频率特征的^[3]等有意义的特征。由于手势特征的类别多种多样,为了提高手势识别的准确率将从 2 个方面入手^[4]:(1) 手势特征的提取。手势特征的采集容易受环境影响而造成手势的几何特征和纹理特征有较大浮动,因此必须选用一些表达能力较强的特征来描述手势特征^[5]。(2) 识别分类器的选择。例如,尚小晶等^[6]提出一种基于改进概率神经网络(Probabilistic Neural Network, PNN)的手势动作识别方法。吴健健等^[7]提出一种基于多特征融合与支持向量机(Support Vector Machine, SVM)的手势识别方法。

基于以上分析,本文提出了一种融合纹理特征、几何特征和卷积神经网络(Convolutional Neural Network, CNN)等特征进行手势识别的方法,并且将深度卷积神经网络(Deep Convolutional Neural Network, DCNN)作为分类器引入到手语识别中,从而来提高手语的识别率,并将此方法应用于建立的手语数据库-“hand”数据库中进行。

2 手势特征提取

2.1 手势图像预处理

手势图像的特征众多,因此需要对手势图像进行预处理。手势图像预处理的步骤如图 1 所示,首先将手势 RGB 图像转换为手势灰度图像从而消除颜色特征的影响;其次将灰度图像转换成二值图像并选取二值图像的前景部分作为手势提

取区域^[8]。

2.2 局部二值模式特征

LBP 是描述手势图像纹理特征最常见的一种形式。LBP 算子能够从 3×3 大小的窗口中,比较中心点像素与相邻点像素的灰度值^[9]。LBP 局部区域纹理计算方法如公式(1)所示:

$$K_{LBP}(x_c, y_c) = \sum_{p=0}^7 s(g_p - g_c) 2^p, \quad (1)$$

其中: g_p ($p=0, 1, \dots, 7$) 表示中心点相邻点的灰度, (x_c, y_c) 表示中心点的坐标,其中函数 $s(g_p - g_c) = \begin{cases} 1, & g_p \geq g_c \\ 0, & g_p < g_c \end{cases}$, g_c 表示中心点的灰度值。

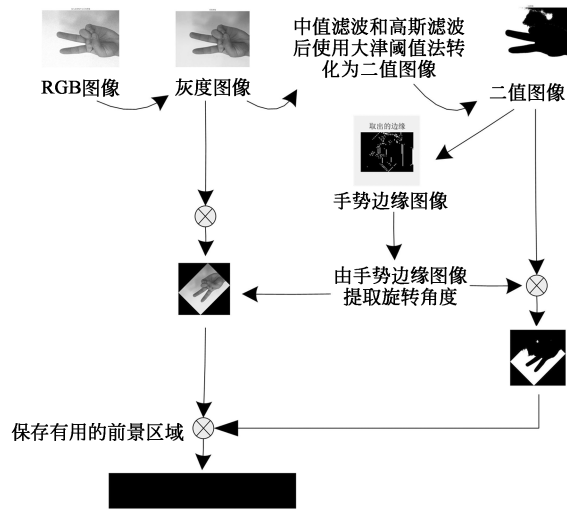


图 1 手势图像预处理的步骤

Fig.1 Gesture image preprocessing steps

2.3 卷积神经网络

手势识别的算法流程图如图 2 所示。

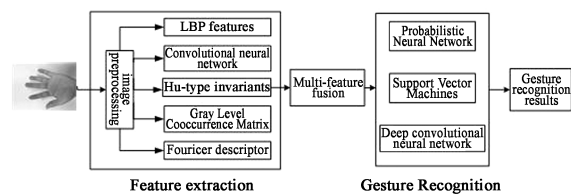


图 2 手势识别的算法流程图

Fig.2 Gesture recognition algorithm flow chart

CNN 具有 2 个最大的特点:(1)局部连接性,图像中距离越近的像素之间的影响也就越大;(2)权值共享性,图像中区域间的权值是可以共享的^[10]。概况的讲,卷积层的计算方法就是根据公式(2):

$$conv = \sigma(imgMat * W + b), \quad (2)$$

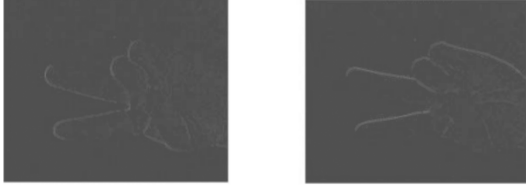


图3 多个卷积核运算提取图像特征

Fig.3 Multiple convolution kernel operations to extract image features

式中:“ σ ”表示激活函数;“ $imgMat$ ”表示灰度图像矩阵;“ $*$ ”表示卷积操作;“ W ”表示卷积核;“ b ”表示偏置值。

图3中我们分别使用了2个不同卷积核进行卷积运算从而提取图像特征,得到了不同的图像特征。实际中,我们会采用多个卷积核来提取图像特征,进而来进行下一步的运算。

2.4 Hu 氏不变量

Hu 氏不变量是具有旋转、平移、尺度不变性的量^[11]。函数 $f(x, y)$ 的二维 Hu 氏 p, q 阶统计量定义如式(3)所示,式中 p, q 为非负整数。

$$M_{pq} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^p y^q f(x, y) dx dy, \quad (3)$$

考虑质心 (x_c, y_c) 时,则式(3)变成式(4):

$$\mu_{pq} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - x_c)^p (y - y_c)^q f(x, y) dx dy, \quad (4)$$

其中: $x_c = \frac{M_{10}}{M_{00}}, y_c = \frac{M_{01}}{M_{00}}$ 。

平移不变性通过以上统计量结果得到,尺度不变性由公式(5)计算确保。

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^\gamma}, \quad (5)$$

式中:标准化因子 $\gamma = \frac{p+q}{2} + 1$ 。

Hu 氏不变量见式(6)~(12),这些特征量可以用来描述手势图像特征。

$$\varphi_1 = \eta_{20} + \eta_{02}, \quad (6)$$

$$\varphi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2, \quad (7)$$

$$\varphi_3 = (\eta_{30} - 3\eta_{12})^2 + (\eta_{03} - 3\eta_{21})^2, \quad (8)$$

$$\varphi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{03} + \eta_{21})^2, \quad (9)$$

$$\varphi_5 = (\eta_{30} - \eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{21} + \eta_{03})^2 - (\eta_{21} + \eta_{03})^2], \quad (10)$$

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

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

在我们的实验中,为了满足深度卷积神经网络的输入,特征选取 7 个 Hu 不变量的绝对值的对数值。

2.5 灰度共生矩阵

灰度共生矩阵(Gray Level Cooccurrence Matrix, GLCM)用来反映手势灰度图像的相邻像素之间的梯度信息量^[4]。式(13)是灰度共生矩阵的计算公式:

$$P_{GLCM}(x, y) = \frac{G_{GLCM}(x, y)}{\sum_{x=0}^{L-1} \sum_{y=0}^{L-1} G_{GLCM}(x, y)}, \quad (13)$$

式中: $P_{GLCM}(x, y)$ 为归一化后的 GLCM 值, $G_{GLCM}(x, y)$ 为原 GLCM 值, L 为 GLCM 级数。

对比度、逆差矩、能量、熵和相关性这 5 个量^[12]的计算方法依次如式(14)~(18)所示。

$$f_1 = \sum_{x=y=0}^{L-1} (x - y)^2 \left\{ \sum_{x=0}^{L-1} \sum_{y=0}^{L-1} P_{GLCM}(x, y) \right\}, \quad (14)$$

$$f_2 = \sum_{x=0}^{L-1} \sum_{y=0}^{L-1} \frac{P_{GLCM}(x, y)}{1 + (x - y)^2}, \quad (15)$$

$$f_3 = \sum_{x=0}^{L-1} \sum_{y=0}^{L-1} P_{GLCM}(x, y)^2, \quad (16)$$

$$f_4 = - \sum_{x=0}^{L-1} \sum_{y=0}^{L-1} P_{\text{GLCM}}(x, y) \lg P_{\text{GLCM}}(x, y), \quad (17)$$

$$f_5 = \frac{\sum_{x=0}^{L-1} \sum_{y=0}^{L-1} P_{\text{GLCM}}(x, y) - u_1 u_2}{\delta_1^2 \delta_2^2}, \quad (18)$$

$$\begin{aligned} \text{式中: } u_1 &= \sum_{x=0}^{L-1} x \sum_{y=0}^{L-1} P_{\text{GLCM}}(x, y), \\ u_2 &= \sum_{y=0}^{L-1} y \sum_{x=0}^{L-1} P_{\text{GLCM}}(x, y), \\ \delta_1^2 &= \sum_{x=0}^{L-1} (x - u_1)^2 \cdot \sum_{y=0}^{L-1} P_{\text{GLCM}}(x, y), \\ \delta_2^2 &= \sum_{y=0}^{L-1} (y - u_2)^2 \cdot \sum_{x=0}^{L-1} P_{\text{GLCM}}(x, y). \end{aligned}$$

2.6 傅里叶级数特征

傅里叶级数特征是一种常用的表示手势轮廓的方法,傅里叶级数特征是将手势轮廓变换到频域上进行分析,其高频部分可以清晰表达图像的形状纹理特征^[13]。将预处理后的手势边缘图像排序为序列 $(x_0, y_0), (x_1, y_1), \dots, (x_{K-1}, y_{K-1})$ 。其中 (x_0, y_0) 是首坐标, K 是范围内的点数。每个坐标都是 1 个复数 $s(k) = x(k) + j(k)$ ($k = 1, 2, \dots, K$), 对其进行傅里叶变换可得到傅里叶系数 $a(u)$ ($u = 0, 1, 2, \dots, K-1$) :

$$a(u) = \frac{1}{K} \sum_{k=0}^{K-1} s(k) e^{-j \frac{2\pi u k}{K}}. \quad (19)$$

手势轮廓的傅里叶级数特征是 $a(u)$ 傅里叶反变换。

$$s(k) = \sum_{u=0}^{K-1} a(u) e^{j \frac{2\pi u k}{K}}. \quad (20)$$

3 深度卷积神经网络

深度卷积神经网络具有更复杂的网络结构以及更强大的特征学习和表达能力^[14]。DCNN 可以看作是一种机器学习和深度学习交叉的产物。1990 年 LeCun 等^[15]提出了使用梯度反向传播算法训练的卷积神经网络模型。本文将 DCNN 作为分类器应用于手势识别从而解决手势识别问题。

3.1 DCNN 训练方法

深度卷积神经网络(DCNN)是由若干个单层卷积神经网络相互嵌套组成的,主要处理对象是二维输入数据,是一种监督型深度学习网络结

构^[16]。如图 4 所示一种最常见的深度卷积神经网络模型,其由 2 个卷积层(C_1, C_2)、2 个下采样层(S_1, S_2)和 1 个全连接层堆叠而成。单层卷积神经网络的权值共享是通过约束神经元来实现的。各网络层的输入和输出都是由一组维度确定的特征向量来表示,有利于卷积运算降低时耗。

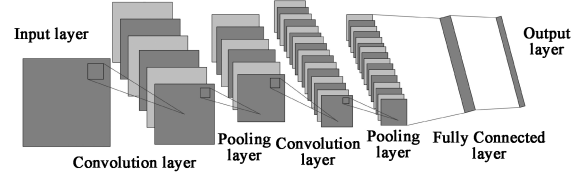


图 4 深度卷积神经网络架构

Fig.4 Deep convolutional neural network architecture

3.2 使用“dropout”方法的 DCNN

过拟合现象是训练中极易产生的现象,训练时若测试样本过少,样本参数可能会过度拟合训练数据集,这样容易影响训练数据集的泛化性能,造成过拟合现象从而对测试样本产生错误的分类结果。为了削弱这种情况,本文采用“dropout”方法的深度卷积神经网络通过在训练过程中随机地忽略隐含层中一定比例的节点响应,有效地提高了网络的鲁棒性。

4 实 验

4.1 创建“hand”数据库

本文中的手势数据均由本课题组人员自行录制,针对不同人以及不同背景下的手势分别进行了采集,并将手势图片保存成 jpg 格式,创建了“hand”数据库。为了验证方法的有效性,我们在“hand”数据库中进行试验。试验使用 Lenovo 笔记本电脑 window 10 系统, MATLAB(R2016a 版本)环境,处理器为英特尔 Core i5-7300HQ @ 2.5GHz。

本文所建立的“hand”手势数据库由 5 220 个样本构成,让 15 个不同的人把使用频率较高的 86 个手语词用摄像机采集数据。将所采集的手势图像统一调整成 512×512 (像素)大小,以方便进行操作和处理,并按照手语词对应的意思为图像文件命名,储存在本地硬盘中以备程序调用和对比^[17]。我们随机选择使用“hand”数据库每个 45 张手势图像作为训练样本,5 张作为测试样本。

4.2 试验结果

将多特征进行融合并采用不同的分类器进行试验对比,其中类别数分别为100维LBP特征,7个Hu氏不变量,7层深度卷积神经网络,10个灰度共生矩阵特征量,400维傅里叶特征,总特征维数为520维。

试验结果见表1。通过对比分析得出文献

[6]中的方法的识别率是最低的,说明PNN作为分类器的识别率最低,其识别效果最差,识别时间过长;文献[7]中的方法识别率中等,SVM的识别效果高于PNN;本文方法采用的分类器DCNN识别率最高可以达到97.73%,识别效果最好且大幅缩短了识别时间。同时可以看出随着特征融合的种类数的增多,其方法识别率随之升高。

表1 “hand”数据库试验结果
Tab.1 “hand” database test result

测试序列	类别数	TLD Classifier	识别率/%	时间消耗/s
形状特征+纹理特征+颜色特征+ Zernike统计矩 Shape feature + Texture feature + Color feature + Zernike moment	32	PNN	92.15	2.365
形状特征+Fourier特征+颜色特征+ 多尺度距离矩阵 Shape + Fourier+Multiscale distance matrix	32	SVM	93.78	2.224
LBP特征+Hu氏不变量+卷积神经网络+ 灰度共生矩阵特+Fourier特征 LBP+Hu+CNN+GLCM+Fourier	32	DCNN	95.26	1.958
LBP特征+Hu氏不变量+卷积神经网络+ 灰度共生矩阵特+Fourier特征 LBP+Hu+CNN+GLCM+Fourier	32	DCNN “dropout”	97.73	1.732

5 结 论

本文提出了一种基于深度卷积神经网络的多特征融合的手势识别算法,将深度卷积神经网络作为分类器融入到手势识别中。伴随着融合特征复杂度

提高,“dropout”方法能够在DCNN中避免过拟合现象,进而有效地提高识别率,对比概率神经网络和支持向量机作为分类器大幅缩短了识别时间,做到了高效分类。同时伴随着深度卷积神经网络层数的增加,“dropout”方法能够通过随机忽略各网络层中一定比例的节点响应,提高算法的鲁棒性。

参 考 文 献:

- [1] 王民,石新源,王雅慧,等.基于Vibe和时空上下文的运动手势跟踪算法[J].液晶与显示,2018,33(1):92-98.
WANG M, SHI X Y, WANG Z H, et al. Moving gesture tracking algorithm based on Vibe and space-time context [J]. *Chinese Journal of Liquid Crystals and Displays*, 2018, 33(1): 92-98. (in Chinese)
- [2] 刘淑琴.图像特征提取方法及其应用研究[D].西安:西北大学,2016.
LIU S Q. Research on Image feature extraction and its application [D]. Xi'an: Northwest University, 2016. (in Chinese)
- [3] 李丹娇,彭进业,冯晓毅,等.结合CSS与傅里叶描述子的手势特征提取[J].计算机工程,2012,38(6):178-180.
LI D J, PENG J Y, FENG X Y, et al. Gesture feature extraction combining CSS with fourier descriptor [J]. *Computer Engineering*, 2012, 38(6): 178-180. (in Chinese)

- [4] 刘念, 阙江明. 基于多特征融合和深度信念网络的植物叶片识别[J]. 北京林业大学学报, 2016, 38(3): 110-119.
LIU N, KUANG J M. Plant leaf identification based on the multi-feature fusion and deep belief networks method [J]. *Journal of Beijing Forestry University*, 2016, 38(3): 110-119. (in Chinese)
- [5] 高良, 闫民, 赵方. 基于多特征融合的植物叶片识别研究[J]. 浙江农业学报, 2017, 29(4): 668-675.
GAO L, YAN M, ZHAO F. Plant leaf recognition based on fusion of multiple features [J]. *Acta Agriculturae Zhejiangensis*, 2017, 29(4): 668-675. (in Chinese)
- [6] 尚小晶, 田彦涛, 李阳, 等. 基于改进概率神经网络的手势动作识别[J]. 吉林大学学报(信息科学版), 2010, 28(5): 459-466.
SHANG X J, TIAN Y T, LI Y, et al. Recognition of gestures and movements based on MPNN [J]. *Journal of Jilin University (Information Science Edition)*, 2010, 28(5): 459-466.
- [7] 吴健健, 陈玮. 基于多特征融合与支持向量机的手势识别[J]. 电子科技, 2015, 28(5): 127-131.
WU J J, CHEN H. Hand gesture recognition based on multi-feature fusion and support vector machines [J]. *Electronic Science and Technology*, 2015, 28(5): 127-131. (in Chinese)
- [8] 刘念. 基于深度信念网络的植物叶片识别研究[D]. 北京: 北京林业大学, 2016.
LIU N. The study of plant leaf recognition based on the deep belief networks [D]. Beijing: Beijing Forestry University, 2016. (in Chinese)
- [9] 黄非非. 基于 LBP 的人脸识别研究[D]. 重庆: 重庆大学, 2009.
HUANG F F. Research on face recognition based on LBP operator [D]. Chongqing: Chongqing University, 2009. (in Chinese)
- [10] 李彦冬, 郝宗波, 雷航. 卷积神经网络研究综述[J]. 计算机应用, 2016, 36(9): 2508-2515, 2565.
LI Y D, HAO Z B, LEI H. Survey of convolutional neural network [J]. *Journal of Computer Applications*, 2016, 36(9): 2508-2515, 2565. (in Chinese)
- [11] WANG X F, HUANG D S, DU J X, et al. Classification of plant leaf images with complicated background [J]. *Applied Mathematics and Computation*, 2008, 205(2): 916-926.
- [12] 高程程, 惠晓威. 基于灰度共生矩阵的纹理特征提取[J]. 计算机系统应用, 2010, 19(6): 195-198.
GAO C C, HUI X W. GLCM-based texture feature extraction [J]. *Computer Systems & Applications*, 2010, 19(6): 195-198. (in Chinese)
- [13] 晋杰. 基于多特征融合的服装图片检索关键技术研究[D]. 昆明: 昆明理工大学, 2017.
JIN J. Research on key techniques of garment image retrieval based on multi-feature fusion [D]. Kunming: Kunming University of Science and Technology, 2017. (in Chinese)
- [14] 卢宏涛, 张秦川. 深度卷积神经网络在计算机视觉中的应用研究综述[J]. 数据采集与处理, 2016, 31(1): 1-17.
LU H T, ZHANG Q C. Applications of deep convolutional neural network in computer vision [J]. *Journal of Data Acquisition and Processing*, 2016, 31(1): 1-17. (in Chinese)
- [15] LE CUN Y, BOSER B, DENKER J S, et al. Handwritten digit recognition with a back-propagation network [M]//TOURETZKY D. *Advances in Neural Information Processing Systems*. San Francisco, CA, USA: Morgan Kaufmann Publishers, 1990: 396-404.
- [16] 吴尧, 邱卫根. 基于改进的深度卷积神经网络的人脸识别[J]. 计算机工程与设计, 2017, 38(8): 2246-2250.
WU Y, QIU W G. Face recognition algorithm based on improved deep convolutional neural network [J]. *Computer Engineering and Design*, 2017, 38(8): 2246-2250. (in Chinese)
- [17] 王雅. 基于改进的 HMM 和自适应技术的非特定人手语识别的研究[D]. 西安: 西安建筑科技大学, 2017.
WANG Y. Research on signer-independent sign language recognition based on improved HMM and adaptive technology [D]. Xi'an: Xi'an University of Architecture and Technology, 2017. (in Chinese)

作者简介:



尹卫国(1962—), 男, 山西运城人, 硕士, 教授, 硕士生导师, 1988 年于西安建筑科技大学获得硕士学位, 主要从事智能识别与控制方面的研究。E-mail: 1372245990@qq.com



史其琦(1993—), 男, 陕西咸阳人, 硕士研究生, 2016 年于西安建筑科技大学获得学士学位, 主要从事智能信息处理相关技术的研究。E-mail: sqq18700978170@163.com