

基于层次形状特征提取模型的图像分类^①

张盛博^② 刘 娜 霍 宏 方 涛^③

(上海交通大学自动化系控制与信息处理教育部重点实验室 上海 200240)

摘 要 研究了基于层次形状特征提取的图像分类。针对从初级视觉皮层(V1)提取的条形特征对目标形状的描述不充分,提出了一种层次模型(V1-V2-V4),以进一步提取角形、曲率特征。模型中 V1 层的条形特征提取采用 Gabor 模拟;V2 层结合了抑制噪声的 3D 高斯差分(DOG),并使用滤波方向相差 90 度的 Gabor 滤波器组提取多尺度角形特征;V4 层通过曲率域计算来描述目标轮廓的形变程度,并最终提取融合曲率与梯度方向的直方图特征。该模型的优势在于,通过角形以及曲率计算的层次表达,可有效增强目标形状的关键特征点(如角点位置)的提取,并且结合曲率与梯度的直方图描述,也可有效弥补单一曲率或梯度特征局部描述不足的问题。在 MNIST 手写数字与 21 类遥感影像上的实验表明,曲率与梯度的融合特征的运用可获得 98.94% 的数字识别精度,同时在遥感影像分类中也可获得较好的分类效果。

关键词 腹侧通路, 形状特征, 曲率, 层次性, 图像分类

0 引 言

图像分类是利用计算机视觉技术,根据目标在图像信息中所反映出的特征,把不同类别的目标识别、区分开来的图像处理办法。人的视觉系统可以快速、准确地处理大量视觉信息,相比于传统的计算机视觉办法,可以更好地完成目标识别任务。视觉系统对目标形状的信息处理是从视网膜开始,经过侧膝体(LGN)、V1、V2、V4、IT 等层次的过程^[1]。受初级视觉皮层 V1 启发,余鹏等^[2]提出了一种前馈深度层次模型。该模型利用稀疏连接模拟感受野特性,结合方向图实现对目标的抽象表示。文献[3]基于 V1 简单细胞特性与极大化稀疏原理,提出了快速收敛的独立成分分析(independent component analysis, ICA)模型,该模型较传统的 ICA 模型收敛速度更快,但是受噪声影响较大。最具代表性的模

型是 Poggio 等^[4]提出的分层最大化(HMAX)模型。该模型的简单细胞层执行“线性和”操作,复杂细胞层执行“最大化”操作,从而实现位置、尺度等的不变性。该模型与 V1 实现了较好对应,而对更高级皮层的模拟比较粗糙。以上模型主要停留在 V1,忽略了较高级皮层神经机制对目标表达的重要作用^[5]。近年来,涌现出了许多结合高级皮层进行模拟的层次模型:Kavukcuoglu 等^[6]在 ICA 基础上,把非线性变换引入稀疏编码模型,模拟 V1 复杂细胞特性,但是在特征提取部分需要执行迭代算法,计算量较大;Yang 等^[7]基于尺度不变特征变换(SIFT)特征提出了一种监督的稀疏编码模型,该模型通过 max pooling 操作实现对 V1 复杂细胞不变性的模拟,但此模型使用的基于反投影的随机优化算法计算速度较慢,并且 max pooling 操作会丢失较多信息;Kim 等^[8]提出了一种模拟 V1、V4 的形状编码方

① 973 计划(2012CB719903),国家自然科学基金委创新研究群体(X198144),国家自然科学基金青年科学基金(41101386)和国家自然科学基金(41071256)资助项目。

② 男,1990 年生,硕士;研究方向:计算机视觉,数字图像处理;E-mail:sheng_bo_cool@163.com

③ 通讯作者,E-mail:tfang@sjtu.edu.cn

(收稿日期:2015-10-28)

法,该方法结合视觉区域检测和形状编码提取目标单一尺度下的曲率方向特征并初步应用在目标识别中,但是此模型忽略了 V2 区对角形特征提取的重要作用^[9];Tschechne 等^[10]受神经元间存在反馈连接的启发,提出了对表面和目标边界编码的循环计算网络模型,此模型增加了对 V2 区模拟并应用于目标边界提取。

形状特征是用于描述目标的基本特征,常见的有基于轮廓和基于区域的形状描述^[11]。其中,基于轮廓的描述应用更为广泛,原因在于视觉系统通常是基于轮廓区分形状。此外,曲率作为判别形状间相似性的重要轮廓特征^[12],可在视皮层中的 V4 区进行感知^[5]。但是,基于轮廓的形状描述也存在一些限制,比如轮廓受噪声影响较大,并且目标轮廓常常难以提取。不同于基于轮廓或区域的描述方法,

层次性的形状描述可以逐层地过滤掉不匹配的图像特征^[11]。基于此,本文提出了一种层次形状特征提取模型(V1-V2-V4)以弥补 V1 条形特征对形状描述的不充分。该模型与上述文献的模型不同:首先是结合噪声抑制与 V2 角形特征,通过增强角形特征点如角点位置的提取,进而将其应用于目标与场景分类中;其次是对 V4 层引入曲率计算以描述目标轮廓的形变程度,并对 V4 输出提取融合曲率及梯度方向的直方图特征以弥补单一特征对目标或场景分类的不足。

1 模型框架

模型共三层,分别为 V1 层、V2 层、V4 层,层与层之间采用前馈连接,如图 1 所示。

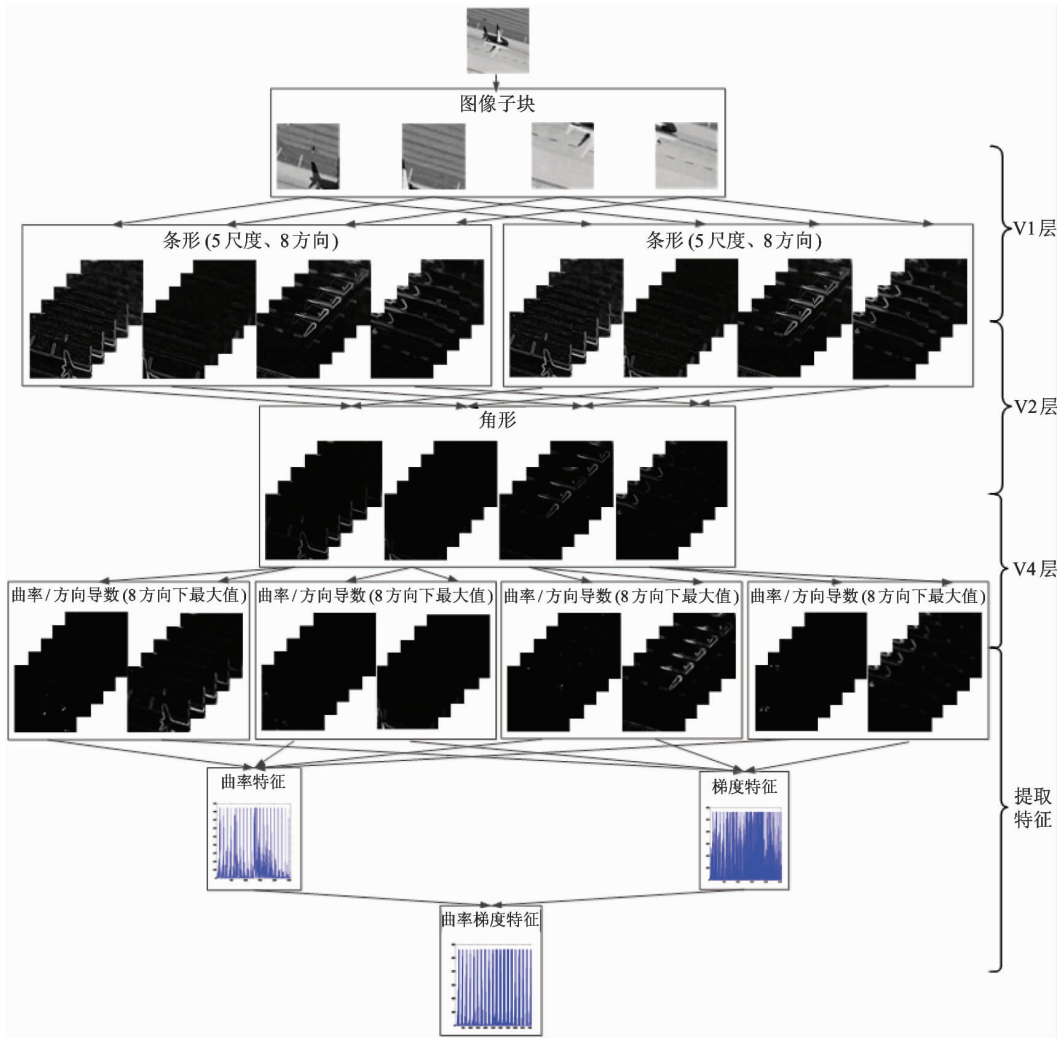


图 1 层次形状特征提取模型

对每个图像子块,V1 层使用 Gabor 滤波器组在 5 尺度 8 方向下提取两组条形特征,每组条形特征的滤波方向相差 90 度。V2 层在每个尺度下对 V1 输出的两组条形特征进行组合,实现对角形特征的提取。V4 层计算不同方向下 V2 输出的曲率与方

向导数。特征提取阶段则是累积每个尺度的不同方向下 V4 输出的曲率与方向导数的最大值作为目标的曲率特征与梯度特征,简称为“曲率 + 梯度”特征。

图 2 为各层输出效果图。

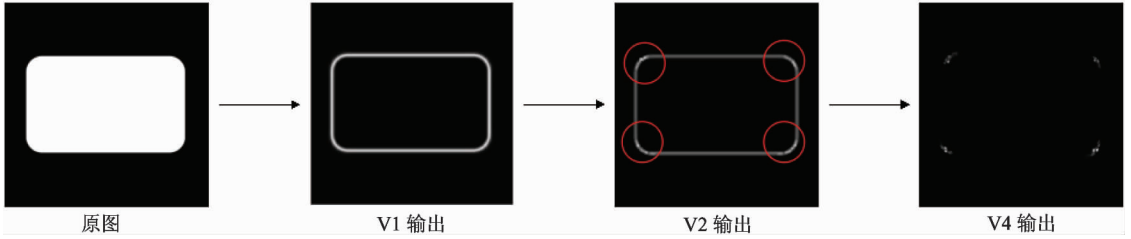


图 2 各层输出效果图

1.1 V1 层

使用 Gabor 函数模拟 V1 层简单细胞对条形刺激的响应作用^[13]。二维 Gabor 函数的表达式为

$$g_{\lambda, \sigma_1, \theta, \varphi}(x, y) = \exp\left(-\frac{\tilde{x}^2 + \tilde{y}^2}{2\sigma_1^2}\right) \cdot \cos\left(2\pi \frac{\tilde{x}}{\lambda} + \varphi\right) \quad (1)$$

$$\begin{cases} \tilde{x} = x \cos\theta + y \sin\theta \\ \tilde{y} = y \cos\theta - x \sin\theta \end{cases} \quad (2)$$

其中尺度参数 σ_1 决定 V1 层简单细胞感受野的大小,角度参数 θ 模拟神经元的最优响应方向。参数 r 决定神经元感受野形状,参数 φ 决定偶数的奇偶性, $1/\lambda$ 表示余弦因子频率设输入图像为 $f(x, y)$, 则模拟简单细胞响应结果为

$$R_{\lambda, \sigma_1, \theta, \varphi}(x, y) = (f \times g_{\lambda, \sigma_1, \theta, \varphi})(x, y) \quad (3)$$

使用 Gabor 能量算子模拟复杂细胞的响应结果:

$$E_{\lambda, \sigma_1, \theta}(x, y) = \sqrt{R_{\lambda, \sigma_1, \theta, 0}^2(x, y) + R_{\lambda, \sigma_1, \theta, -\frac{\pi}{2}}^2(x, y)} \quad (4)$$

使用 N_θ 个不同方向 θ_i 下的 Gabor 能量的向量和来模拟 V1 层输出,将多个方向合并为 1 个主方向:

$$\vec{E}_{\theta_i}(x, y) = \sum_{i=1}^{N_\theta-1} E_{\lambda, \sigma_1, \theta_i}(x, y) \cdot (\cos\theta_i, \sin\theta_i)^T \quad (5)$$

其中 $\theta_i = \frac{(i-1)}{N_\theta}\pi$, $i = 1, 2, \dots, N_\theta - 1$ 。记 $\vec{E}_{\theta_i}(x, y)$ 在尺度 s 下的幅值为 $R_{\theta_i}^{V1}(x, y, s)$ 。

1.2 V2 层

模型 V2 层的目的是在获得条形特征的基础上

提取角形^[9]特征,进而有效增强目标在关键特征点(如角点)位置处的特征描述。

首先,本文使用轮廓提取算法的 3D-DOG (difference of Gaussian, 高斯差分)^[14]函数对 $R_{\theta_i}^{V1}(x, y, s)$ 进行抑制以消除图像噪声,用于克服基于轮廓的形状特征提取受噪声影响较大的弊端^[11]。

3D-DOG 与 $R_{\theta_i}^{V1}(x, y, s)$ 的卷积记为抑制因子 W :

$$W(x, y, s) = (R_{\theta_i}^{V1} \times DOG)(x, y, s) \quad (6)$$

$$R_{\theta_i}(x, y, s) = H(R_{\theta_i}^{V1}(x, y, s) - \alpha_1 W(x, y, s)) \quad (7)$$

$$H(x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases} \quad (8)$$

参数 α_1 控制抑制强度的大小。

其次,采用最优响应角度相差 $\pi/2$ 的 $R_{\theta_i}(x, y, s)$ 模拟 V2 层对角形特征的响应(图 3):

$$R^{V2} = R_{\theta_i} \cdot R_{\theta_i + \pi/2} \quad (9)$$

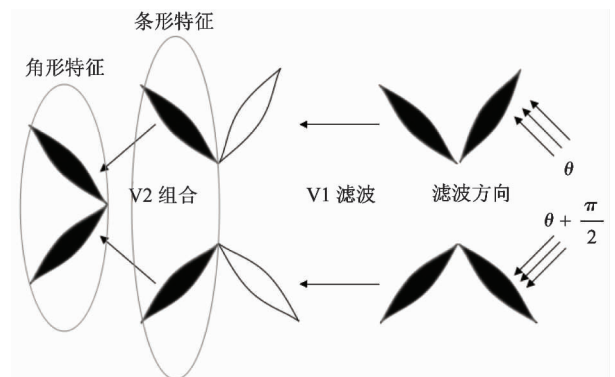


图 3 V2 层角形特征提取

其中, R_{θ_i} 与 $R_{\theta_i+\pi/2}$ 分别表示 V1 层不同方向下的条形特征。将对应像素点的灰度值相乘, 则结果只保留 θ_i 与 $(\theta_i + \pi/2)$ 方向组合的条形特征, 即角形特征 R^{V2} 。根据图 2 可以看出, V2 输出后四角位置处亮度相对更高, 角形部分得到加强。

此外, 由于单一尺度下会造成目标尺度信息的缺失, 因此, 与文献[10]不同的是, 本文将单一尺度下的角形特征提取扩展到多尺度, 该扩展方式有助于增强目标尺度变化的稳健性。

1.3 V4 层

V4 层的目的是在获得目标条形、角形特征基础上提取曲率^[5]方向特征。采用构建方向域与曲率域的方法^[8]模拟 V4。

曲率的计算方式如下: 首先借助梯度向量的能量方向构建方向域:

$$OF_{\text{flow}}(i, j) = \frac{1}{2} \tan^{-1} \left\{ \frac{2I_x I_y}{I_x^2 - I_y^2} \right\} + \frac{\pi}{2} \quad (10)$$

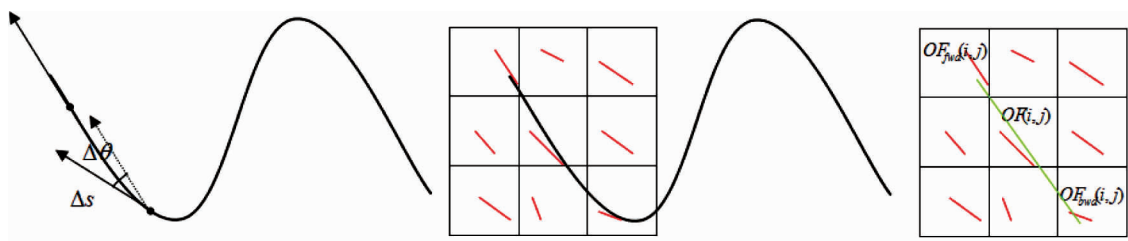


图 4 135°方向的曲率信息近似计算

1.4 特征提取

对于每个分解后的某一尺度下的图像子块, 采用一个 4×4 的网格进行进一步分解。首先比较每个亚子块内每个点在 $0^\circ, 45^\circ, 90^\circ, 135^\circ$ 方向的 $R_{CF}^{V4}(\theta_j)$, 在这 4 个方向累积 $R_{CF}^{V4}(\theta_j)$ 的最大值, 提取基于曲率的 64 维特征向量。为避免图像尺度信息的丢失, 在文献[8]的基础上, 我们进行了多尺度、多方向的扩展: 根据实际情况, 如需要更为精细的图像特征, 可根据每个点的 I_x 正负进一步将 4 个方向拓展为 8 个方向: $0^\circ/180^\circ, 45^\circ/225^\circ, 90^\circ/270^\circ, 135^\circ/315^\circ$, 进而提取基于曲率的 128 维特征向量。接着根据上一步选择的不同方向数, 使用相似方式提取基于方向导数的 64/128 维特征向量, 简称为梯度特征。综上, 对每个图像子块即可提取出一个 128/256 维的“曲率+梯度”的特征向量。

其中 I_x, I_y 表示对 $R^{V2}(i, j)$ 分别求 x, y 方向偏导。则推导近似曲率 (k_{sel}) 如图 4 所示:

$$k_{\text{sel}}(i, j) = \frac{\Delta\theta}{\Delta S} = \frac{\|OF_{\text{fwd}}(i, j) - OF_{\text{bwd}}(i, j)\|}{2} \quad (11)$$

通过添加梯度幅值 ($M_{\text{fwd}}, M_{\text{bwd}}$) 修正上述方程, 得到在 (i, j) 处曲率域表达式:

$$CF(i, j) = (1 - \cos(k_{\text{sel}}(i, j))) \cdot M_{\text{fwd}} \cdot M_{\text{bwd}} \quad (12)$$

方向导数的计算方式如下:

$R^{V2}(i, j)$ 在点 (x, y) 处沿方向 $l = (\cos\theta, \sin\theta)$ 的方向导数为

$$\frac{\partial R^{V2}(i, j)}{\partial l} = I_x \cos\theta + I_y \sin\theta \quad (13)$$

记 V4 输出的 θ_j 方向的曲率为 $R_{CF}^{V4}(\theta_j)$, 方向导数为 $R_{CF}^{V4}(\theta_j)$ 。从图 2 中 V4 输出效果图可以看出, 目标的高曲率部分被保留, 低曲率部分被过滤掉。

图 5 给出了特征提取的示意图: 计算每个点在四个方向下的曲率/梯度幅值, 累积最大值。

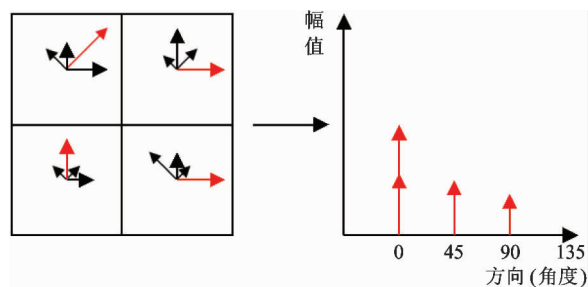


图 5 特征提取

2 实验及分析

为验证本文模型的有效性, 实验采用两种类型的数据集 (MNIST 手写数字识别^[15] 和 21 类遥感影

像分类^[16])。实验环境如下:CPU: Intel Xeon E5-2630 2.40GHz;内存:64.0GB;系统类型:64 位 Windows 7 企业版;开发平台:MATLAB R2013a。

2.1 MNIST 手写数字识别

MNIST 共 10 类(28 × 28 像素),如图 6 所示。其中训练样本共 60000 幅,测试样本共 10000 幅。

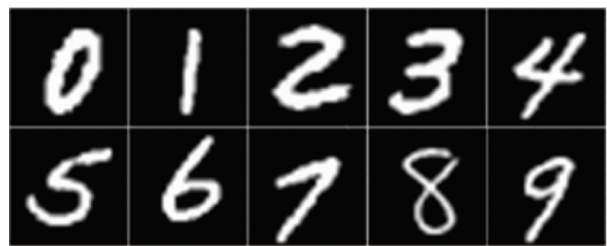


图 6 MNIST 数据集样图

实验中,首先将每幅图像分割成 2 × 2 个子块,使用最近邻插值算法,将每个图像子块大小缩放为 16 像素 × 16 像素。设定式(1)中 $\gamma = 0.3, \lambda = 8, \sigma_1 \in \{0.6, 1.7, 2.8, 3.9, 5.0\}$, 式(2)中 $\theta \in \{0, \pi/8, 2\pi/8, 3\pi/8, 4\pi/8, 5\pi/8, 6\pi/8, 7\pi/8\}$, 式(5)中 $N_\theta = 8$, 式(7)中 $\alpha_1 = 1$ 。

使用 1.4 节的方式提取特征,每个亚子块提取 8 方向 16 维的“曲率 + 梯度”特征,则每个子块提取 $16 \times (4 \times 4) = 256$ 维特征。共 5 个尺度、2 × 2 个子块,每幅图像提取的特征维数为 $256 \times 5 \times (2 \times 2) = 5120$ 维。最终采用支持向量机(SVM)进行训练与测试。表 1 给出了本文方法识别精度的混淆矩阵,表 2 示出了本文模型与其它基于生物视觉启发方法的识别精度,表 3 给出了方向数分别为 4 和 8 时单独的层次性梯度、曲率特征及“曲率 + 梯度”特征的识别精度。

表 1 MNIST 数字识别混淆矩阵

类别 \ 预测结果										
	0	1	2	3	4	5	6	7	8	9
0	975	0	0	0	0	0	1	1	3	0
1	0	1132	1	0	0	0	1	0	1	0
2	1	1	1019	0	1	0	0	6	4	0
3	0	0	2	1002	0	2	0	2	2	0
4	0	0	0	0	976	0	0	0	1	5
5	1	0	0	5	0	882	2	0	1	1
6	1	2	1	1	0	1	949	0	3	0
7	0	3	4	3	2	0	0	1012	1	3
8	1	0	1	3	1	1	0	3	961	3
9	1	3	0	0	5	4	0	6	4	986

表 2 不同方法在 MNIST 上的精度

方法	错误率%
HMAX 改进 ^[17]	3.5
ICA 直方图 ^[3]	1.28
IPSD ^[6]	1.0
SIFT 稀疏编码 ^[7]	0.84
本文方法	1.06

表 3 两种方向下不同特征的精度

特征 \ 方向数			
	梯度(%)	曲率(%)	曲率 + 梯度(%)
4	84.80	78.92	97.47
8	87.14	82.12	98.94

表 1 给出了 10 个测试集的识别情况,表中行代表各个类别(0 ~ 9),列表代表预测结果。对角线上的表示正确分类的样本数,非对角线表示错分样本。

总体上看,实验取得了理想的识别精度。具体而言,“0”、“1”等类别识别效果相对更好,“0”只有 5 个样本分错,“1”只有 3 个样本分错。因为这些类别在形状上区分度较高,可以更准确地使用形状特征进行描述;而“9”的识别效果相对较差,共有 23 个样本分错。其中有 5 个样本分成了“4”,有 6 个样本分成了“7”,因为在形状上“9”与“4”、“7”等类别具有相似性,因此用形状特征区分其差别具有一定难度。

根据表 2,本文模型比 HMAX 改进模型识别精度高 2.44%,因为 HMAX 在提取条形特征后缺少对更复杂形状特征的描述,因而建立的特征词典不够完备;比 ICA 直方图方法高 0.22%,因为 ICA 方法容易受噪声影响;与 IPSD 方法精度相差 0.06%,因为 IPSD 在 ICA 的特征提取等部分进行了有效优化;SIFT 稀疏编码精度最高,因为 SIFT 特征本身实现了较好的局部不变性,可以有效区分目标的形状特征。

表 3 验证了本文模型在方向拓展、特征融合上的有效性。4 方向拓展为 8 方向后,梯度特征识别精度提高了 2.34%,曲率特征精度提高了 3.2%,“曲率+梯度”特征提高了 1.47%。特征融合后,4 方向下“曲率+梯度”特征分别比单独的梯度特征、单独的曲率特征识别精度高 12.67%、18.55%;8 方向下高 11.8%、16.82%。

2.2 21 类遥感影像分类

21 类遥感影像共 21 类,每类包含 100 幅图像(256 像素×256 像素)。数据集包含了较为复杂的地物结构,类间具有形状上的相似性,并且图像的空间结构关系较为复杂,存在尺度、旋转变化等因素影响,如图 7 所示。



图 7 21 类遥感影像部分样图

实验中,模型参数设置与 2.1 部分大致相同,区别在于:此处在 3 个尺度下($\sigma_1 \in \{0.6, 1.8, 3\}$)使用滑动窗口的方式提取特征,每个尺度下的滤波结果对应的滑动窗口数与滑动步长分别选取为 32/8,36/10,40/18,则每幅图像滑动的窗口数为 $29 \times 29 + 23 \times 23 + 13 \times 13 = 1539$ 。对每个窗口分为 4×4 的亚子块,每个亚子块提取 4 方向的 8 维曲率梯度特征,则每个窗口提取 $8 \times (4 \times 4) = 128$ 维特征,则对每幅图像提取 $128 \times 1539 = 196992$ 维特征。

此外,使用 vlfeat^[18] 工具箱提供的 VLAD encoding 方法对特征进行聚集产生更为紧凑的全局描述子,并采用 SVM 分类器进行训练与测试。实验中,每类随机选取 50 幅图像作为训练样本,其余 50 幅作为测试样本,重复 10 次进行交叉验证,最后将 10 次分类结果求均值。

表 4 比较了其它基于直方图特征的提取方法在 21 类遥感影像数据集上的分类结果。其中 SCK 主要应用了 SIFT、空间共生核(Co-occurrence Kernel)方法,SPCK 在 SCK 方法基础上增加了空间金字塔共生核方法,BOVW + SCK 是在 SCK 方法基础上增加了词包(bag of visual words, BOVW),SPMK 主要应用了 SIFT 和空间金字塔匹配核方法。本文模型比 SPCK 方法高 4.87%,比 SPMK 方法高 2.72%,比 SCK 方法高 2.49%,比 BOVW + SCK 方法高 0.3%。

表 4 不同方法在 21 类遥感影像上的精度

方法	分类精度(%)
SPCK ^[19]	73.14
SPMK ^[20]	75.29
SCK ^[16]	75.52
BOVW + SCK ^[16]	77.71
本文方法	78.01

此外,由于该数据集中目标具有较为复杂的背景与噪声信息,适合验证层次模型中每层的有效性。因此,我们分别比较了无 V1、V2 层(即直接提取曲率特征)、无 V2 层的“曲率+梯度”特征、单独的层次性曲率特征、梯度特征与层次性的“曲率+梯度”特征的分类结果,如图 8 所示。结果表明,4 种层次性特征分类结果均明显高于直接提取的曲率特征,

平均差距最高达到 25% 左右,表明层次性形状特征对于处理具有复杂背景与噪声信息的目标更加有效。相比于无 V2 层“曲率+梯度”特征,添加 V2 层后可以将分类精度提高 1% 左右,表明在条形特征基础上引入多尺度角形特征更为有效。同时“曲率+梯度”特征分类精度比单独的曲率特征高 16% 左右,比单独的梯度特征高 3% 左右,再次验证了特征融合的有效性。

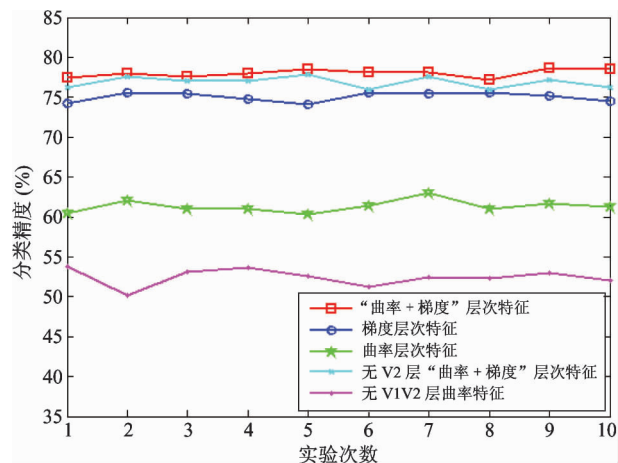


图8 层次性特征有效性验证

3 结 论

本文建立了一种层次模型(V1-V2-V4),用于提取目标的条形、角形、曲率特征。其中 V2 层结合了 3D-DOG 抑制图像噪声,使用滤波方向相差 90 度的条形特征提取多尺度角形特征;V4 通过曲率域提取目标轮廓形变程度较大的部分,最终提取融合曲率与梯度方向的直方图特征。实验表明,相比于条形特征,引入的角形、曲率特征可以更充分地描述目标形状,层次性的形状特征在目标识别上可以取得更好效果;曲率与梯度的融合特征可以有效弥补单一曲率或梯度特征局部描述的不足,并且特征方向的拓展可以进一步提高目标识别效果。未来的工作将继续研究更高级皮层(如 IT)对形状编码的机制,完善模型的层次结构,进一步提取目标更为完善的形状特征。

参考文献

- [1] Rousselet G A, Thorpe S J, Fabre-Thorpe M. How parallel is visual processing in the ventral pathway. *Trends in Cognitive Sciences*, 2004, 8(8): 363-370
- [2] 余鹏, 万里红, 霍宏等. 基于层次特征映射模型的目标识别. *高技术通讯*, 2014, 24(4): 414-419
- [3] Hyvärinen A. Fast and robust fixed-point algorithms for independent component analysis. *Neural Networks, IEEE Transactions on*, 1999, 10(3): 626-634
- [4] Riesenhuber M, Poggio T. Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 1999, 2(11): 1019-1025
- [5] Connor C E, Brincat S L, Pasupathy A. Transformation of shape information in the ventral pathway. *Current Opinion in Neurobiology*, 2007, 17(2): 140-147
- [6] Kavukcuoglu K, Ranzato M A, Fergus R, et al. Learning invariant features through topographic filter maps. In: *Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR)*, Miami, America, 2009. 1605-1612
- [7] Yang J, Yu K, Huang T. Supervised translation-invariant sparse coding. In: *Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR)*, San Francisco, America, 2010. 3517-3524
- [8] Kim S, Kwon S, Kweon I S. A perceptual visual feature extraction method achieved by imitating V1 and V4 of the human visual system. *Cognitive Computation*, 2013, 5(4): 610-628
- [9] Boynton G M, Hegd  J. Visual cortex: The continuing puzzle of area V2. *Current Biology*, 2004, 14(13): R523-R524
- [10] Tschechne S, Neumann H. Hierarchical representation of shapes in visual cortex—from localized features to figural shape segregation. *Frontiers in Computational Neuroscience*, 2014, 8: 93
- [11] Zhang D, Lu G. Review of shape representation and description techniques. *Pattern Recognition*, 2004, 37(1): 1-19
- [12] Yang M, Kpalma K, Ronsin J. A survey of shape feature extraction techniques. *Pattern Recognition*, 2008, 41(1): 43-90

- [13] Park J G, Kim K J. Design of a visual perception model with edge-adaptive Gabor filter and support vector machine for traffic sign detection. *Expert Systems with Applications*, 2013, 40(9): 3679-3687
- [14] Deng S, Liu N, Huo H, et al. Contour detection based on multi-scale spatial inhibition and contextual modulation. In: Proceedings of the IEEE 2015 Seventh International Conference on Advanced Computational Intelligence (ICACI), Wuyi, China, 2015. 372-377
- [15] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 1998, 86(11): 2278-2324
- [16] Yang Y, Newsam S. Bag-of-visual-words and spatial extensions for land-use classification. In: Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM GIS), San Jose, USA, 2010. 270-279
- [17] Borji A, Hamidi M, Mahmoudi F. Robust handwritten character recognition with features inspired by visual ventral stream. *Neural Processing Letters*, 2008, 28(2): 97-111
- [18] Vedaldi A, Fulkerson B. VLFeat: An open and portable library of computer vision algorithms. In: Proceedings of the International Conference on Multimedia, Firenze, Italy, 2010. 1469-1472
- [19] Yang Y, Newsam S. Spatial pyramid co-occurrence for image classification. In: Proceedings of the International Conference on Computer Vision (ICCV), Barcelona, Spain, 2011. 1465-1472
- [20] Lazebnik S, Schmid C, Ponce J. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR), New York, America, 2006. 2169-2178

Image classification based on a hierarchical shape feature extraction model

Zhang Shengbo, Liu Na, Huo Hong, Fang Tao

(Key Laboratory of System Control and Information Processing, Ministry of Education,
Department of Automation, Shanghai Jiao Tong University, Shanghai 200240)

Abstract

The image classification based on hierarchical shape feature extraction was studied. Considering that the bar features extracted from the primary visual cortex of V1 can not sufficiently represent the shape of an object, a new hierarchical model of V1-V2-V4 was presented to further extract the features of angle and curvature. Under the model, the Gabor function is adopted to simulate V1 to extract the bar features, the noisy inhibition using 3D-DOG (difference of Gaussian) combined with a group of Gabor filters with the filtering direction difference of 90° , is introduced to extract the multi-scale angle features in V2, and in V4, the deformation of the shape of an object is described through the computation of the curvature field, to finally extract the histogram features fusing the curvature and gradient orientation. The advantage of the proposed model is that the hierarchical representation based on the computation of angle and curvature helps to extract the key points of object shape such as the corners. Moreover, the use of fusion features also remedies the inadequacy of the object description only using curvature or gradient feature. The experiments on MNIST handwritten digits and 21 remote sensing images demonstrated that the use of the features fusing curvature and gradient orientation achieved the digit recognition accuracy of 98.94%, and obtained the better result in classification of remote sensing images.

Key words: ventral pathway, shape feature, curvature, hierarchy, image classification