



一种基于分层复杂性度量的有效的形状图像检索方法

王斌^{1,2}

1. 南京财经大学信息工程学院, 南京 210023
 2. 南京财经大学电子商务省级重点实验室, 南京 210023
- E-mail: wangbin@njue.edu.cn

收稿日期: 2016-12-31; 接受日期: 2017-03-13; 网络出版日期: 2017-08-30

国家自然科学基金(批准号: 61372158)、江苏省自然科学基金(批准号: BK20141487)、江苏省“333工程”高层次人才工程(批准号: BRA2015351)、江苏省科技计划(产学研合作前瞻性联合研究)(批准号: BY2016009-03)和江苏高校优势学科建设工程(PA-PD)资助项目

摘要 本文提出了一种基于分层复杂性度量的形状描述方法(简称 HCMD), 以应用于一般的形状图像检索任务. 该方法属于基于区域的形状描述方法, 不需要专门提取形状的边界完成特征抽取任务. 本文的主要贡献是提出了一种分层的形状描述框架, 基于该框架, 形状沿各个方向被迭代地切割成小的区块, 每一层级的形状区块被施以各种几何特性的度量, 以刻画其形状复杂性. 这种分层抽取形状复杂性特征的描述机制使得 HCMD 具有由粗到细的形状描述能力, 能有效地表达形状内部的复杂结构特性. 基于 HCMD 的形状匹配, 独立于形状的旋转、缩放、平移和镜像变换, 而且计算简单, 是一种能有效处理轮廓线形状和区域形状识别的通用方法. 用 MPEG-7 的 CE-1 轮廓线形状图像库、CE-2 区域形状图像库和美国哥伦比亚图像库 COIL-20 这 3 个标准测试集对 HCMD 进行性能评估, 并与其他形状描述方法进行了广泛的比较, 包括 5 种基于区域的形状描述子、4 种基于点集的形状描述子和两种基于曲线的描述子. 实验结果表明 HCMD 方法在综合考虑检索率、计算效率和一般应用能力指标下, 其性能要优于参与比较的各类方法, 证明了该方法的有效性.

关键词 形状描述, 分层复杂性度量, 特征抽取, 目标识别, 图像检索

1 引言

在模式识别和图像分析中, 形状作为目标的一个非常重要的视觉属性, 为目标的分类和识别提供了非常重要的信息^[1,2]. 因此, 形状分析一直是计算机视觉领域一个非常活跃的研究方向, 并已在文本分析、神经科学、农业、生物医药和工程技术等领域产生了大量的应用^[3].

引用格式: 王斌. 一种基于分层复杂性度量的有效的形状图像检索方法. 中国科学: 信息科学, 2017, 47: 1674-1693, doi: 10.1360/N112016-00156
Wang B. Hierarchical complexity measures for effective shape-based image retrieval (in Chinese). Sci Sin Inform, 2017, 47: 1674-1693, doi: 10.1360/N112016-00156

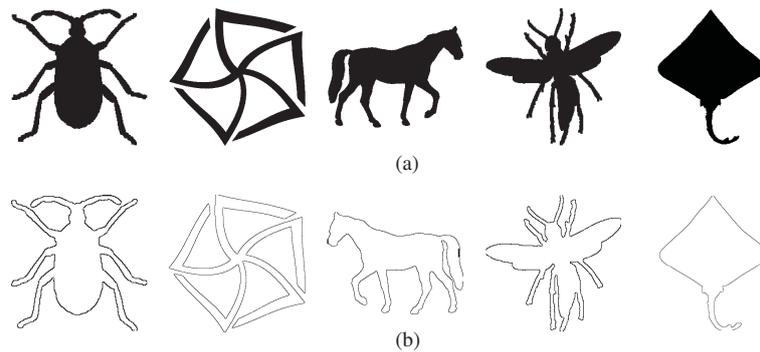


图 1 (a) 取自于 MPEG-7 CE-1 轮廓线形状库的 5 幅形状图像及 (b) 从中抽取的边界像素点

Figure 1 (a) Five samples from the MPEG-7 CE-1 contour shape database and (b) the edge points extracted from them

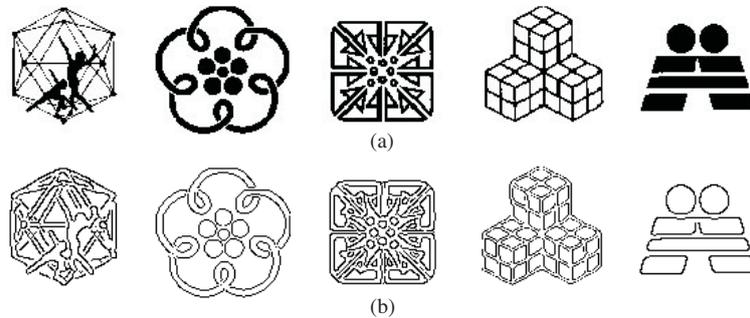


图 2 (a) 取自于 MPEG-7 CE-2 区域形状库的 5 幅形状图像及 (b) 从中抽取的边界像素点

Figure 2 (a) Five samples from the MPEG-7 CE-2 region shape database and (b) the edge points extracted from them

从一个场景中分割出来的目标形状一般地可以表示为一个分布在目标区域的所有像素点的集合, 如果可以将之简化为一条仅由边界像素点构成的曲线, 而且形状信息完全保持, 称该类形状为轮廓线形状, 否则, 该类形状为区域形状. 图 1 和 2 分别给出了 5 幅轮廓线形状图像和 5 幅区域形状图像的例子, 它们分别取之于 MPEG-7 的轮廓线形状图像库 (MPEG-7 CE-1) 和区域形状图像库 (MPEG-7 CE-2), 图中同时给出了从每一幅图像抽取的边界像素点图. 通过比较图 1 和 2 中的形状图像可以看出, 对于图 1 中的目标图像, 可以依据其边界像素点的邻接关系, 将其连接成一条简单的曲线 (无交叉), 而图 2 中的形状图像则有着复杂的结构, 从左往右的第 1, 2 和 4 幅图像都有着复杂的内部结构, 而第 3 和 5 幅图则由多个相互分开的目标区域构成. 显然, 对于图 2 中的形状图像, 将其边界像素点 (见图 2(b)) 连成一条曲线是不可能的. 轮廓线形状的本质是一维的, 可以用参数化曲线来表示, 而复杂的区域形状一般无法用参数化的曲线表示, 轮廓线形状可以看作是区域形状的特例.

形状描述是形状分析的一个非常重要的步骤, 旨在抽取目标形状的有效特征, 用于后续的目标形状的匹配、识别、分类和检索等形状分析任务^[3]. 根据形状特征抽取所考虑的像素点, 可以将其分为基于边界的形状描述方法和基于区域的形状描述方法两大类. 前者在抽取形状特征时, 仅考虑目标形状的边界像素点, 从边界像素点的几何关系中, 抽取形状特征. 而后者则将分布在整个目标形状区域的所有像素点作为特征抽取算法的输入, 产生形状描述子. 基于边界的形状描述方法需要预先抽取目标的边界像素点, 根据是否对边界像素点进行序化操作, 可以将该类方法进一步地分为基于曲线的描述方法和基于点集的描述方法. 这里序化操作是指, 从边界的某一点出发, 跟踪边界, 根据边界像素点

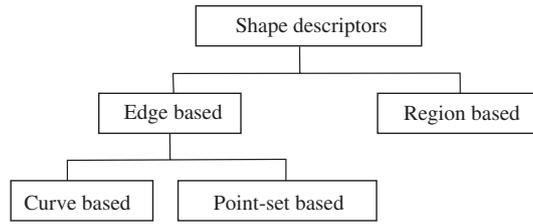


图 3 形状描述子的分类

Figure 3 Classification of shape descriptors

的邻接关系, 将边界的所有像素点串接在一起, 构成一个边界点的有序集. 而基于点集的描述方法, 把边界像素点看成一个一般的点的集合, 不需要进行序化操作. 图 3 给出的是形状描述子的分类示意图. 显然, 无论是基于区域的描述方法, 还是基于边界的描述方法, 都能处理图 1 所示的轮廓线形状, 而对于图 2 中区域形状, 由于无法对其边界像素点进行序化处理, 即无法表示成一条简单的曲线, 所以基于曲线的描述方法不能处理该类形状, 而基于点集的方法把边界当成一个一般的点集, 只要能得到边界像素点, 该算法就能工作, 所以该类方法可以处理区域形状. 而基于区域的描述方法在特征抽取过程中, 考虑形状区域的所有像素点, 且无需对它们做特别的预处理, 因此能有效处理图 2 中具有复杂内部结构的区域形状. 本文提出的方法正是属于基于区域的形状描述方法.

针对轮廓线形状, 因其可以被简化为一条曲线, 各种曲线分析方法可以应用于特征抽取, 由此产生了大量的基于曲线的形状描述方法. Fourier 描述子^[4,5]和曲率尺度空间^[6,7]是经典的轮廓线形状描述子, 后者被 MPEG-7 推荐为标准的轮廓线形状描述子, 其在 MPEG-7 CE-1 轮廓线形状测试集上的图像检索实验中, 获得了 75.44% 的检索精确率^[6]. 近些年来, 针对轮廓线形状又提出了形状树^[8]、多尺度凸凹性^[9]、内部距离^[10]、三角形面积^[11]、轮廓线柔韧性^[12]、分段多项式描述子^[13]、局部仿射不变描述子^[14], 以及一些其他的轮廓线形状描述子^[15~22]. 这些方法形式优美, 性能优越, 在 MPEG-7 CE-1 轮廓线形状测试集上报告的检索精确率大多超过了 85%. 但是这些方法是专门为轮廓线形状设计的, 在实际运用中, 需要首先获取形状图像的边界像素点, 并需要依据边界像素点的邻接关系, 将其表示成一个边界点的有序集^[4~7,9~12,15~19], 因此这类方法的性能严重依赖于形状图像轮廓线的提取质量, 从而不具稳定性. 为说明这一问题, 图 4 给出一个受椒盐噪声干扰的形状图像和从该图像提取的轮廓线的示例 (所用的提取轮廓线的程序代码来自于文献 [10] 的作者的学术网站¹⁾). 从该图可以看出, 由于受到噪声的干扰, 从该形状图像中抽取的轮廓线产生了大量的残缺, 即相当不完整. 因此, 提取的轮廓线的不稳定性, 导致了该类方法对形状描述的不准确. 另外, 如前所述, 基于曲线的描述方法不能处理具有复杂结构的区域形状, 因此该类方法不能用于一般的形状识别任务.

基于点集的形状描述方法近些年来也得到了一定的研究, 著名的方法是形状上下文方法 (Shape Contexts)^[23]. 该方法先将形状的边界点重新采样成指定个数的点 (一般 100 ~ 300 个点), 然后对于每一个采样点, 利用极坐标栅格, 统计其他的点相对于该点的空间分布. 该方法相比于曲线描述方法, 因为不需要预先对边界点进行序化处理, 所以更加稳定. 该方法在 MPEG-7 CE-1 轮廓线形状测试集上, 获得了 76.51% 的检索精确率. 距离集描述方法 (distance sets)^[24] 则对边界重采样得到的 N 个点中的每一个点, 计算其与最邻近的 n ($n \ll N$) 个点的距离, 构成一个距离集, 以描述该点与其邻近各点的空间关系. 由各点的距离集构成的集合, 构成了描述整个点集的空间安排. 该方法在 MPEG-7 CE-1 轮廓线形状测试集上, 报告的检索精确率为 78.38%. 我们在文献 [25] 中, 将目标边界点看成一个无序的点集,

1) http://www.dabi.temple.edu/~hbling/code_data.htm.

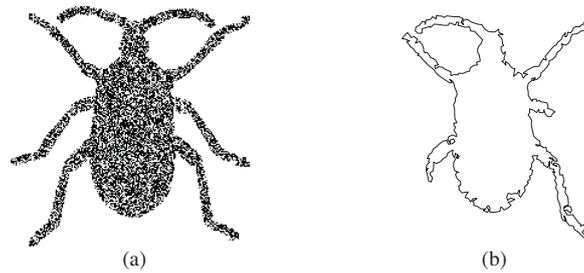


图 4 (a) 一个受椒盐噪声干扰的形状图像和 (b) 从中抽取的轮廓线 (运行文献 [10] 的作者所提供的程序代码)
 Figure 4 (a) A shape image distorted by salt pepper noise and (b) the contour extracted from it

提出了一种自组织的复杂网络动态演化模型,以产生目标形状描述子,该方法在 MPEG-7 CE-1 报告了 78.18% 的检索精确率.而在 MPEG-7 CE-1 轮廓线形状测试集上,报告的检索精确率超过了 85% 的基于点集的形状描述方法是 Peter 等^[26]提出的基于小波变换的描述子(文中称为形状 L'Âne Rouge).该方法的核心思想是将形状的点集描述为概率密度平方根的小波基展开.基于小波展开,密度表示为单位超球面上的点,密度之间的距离表示为对应的点沿球面到达对方点的最短弧长.但是这些方法计算非常耗时.形状上下文方法^[23]和距离集方法^[24]在形状匹配阶段,都需要计算点到点的优化匹配,该优化问题可归结为经典的二次指派问题,而求解该问题的计算复杂度为 $O(N^3)$ (采用匈牙利算法),这里变量 N 是点集的规模.文献 [24] 报告,执行两个形状(皆表示为 250 个采样点)的匹配,需耗时 0.7 s.因此,受限于匹配算法的计算复杂度,该类方法处理的点集的规模不能太大(一般不超过 300 点).而对于具有复杂结构的区域形状,过少的边界采样点显然不足以表达其复杂的结构信息.所以,形状上下文方法^[23]和距离集方法^[24]虽然能处理区域形状,但受限于其计算效率.对于形状 L'Âne Rouge 方法^[26]也存在计算量大的问题.据该方法报告的结果,在形状描述阶段,对于每个形状进行小波密度估计计算,需耗时 2~3 min,而在形状匹配阶段,仅计算小波系数重排优化问题,就耗时近 5 s.所以,基于点集的方法存在计算效率不高的问题.

相对于现有的基于边界的描述方法,基于区域的形状描述子提出的还相对较少(代表性的方法将在第 2 节予以介绍),虽然这些方法都能处理轮廓线形状(当成一般的区域形状处理)和区域形状这两类形状,但是这些方法在应用于轮廓线形状的识别任务却并不理想,它们在 MPEG-7 CE-1 轮廓线形状测试集上报告的检索精确率不到 70%. MPEG-7 对适宜于形状检索的描述子的评判标准是:(1)好的检索精确率;(2)一般的应用能力;(3)描述的可靠性;(4)低的计算复杂度;(5)由粗到细的分层描述能力^[27].为克服现有的形状描述方法的局限性,设计一种能满足 MPEG-7 评估标准,适宜于一般形状图像检索任务的描述子,本文提出了一种基于分层复杂性度量的形状描述方法(hierarchical complexity measures descriptor),简称为 HCMD.该方法属于基于区域的形状描述子.HCMD 提出了一种旋转分层的形状描述框架,该框架迭代地将形状区域切割成小的形状区块,并通过对这些形状区块的复杂性度量,有效地描述了形状的复杂内部结构.该方法的突出优点在于其能有效完成两种不同类型的形状识别任务,而且具有高的识别准确率和较低计算开销.该方法在 MPEG-7 的标准的轮廓线形状 CE-1 和区域形状测试集 CE-2 的检索性能评估中,分别获得了 86.36% 和 94.84% 的检索精确率,在 CE-1 上,高出同类方法 17 个百分点,在 CE-2 则高出同类方法 13 个百分点,表明该方法的有效性.

2 相关工作

HCMD 方法属于基于区域的形状描述方法,因此与本文工作相关的是各种基于区域形状的描述

方法. 常用的区域形状描述子是几何不变矩^[28]、Zernike 矩^[29] 和伪 Zernike 矩^[30] 等各种基于矩的形状描述方法. 其中, Zernike 矩被 MPEG-7 采纳为标准的基于区域的形状描述子^[27], 并获得了大量的应用. Zernike 矩是用一个定义在单位圆内的完备正交集 (称为 Zernike 多项式集) 所产生的形状描述子. 用 Zernike 矩可以构造任意高阶的矩, 其中低阶矩描述目标的整体形状, 高阶矩描述目标形状的细节, 且 Zernike 矩的模满足旋转不变性, 使得其在目标识别应用中具有很好的效果. 但由于 Zernike 的核函数的计算较为复杂, 而且在计算 Zernike 矩之前, 需要先将原图像映射到单位圆, 使得该方法的计算效率不高. 近年来, 文献 [31] 提出了一种二维极谐波变换用于形状的描述, 相较于 Zernike 矩, 其核函数的计算非常简单, 且没有数字稳定性问题.

用 Fourier 变换抽取形状的特征是另一类常用的方法. 经典的 Fourier 描述子^[32] 是针对轮廓线形状提出的. 该类方法首先从目标轮廓线获取一维复信号或实信号, 然后用其一维 Fourier 变换系数的模构成描述形状的特征向量. 为使 Fourier 描述子能适合区域形状的识别问题, 文献 [33] 提出一种通用的 Fourier 描述子 (GFD). 使用该方法, 需要首先将图像进行预处理, 将原始图像变换为极坐标图像, 再对其进行二维 Fourier 变换, 用其变换系数的模作为描述子. GFD 满足对目标旋转、缩放和平移的不变性, 而且适合于一般形状图像的检索应用.

近些年来, 一类方法着眼于用 Radon 变换^[34] 对目标形状进行几何分析. Radon 变换的主要思想是对目标形状沿平面内所有可能的直线作线积分, 得到一个二维函数 $R(\lambda, \theta)$, 该函数称为目标形状图像的 Radon 变换, 这里 λ 和 θ 是直线的两个参数, 分别表示坐标原点到直线的欧式距离和直线的方向角. 但由于一个目标形状的 Radon 变换并不满足对旋转、缩放和平移的不变性, Radon 变换并不能直接用于目标形状的识别, 需要在 Radon 变换所得平面内进一步地抽取不变的形状特征. 沿着该方向, 文献 [35] 提出一种 R- 变换, 文献 [36] 提出一种泛化的 R- 变换, 文献 [37] 则提出一种组合的 Radon 特征, 以在 Radon 变换的基础上, 产生不变的形状描述子.

与本文想法最为接近的是各种基于形状区域分割的方法. 将形状区域分割成小的区块进行描述, 这种机制跟上述方法相比其优点是能有效地抽取形状的结构特征. 文献 [38] 提出一种递归的形状区域分割算法. 该算法通过若干次迭代将形状区域分割成越来越小的区块. 在每一个迭代步, 对当前的形状区块计算其质心, 再依据获得的质心将其分割成 4 个更小的区块. 汇集每一个迭代步产生的区块的质心坐标, 构成了描述形状的特征向量. 沿着这条技术路线, 近来, 文献 [39] 提出一种称为适应性分层密度直方图的区域形状描述方法. 该方法用每一个区块的密度特征代替质心坐标产生特征向量, 以刻画形状区域像素点的分布特性. 但是由于其形状的分割方向依赖于图像的坐标系统, 因此不满足旋转不变性. 此外, 由于在迭代的过程中, 分割块数呈指数级的增长, 使得计算的复杂度很高, 无法满足实时应用的需要.

3 分层复杂性度量描述子 (HCMD)

形状的一个重要几何特性是它的复杂性, 该特性反应的是目标形状的空间覆盖能力 (也称空间填充能力), 如圆度、方度和面积周长比等都是跟目标形状的复杂性有关的几何特征^[3]. 虽然该类特征在形状的认识、分类和检索中经常使用, 但由于其只对目标形状的整体作复杂性测量, 抽取的是目标形状的全局特征, 只能区分具有较大差异的目标形状, 因此在实际应用中, 该类特征一般不单独使用, 要和其他的形状描述子结合起来使用才能完成目标形状识别任务^[27]. 本节提出一种分层的形状复杂性描述框架, 通过迭代地对形状进行切割, 由粗到细地抽取形状的复杂性特征, 以描述形状的整体特征和局部细节.

3.1 分层的描述框架

一个目标形状 f 的图像可以表示为一个二维的二值函数:

$$f(x, y) = \begin{cases} 1, & \text{if } (x, y) \in D, \\ 0, & \text{otherwise,} \end{cases} \quad (1)$$

这里 x 和 y 分别是像素点的横、纵坐标, D 是形状 f 在图像中的分布区域. 形状 f 的面积 S 定义为

$$S = \iint_D f(x, y) dx dy. \quad (2)$$

形状 f 的质心 (\bar{x}_s, \bar{y}_s) 定义为

$$\bar{x}_s = \frac{1}{S} \iint_D x \cdot f(x, y) dx dy, \quad (3)$$

$$\bar{y}_s = \frac{1}{S} \iint_D y \cdot f(x, y) dx dy. \quad (4)$$

将形状 f 的质心移到坐标系的原点, 则形状 f 的数学表示变为

$$\bar{f}(x, y) = f(x - \bar{x}, y - \bar{y}). \quad (5)$$

再将形状 \bar{f} 沿逆时针旋转 θ 角后的形状的数学表示为

$$\bar{f}_\theta(x, y) = \bar{f}(x \cos \theta - y \sin \theta, y \cos \theta + x \sin \theta). \quad (6)$$

可以证明 \bar{f} 与 \bar{f}_θ 具有相同的质心 $(0, 0)$. 过该质心, 沿 y 轴方向对形状 \bar{f}_θ 进行切割所得到的左边的形状区块的数学表示为

$$\bar{f}_\theta^{(1)}(x, y) = \begin{cases} \bar{f}_\theta(x, y), & \text{if } x \leq 0, \\ 0, & \text{otherwise.} \end{cases} \quad (7)$$

记 $(\bar{x}_\theta^{(1)}, \bar{y}_\theta^{(1)})$ 为形状区块 $\bar{f}_\theta^{(1)}$ 的质心. 过该质心, 对形状区块 $\bar{f}_\theta^{(1)}$ 进行沿 y 轴方向的切割, 将切割得到的左边形状区块记为 $\bar{f}_\theta^{(2)}$. 如此迭代地对形状 $\bar{f}_\theta^{(2)}$ 继续切割下去, 将经第 k 次切割后得到的左边的形状区块记为 $\bar{f}_\theta^{(k)}$. 将 \bar{f}_θ 记为 $\bar{f}_\theta^{(0)}$, 则对形状区块 $\bar{f}_\theta^{(k)}$ 迭代的定义为

$$\bar{f}_\theta^{(k)}(x, y) = \begin{cases} \bar{f}_\theta^{(k-1)}(x, y), & \text{if } x \leq \bar{x}_\theta^{(k-1)}, \\ 0, & \text{otherwise,} \end{cases} \quad (8)$$

这里 $\bar{x}_\theta^{(k-1)}$ 表示形状区块 $\bar{f}_\theta^{(k-1)}$ 的质心的横坐标.

记 $\varphi_\theta^{(K)} = \{\bar{f}_\theta^{(0)}, \bar{f}_\theta^{(1)}, \dots, \bar{f}_\theta^{(K-1)}\}$, 则 $\varphi_\theta^{(K)}$ 构成了对形状 \bar{f} 沿 θ 方向的一个 K 层分割, 这里 $K \geq 1$ 是预先设定的一个整数, 表示要进行分割的层数. 让 θ 从弧度 0 变化到 2π , 则 $f^{(K)} = \{\varphi_\theta^{(K)} | \theta \in [0, 2\pi)\}$ 构成了对形状 f 的一个层级数为 K 的分层的描述框架. 图 5 给出了本文提出的分层描述框架的一个示例.

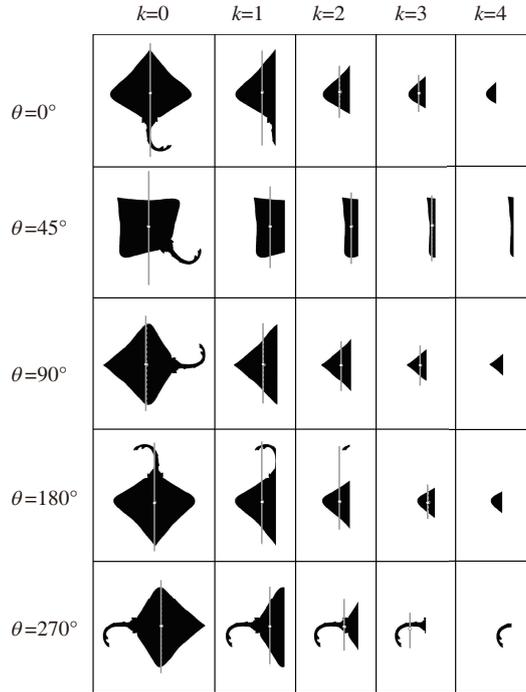


图 5 分层描述框架示例, 一个逆时针方向旋转了 θ 角的形状图像, 经过 4 次迭代切割, 得到了 5 个层级 ($k = 0, 1, 2, 3, 4$) 的形状区块 (灰线是切割线, 白点是当前形状区块的质心)

Figure 5 An example of hierarchical description framework. In the figure, a shape being rotated by angle θ counterclockwise is partitioned into five levels of blocks ($k = 0, 1, 2, 3, 4$) by four iterative cuts, where the cut line is marked by grey color and the centroid of the current region block is marked by white dot

3.2 形状区块的复杂性度量

本节对前述 K 层描述框架中的每一个形状区块的复杂性进行度量. 对形状区块 $\bar{f}_\theta^{(j)}$, 其像素点的 x 坐标构成的集合可表示为 $X_\theta^{(j)} = \{x | \bar{f}_\theta^{(j)}(x, y) = 1, x \in R, y \in R\}$, 像素点的 y 坐标构成的集合可表示为 $Y_\theta^{(j)} = \{y | \bar{f}_\theta^{(j)}(x, y) = 1, x \in R, y \in R\}$. 令 $x_L = \min(X_\theta^{(j)})$, $x_R = \max(X_\theta^{(j)})$, $y_L = \min(Y_\theta^{(j)})$, $y_R = \max(Y_\theta^{(j)})$. 则形状区块 $\bar{f}_\theta^{(j)}$ 的面积可用下式计算:

$$S_\theta^{(j)} = \int_{y_L}^{y_R} dy \int_{x_L}^{x_R} \bar{f}_\theta^{(j)}(x, y) dx. \quad (9)$$

对形状区块的 $\bar{f}_\theta^{(j)}$ 的质心的横坐标 $\bar{x}_\theta^{(j)}$ 可计算为

$$\bar{x}_\theta^{(j)} = \frac{\int_{y_L}^{y_R} dy \int_{x_L}^{x_R} x \cdot \bar{f}_\theta^{(j)}(x, y) dx}{S_\theta^{(j)}}. \quad (10)$$

对形状区块 $\bar{f}_\theta^{(j)}$ 定义一组公式来度量其复杂性. 下面两个公式是基于面积属性的复杂性度量:

$$\xi_\theta^{(j)} = \frac{S_\theta^{(j)}}{(x_R - x_L)(y_R - y_L)}, \quad (11)$$

$$\zeta_\theta^{(j)} = \frac{\int_{y_L}^{y_R} dy \int_{x_L}^{\bar{x}_\theta^{(j)}} \bar{f}_\theta^{(j)}(x, y) dx}{S_\theta^{(j)}}. \quad (12)$$

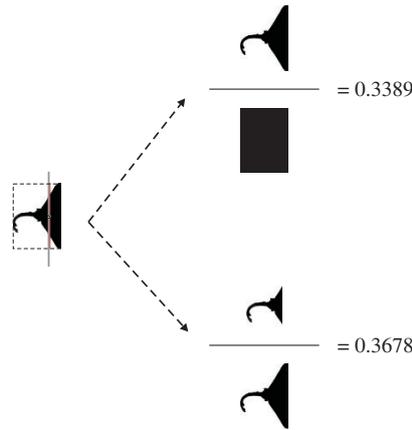


图 6 对形状区块进行复杂性度量的示意图. 左图: 被测量的形状区块 (灰线为过其质心的切割线, 虚线为其绑定矩形框); 右上图: 度量形状区块的方度 (式 (11)); 右下图: 度量形状区块的平衡性 (式 (12)), 图中的数值为度量值

Figure 6 An example of the complexity measures for a shape region block. Left figure: the measured shape region block (the grey line is the cut line crossing the centre of the shape region block and the rectangle shown in dash line is the bounding box of the shape region block; Right figure (upper): measuring the rectangle degree (Eq. (11)); Right figure (bottom): measuring the balance degree (Eq. (12)), the numbers (right-most) are the values of the complexities of the region blocks

由式 (11) 和 (12), 有 $0 < \xi_\theta^{(j)} \leq 1$ 和 $0 < \zeta_\theta^{(j)} \leq 1$. $\xi_\theta^{(j)}$ 度量的是形状区块 $\bar{f}_\theta^{(j)}$ 的方度, 若 $\xi_\theta^{(j)} = 1$, 则表示形状区块 $\bar{f}_\theta^{(j)}$ 是矩形的, 即其充满了由 4 个点 $(x_L, y_L), (x_L, y_R), (x_R, y_L), (x_R, y_R)$ 作为顶点确定的矩形区域. $\zeta_\theta^{(j)}$ 度量的是形状区块 $\bar{f}_\theta^{(j)}$ 的平衡性, 这里将形状区块的质心作为支点. 若 $\zeta_\theta^{(j)} = 0.5$, 则表明 $\bar{f}_\theta^{(j)}$ 在左右两边 (以直线 $x = \bar{x}_\theta^{(j)}$ 为界) 分布的面积相等, 此时形状区块处于平衡状态, 反之则处于非平衡态, 若 $\zeta_\theta^{(j)} > 0.5$, 则表明形状区块偏向左边, 而 $\zeta_\theta^{(j)} < 0.5$, 则表明形状区块偏向右边. 图 6 给出了用式 (11) 和 (12) 进行形状区块复杂性度量的一个示意图.

除了上述基于面积属性的复杂性度量, 为更精确地描述形状区块的几何特性, 还提出了用下面两个公式对形状区块进行基于长度属性的复杂性度量:

$$\eta_\theta^{(j)} = \frac{x_R - x_L}{y_R - y_L}, \tag{13}$$

和

$$\mu_\theta^{(j)} = \frac{x_R - \bar{x}_\theta^{(j)}}{\bar{x}_\theta^{(j)} - x_L}, \tag{14}$$

其中, 式 (13) 用来度量形状区块的长宽比, 而式 (14) 用来度量形状区块的质心相对于区块左右边界的距离比. 它们的值有可能大于 1, 为便于与前面两个公式组合使用, 3.3 小节将把它们归一到区间 $[0, 1]$.

3.3 形状签名的产生与不变性处理

3.2 小节对形状区块 $\bar{f}_\theta^{(j)}$ 给出了描述其几何特性的形状复杂性度量 $\xi_\theta^{(j)}, \zeta_\theta^{(j)}, \eta_\theta^{(j)}$ 和 $\mu_\theta^{(j)}$. 现固定 j , 让 θ 从弧度 0 变化到 2π , 则 $\xi_\theta^{(j)}, \zeta_\theta^{(j)}, \eta_\theta^{(j)}$ 和 $\mu_\theta^{(j)}$ 可分别看成变量 θ 的函数, 这里将它们分别改记为 $\xi_j(\theta), \zeta_j(\theta), \eta_j(\theta), \mu_j(\theta)$. 因此对于 K 层描述框架, 由 $j = 0, 1, \dots, K - 1$ 得到了 $4K$ 个一维函数. 从而对二维目标形状 f , 已构造了 $4K$ 个形状签名来描述其几何特性.

下面分析这些形状签名对目标的旋转、缩放、平移和镜像变换的敏感性. 因为在产生形状 f 的形状签名描述之前, 已将形状 f 的质心移到了坐标系的原点, 从而对其进行了位置归一, 因此抽取的形状签名都满足对目标形状的平移不变性. 而由形状复杂性度量 $\xi_\theta^{(j)}$ 和 $\zeta_\theta^{(j)}$ 的定义 (式 (11) 和 (12)), 两个公式计算的都是面积的比值, 因此当目标形状发生缩放变换时, 缩放因子在式 (11) 和 (12) 的计算中都会被消掉, 使得形状复杂性度量 $\xi_\theta^{(j)}$ 和 $\zeta_\theta^{(j)}$ 具有内在的缩放不变性. 同样, $\eta_\theta^{(j)}$ 和 $\mu_\theta^{(j)}$ (式 (13) 和 (14)) 计算的是长度的比值, 因此也具有内在的缩放不变性. 当目标形状发生旋转时, 如逆时针旋转 θ_0 , 形状签名 $\xi_j(\theta)$, $\zeta_j(\theta)$, $\eta_j(\theta)$ 和 $\mu_j(\theta)$ 将变为 $\xi_j(\theta + \theta_0)$, $\zeta_j(\theta + \theta_0)$, $\eta_j(\theta + \theta_0)$ 和 $\mu_j(\theta + \theta_0)$, 即它们的函数曲线都将在 θ 轴方向发生平移, 且平移量为 θ_0 . 该问题将在后续的形状匹配阶段通过循环移位匹配来消除该影响. 当形状发生了镜像变换, 形状签名 $\xi_j(\theta)$, $\zeta_j(\theta)$, $\eta_j(\theta)$ 和 $\mu_j(\theta)$ 将分别变为 $\xi_j(2\pi - \theta)$, $\zeta_j(2\pi - \theta)$, $\eta_j(2\pi - \theta)$ 和 $\mu_j(2\pi - \theta)$. 该问题也将留在后续的形状匹配阶段来消除. 此外, 为把函数的值域 $\eta_j(\theta)$ 和 $\mu_j(\theta)$ 归到区间 $(0, 1]$, 用它们各自的最大值对它们进行归一处理.

3.4 计算 HCMD 的算法及其时间复杂度

为有效地计算 HCMD, 设计了算法并给出其伪代码 (见算法 1), 该算法有两个参数, 一个是分层的层级数 K , 一个是对角度区间 $[0, 2\pi]$ 采样的频率 T . 现对整个算法进行时间复杂度分析. 对于输入的图像矩阵 $(f_{x,y})_{G \times H}$, 设形状在图像中占据的区域的像素点的个数为 N , 显然有 $N < G \cdot H$. 算法中最耗时的计算是从第 7 ~ 26 步的循环操作, 以及从第 27 ~ 30 步之间的循环操作.

对于前者, 是两层循环的嵌套, 其内层循环 (第 9 ~ 24 步) 每执行一次, 对当前的形状区块 D 进行复杂性度量, 并将当前形状区块 D 分成两块, 留下左边的那块作为下一次循环的当前区块, 其时间复杂度为 $O(S)$, 这里 S 是当前区块 D 中像素点的个数. 由于无法确定在每一次对当前形状区块的分割中, 有多少像素点分在了左边区块, 最坏情况是当前区块的所有像素点都分在了左边. 在这种情况下, 执行内层循环 (第 9 ~ 24 步) 的最坏时间复杂度为 $O(KN)$. 但每次都沿当前形状区块的质心进行切割, 因此这种情况一般不会发生. 由于每次切割的切割线都要过当前形状区块的质心, 可以认为当前区块的像素点落在切割线左右两边的概率相等, 所以执行内层循环 (第 9 ~ 24 步) 的平均时间复杂度为 $O(N + N/2 + \dots + N/2^{K-1}) = O(N)$. 执行第 25 步的时间复杂度为 $O(N)$, 所以执行外层循环 (第 7 ~ 26 步) 的平均时间复杂度为 $O((N + N)T) = O(NT)$, 最坏时间复杂度为 $O((KN + N)T) = O(KNT)$.

对于后者 (第 27 ~ 30 步), 执行第 28 和 29 步的时间复杂度都为 $O(T)$, 所以执行第 27 ~ 30 步的时间复杂度为 $O(KT)$. 从而可得出整个算法的平均时间复杂度为 $O(NT + KT)$, 因为 K 为给定的分层的层级数, 其值要远小于形状区域像素点个数 N , 所以整个算法的平均时间复杂度为 $O(NT)$, 最坏时间复杂度为 $O(KNT + KT) = O(KNT)$.

3.5 HCMD 描述子的距离度量方法

对于给定的两个形状 A 和 B , 用它们的描述子 HCMD, 来度量它们的形状差异度. 设 A 的形状签名为 $\{\xi_j^A(\theta), \zeta_j^A(\theta), \eta_j^A(\theta), \mu_j^A(\theta)\}_{j=0}^{K-1}$, B 的形状签名为 $\{\xi_j^B(\theta), \zeta_j^B(\theta), \eta_j^B(\theta), \mu_j^B(\theta)\}_{j=0}^{K-1}$. 而 A 形状的镜像变换版本 \bar{A} 的形状签名 $\{\xi_j^{\bar{A}}(\theta), \zeta_j^{\bar{A}}(\theta), \eta_j^{\bar{A}}(\theta), \mu_j^{\bar{A}}(\theta)\}_{j=0}^{K-1}$ 不需要重新计算, 因为 $\xi_j^{\bar{A}}(\theta) = \xi_j^A(2\pi - \theta)$, $\zeta_j^{\bar{A}}(\theta) = \zeta_j^A(2\pi - \theta)$, $\eta_j^{\bar{A}}(\theta) = \eta_j^A(2\pi - \theta)$ 和 $\mu_j^{\bar{A}}(\theta) = \mu_j^A(2\pi - \theta)$. 首先定义基于形状签名 ξ_j 和角度偏移量 $\Delta\theta \in [0, 2\pi)$ 的形状距离为

$$d_{\xi_j}^{\Delta\theta}(A, B) = \int_0^{2\pi} |\xi_j^A(\theta) - \xi_j^B(\theta + \Delta\theta)|_1 d\theta, \quad (15)$$

算法 1 Calculating hierarchical complexity measures descriptor (HCMD)**Input:**

$(f_{x,y})_{G \times H}$: the matrix of binary shape image;
 K : the number of hierarchical levels;
 T : the number of sampling angles from range $[0, 2\pi]$;

Output:

$(\xi_{k,t})_{K \times T}$, $(\zeta_{k,t})_{K \times T}$, $(\eta_{k,t})_{K \times T}$, $(\mu_{k,t})_{K \times T}$: HCMD of shape f ;

```

1: Initialize  $(\xi_{k,t})_{K \times T} \leftarrow 0$  and  $(\zeta_{k,t})_{K \times T} \leftarrow 0$ ;
2:  $B \leftarrow \{(x,y) | f_{x,y} = 1\}$ ; /* Put all the pixel points distributed in the shape region into set  $B$  */
3:  $S = |B|$ ; /* Calculate the area of the shape region */
4:  $\bar{x} \leftarrow \frac{1}{S} \sum_{(x,y) \in B} x$ ;  $\bar{y} \leftarrow \frac{1}{S} \sum_{(x,y) \in B} y$ ; /* Calculate the centroid of the shape */
5:  $\bar{B} \leftarrow \{(x - \bar{x}, y - \bar{y}) | (x,y) \in B\}$ ; /* Move the centroid of the shape to the origin of the coordinate system */
6:  $\theta \leftarrow 2\pi/T$ ;
7: for  $t = 1$  to  $T$  do
8:    $D \leftarrow \bar{B}$ ;
9:   for  $k = 1$  to  $K$  do
10:    if  $S > 0$  then
11:       $D_x \leftarrow \{x | (x,y) \in D\}$ ;  $D_y \leftarrow \{y | (x,y) \in D\}$ ;
12:       $x_r \leftarrow \max(D_x)$ ;  $x_l \leftarrow \min(D_x)$ ;  $y_r \leftarrow \max(D_y)$ ;  $y_l \leftarrow \min(D_y)$ ;
13:       $\xi_{k,t} \leftarrow S / ((x_r - x_l + 1)(y_r - y_l + 1))$ ; /* Measure the  $\xi$  complexity of the current shape block  $D$  using Eq. (11) */
14:       $x_c \leftarrow \frac{1}{S} \sum_{(x,y) \in D} x$ ; /* Calculate the x coordinate of the centroid of the current shape block  $D$  */
15:       $D_c \leftarrow \{(x,y) | x \leq x_c \wedge (x,y) \in D\}$ ; /* Partition the current shape block  $D$  into two parts along the line  $x = x_c$  and take the left one */
16:       $S_L = |D_c|$ ; /* Calculate the area of the shape block  $D_c$  */
17:       $\zeta_{k,t} \leftarrow S_L/S$ ; /* Measure the  $\zeta$  complexity of the shape block  $D$  using Eq. (12) */
18:       $\eta_{k,t} = (x_r - x_l) / (y_r - y_l + 1)$ ; /* Measure the  $\eta$  complexity of the shape block  $D$  using Eq. (13) */
19:       $\mu_{k,t} = (x_r - x_c) / (x_c - x_l + 1)$ ; /* Measure the  $\mu$  complexity of the shape block  $D$  using Eq. (14) */
20:       $D \leftarrow D_c$ ;  $S \leftarrow S_L$ ;
21:    else
22:      break;
23:    end if
24:  end for
25:   $\bar{B} \leftarrow \{(x \cos \theta - y \sin \theta, y \cos \theta + x \sin \theta) | (x,y) \in \bar{B}\}$ ; /* Rotate the shape by angle  $\theta$  */
26: end for
27: for  $k = 1$  to  $K$  do
28:    $\eta_{k,1:T} \leftarrow \eta_{k,1:T} / \max(\eta_{k,1:T})$ ; /* scaling the function values to  $(0, 1]$  */
29:    $\mu_{k,1:T} \leftarrow \mu_{k,1:T} / \max(\mu_{k,1:T})$ ; /* scaling the function values to  $(0, 1]$  */
30: end for
31: Return  $(\xi_{k,t})_{K \times T}$ ,  $(\zeta_{k,t})_{K \times T}$ ,  $(\eta_{k,t})_{K \times T}$  and  $(\mu_{k,t})_{K \times T}$ ;

```

这里 $|\cdot|_1$ 表示 $L1$ 距离. 同样定义形状距离 $d_{\zeta_j}^{\Delta\theta}(A, B)$, $d_{\eta_j}^{\Delta\theta}(A, B)$ 和 $d_{\mu_j}^{\Delta\theta}(A, B)$. 组合所有各类形状签名的距离度量则定义为

$$d(A, B) = \min_{\Delta\theta \in [0, 2\pi)} \left(\sum_{j=0}^{K-1} \left(w_1 \cdot d_{\xi_j}^{\Delta\theta}(A, B) + w_2 \cdot d_{\zeta_j}^{\Delta\theta}(A, B) + w_3 \cdot d_{\eta_j}^{\Delta\theta}(A, B) + w_4 \cdot d_{\mu_j}^{\Delta\theta}(A, B) \right) \right), \quad (16)$$

这里 w_1, w_2, w_3, w_4 是权重系数, 用来调节描述子 HCMD 中各类特征在形状差异度量中的贡献, 可以在实际应用中根据经验设定. 通过以上公式, 可以看出, 通过对形状签名在 θ 轴上的平移, 获得了不

依赖于旋转变换的距离度量, 即目标形状发生旋转, 式 (16) 的值不发生改变. 而形状 A 的镜像变换版本与形状 B 的距离 $d(\bar{A}, B)$ 也可以通过式 (16) 计算得到. 那么, 形状 A 和 B 的距离最终定义为

$$D(A, B) = \min(d(A, B), d(\bar{A}, B)). \quad (17)$$

通过式 (17), 进一步获得了不依赖于镜像变换的对形状 A 和 B 的距离度量. 形状 A 和 B 越相似, 则 $D(A, B)$ 的值越小, 而 $D(A, B)$ 的值越大则表明它们的差异越大.

下面分析计算形状距离的时间复杂度. 因为角度区间 $[0, 2\pi]$ 被采样成 T 个角, 则计算式 (15) 的时间复杂度为 $O(T)$, 所以计算式 (16) 的时间复杂度为 $O(KT^2)$, 因此得到计算式 (17) 的时间复杂度为 $O(KT^2)$, 即计算形状距离的时间复杂度为 $O(KT^2)$. 值得指出的是, K 为给定的分层描述的层级数, T 为角度采样频率, 它们都是独立于形状区域像素点个数的算法参数. 在实验中, 它们被设定为 $K = 5$ 和 $T = 32$. 而在 MPEG-7 CE-1 上报告的检索精确率为 87.70% 的基于曲线的形状描述方法——形状树描述子^[8]在形状匹配阶段的时间复杂度为 $O(nm^3)$, 这里 n 和 m 分别为两条曲线的采样点的个数. 而基于点集的形状描述方法——形状上下文描述子^[23]和距离集描述子^[24], 它们在形状匹配阶段的时间复杂度皆为 $O(n^3)$, 这里 n 为点集的规模. 由此, 可以看出, HCMD 方法在形状匹配阶段的计算时间复杂度要低于这些方法.

3.6 HCMD 性能的比较分析

至此, 已给出了描述子 HCMD 的定义、计算方法及距离度量, 下面进一步分析其性能. 由于在相关工作中, 文献 [39] 提出的适应性分层密度直方图方法 (adaptive hierarchical density histogram, AHDH) 与本文工作最为接近, 因此在性能分析中, 将 HCMD 与 AHDH^[39] 进行比较. 目标的旋转、缩放和平移并不改变其形状, 由前面的形状签名的产生与不变性处理和距离度量可知, HCMD 已满足对这些几何变换的不变性. 而 AHDH^[39] 方法则依赖于图像的坐标系统, 不满足旋转不变性.

HCMD 采用分层的描述框架, 在第 0 层 ($k = 0$), HCMD 描述的是整个形状的复杂性, 随着层级的增加, 分割出来的形状区块越来越小, 因而描述的形状信息越来越倾向于形状的细节, 表明 HCMD 具有由粗到细的形状描述能力. 值得指出的是, 虽然 AHDH 采用的也是一种分层描述框架, 但它是单方向的, 因此不能从多个方向上描述形状的结构特征. 而 HCMD 采用的是一种旋转的分层描述框架, 从 0° 到 360° 扫描形状, 沿各个方向分层地描述形状的几何特性, 因此 HCMD 比 AHDH 具有更强的形状识别能力, 后面的实验部分将进一步证明这一点.

下面再从时间复杂度上, 对 HCMD 和 AHDH 加以比较. AHDH 在每一次迭代中, 都将当前的形状区块分割成 4 块, 且对每一个分割出来的形状区块都保留到下一次迭代过程, 这样随着层级的增加, 分割出的形状区块呈指数级的增长, 如当层级 $k = 5$ 时, 对当前的区块进行分割, 得到的小的区块的个数将达到 $4^5 = 1024$. 而 HCMD 在每一次迭代中, 将当前的形状区块分成两块, 且仅保留左边的那块, 右边的那块将被舍弃. 因此, 分割出来并保留的形状区块在每一个层级都只有一块, 这样 HCMD 的算法复杂度要远低于 AHDH. 除了降低算法复杂度的考虑, 作这种设计的另外一个考虑是, 对于分割出来的左右两个形状区块, 左边的那块更加远离整个形状的质心, 而形状边界富含更多的形状信息, 能抽取到更有价值的形状识别特征.

HCMD 比 AHDH 更加紧致. HCMD 的特征个数为 $4KT$, 而 AHDH 的特征个数为 $2(4 + 4^1 + \dots + 4^K) = 8(4^K - 1)/3$. 由此可以看出, HCMD 方法抽取的特征个数仅随层级数线性增长, 而 AHDH 方法则随层级数指数级地增长.

此外, HCMD 跟其他的区域形状描述子一样具有对噪声干扰的鲁棒性, 这是因为 HCMD 描述子在特征抽取的过程中, 形状区域的所有像素点都参与了计算. 而基于边界的形状描述子在计算过程中, 只利用了形状边界像素点的信息, 因此对噪声干扰具有不稳定性, 特别是基于曲线的描述方法, 由于要对边界点进行序化操作, 而该操作会受到噪声的干扰, 该类方法对噪声非常敏感. 因此 HCMD 方法比基于边界的形状描述子更鲁棒. 后面的实验部分, 将进一步验证 HCMD 对噪声干扰的鲁棒性.

4 实验结果和讨论

为评估本文提出的形状描述子 HCMD 的性能, 将其应用于图像检索问题. 在实验中, 采用了 MPEG-7 的两个标准形状图像检索测试集: CE-1 轮廓线形状测试集和 CE-2 区域形状测试集, 对 HCMD 进行性能测试. 因提出的 HCMD 方法属于基于区域的形状描述方法, 所以在实验中, 选择的主要比较对象为基于区域的形状描述子, 它们是: Zernike 矩 (Zernike moments)^[29]、通用的 Fourier 描述子 (generic Fourier descriptors, GFD)^[33]、Radon 组合特征 (Radon composite features, RCF)^[37]、适应性分层密度直方图 (adaptive hierarchical density histogram, AHDH)^[39]、极谐波变换 (polar harmonic transforms, PHTs)^[31]. 其中 Zernike 矩是 MPEG-7 推荐的标准区域形状描述子^[27], 通用的 Fourier 描述子是在区域形状检索应用中性能表现优越的方法^[27,37], Radon 组合特征是近来提出的基于 Radon 变换的方法, 而适应性分层密度直方图是相关工作中与本文工作最为接近的方法. 除了与基于区域的形状描述子进行比较之外, 还选择了两种基于曲线的描述方法: 形状树 (Shape Tree)^[8] 和局部仿射不变描述子 (locally affine invariant descriptors)^[14], 以及 4 种基于点集的形状描述方法: 无序点集描述 (Unordered point-set description, UPSD)^[25]、形状上下文 (Shape Contexts)^[23]、距离集 (Distance Set)^[24] 和形状 L'Âne Rouge 方法^[26], 以进行更为广泛的比较. 在实验中, HCMD 的参数统一设置为: $K = 5$, $T = 32$, $w_1 = 0.18$, $w_2 = 0.59$, $w_3 = 0.06$, $w_4 = 0.18$. 实验平台为一台 CPU 为 Intel Core i7-4770 3.4 GHz, 内存 12 G, 操作系统为 Win 8 的台式计算机, 算法用 Matlab 编程实现.

4.1 MPEG-7 CE-1 轮廓线形状测试集

MPEG-7 CE-1 测试集共包含 1400 幅轮廓线形状图像, 它们被分成 70 种形状类型, 部分样本见图 7(a), 每类 20 个样本 (如图 7(b)). 性能度量是被广泛采用的 “bull's-eye test” 评估方法^[10~12, 15~19, 39]. 采用该方法, 测试集中的每一个样本作为一个查询样本, 跟测试集中的所有样本进行相似性比较, 然后根据形状的差异度排序, 返回差异度最小的前 $2 \times 20 = 40$ 幅图像, 再统计返回的 40 幅图像中有多少跟查询样本属于同一类, 设得到的数据为 r , 显然有 $r \leq 20$, 计算 $r/20$ 的值作为该查询样本的检索率, 将 1400 个检索样本的检索率的平均值作为整个测试的检索率 (bull's-eye test score).

表 1 给出了 HCMD 方法和参与比较的两种基于曲线的描述子、4 种基于点集的描述子和 5 种基于区域的描述子在 MPEG-7 CE-1 测试集上的检索率. 从表中的数据可以看出, 跟同类方法——基于区域的形状描述子 Zernike Moments^[29]、GFD^[33]、RCF^[37]、AHDH^[39] 和 PHTs^[31] 相比, HCMD 方法的检索精确率分别高出了 17.46%, 18.76%, 19.06%, 22.41% 和 15.44%. 跟基于点集的描述方法: 形状上下文^[23]、距离集^[24]、形状 L'Âne Rouge 方法^[26] 和无序点集描述子^[25] 相比, HCMD 方法则分别高出了 9.85%, 7.98%, 1.07% 和 8.18%. 而跟两种基于曲线的描述方法: 形状树^[8] 和局部仿射不变描述子^[14] 相比, HCMD 方法分别低了 1.34% 和 3.26%. 但是 HCMD 方法计算效率要远高于这两种基于曲线的描述子. 在该测试集上, 形状树方法^[8] 报告的结果是, 当曲线仅用 50 ~ 100 个采样

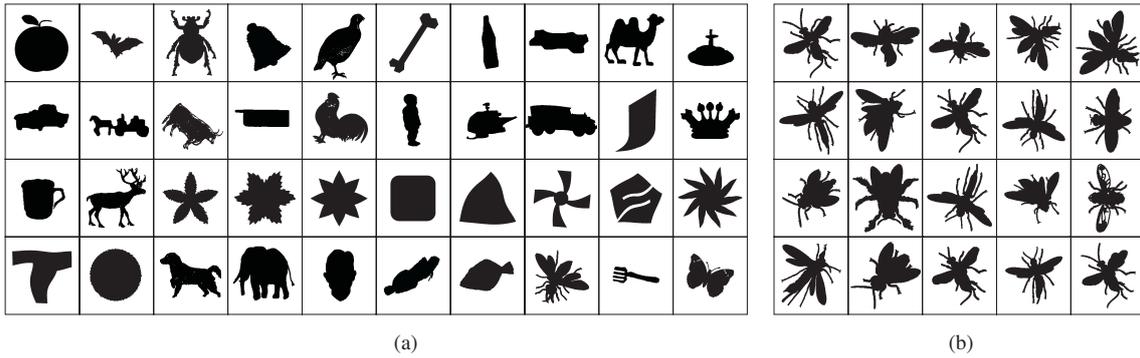


图 7 MPEG-7 CE-1 图像库

Figure 7 MPEG-7 CE-1 image dataset. (a) Samples from the dataset; (b) 20 samples from a class in the dataset

表 1 在 MPEG-7 CE-1 形状测试集上的检索率^{a)}

Table 1 Retrieval rates on the MPEG-7 CE-1 shape dataset^{a)}

	Algorithm	Bulls-eye test score (%)
Curve based	Shape Tree ^[8]	87.70 [†]
	Locally affine invariant descriptors ^[14]	89.62 [†]
Point-set based	Shape Contexts ^[23]	76.51 [†]
	Distance Set ^[24]	78.38 [†]
	Shape L'Âne Rouge ^[26]	85.25 [†]
	Unordered point-set description(UPSD) ^[25]	78.18
	Zernike Moments ^[29]	68.90
Region based	Generic Fourier descriptors (GFD) ^[33]	67.60
	Radon composite features (RCF) ^[37]	67.30
	Adaptive hierarchical density histogram (AHDH) ^[39]	63.95 [†]
	Polar harmonic transforms (PHTs) ^[31]	70.92
	The proposed HCMD	86.36

a) The scores marked by † are directly cited from the published results.

点, 匹配两个形状就耗时 5×10^{-1} s^[8] (算法运行在一台 3 GHz 的计算机上), 而 HCMD 方法匹配两个形状仅需 2.2×10^{-4} s, 比其快 2000 倍以上. 局部仿射不变描述子^[14] 报告在该测试集上完成全部检索, 实验耗时 1.78 h, 而 HCMD 方法完成全部检索实验仅耗时 4.2 min, 比之快 25 倍以上. 更为重要的是, 虽然这两种算法在 CE-1 测试集上分别报告了 87.70% 和 89.62% 的高检索精确率, 但报告这两个结果的算法需要利用形状图像边界点的序化信息, 因此并不能运行于后面的 MPEG-7 CE-2 形状测试集. 而对于在 MPEG-7 CE-1 测试集上报告了 85.25% 检索率的基于点集的形状 L'Âne Rouge 方法^[26], 该方法报告的匹配两个形状的时间达到了近 5 s^[26], 也远高于 HCMD 方法. 从 MPEG-7 CE-1 测试集的实验结果可以看出, 综合平衡检索率和计算效率两项指标, HCMD 方法的性能要好于参与比较的各类方法, 证明了其在轮廓线形状测试中的有效性.

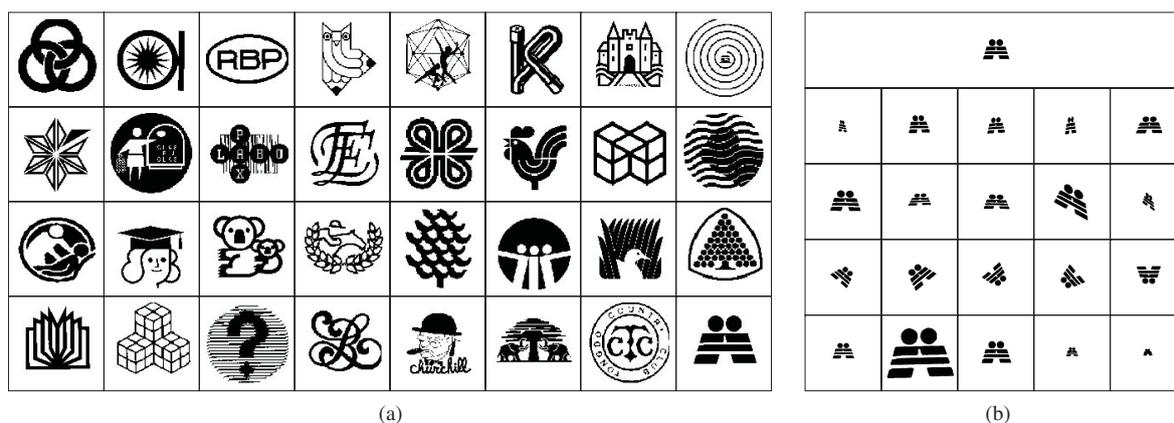


图 8 MPEG-7 CE-2 图像库

Figure 8 MPEG-7 CE-2 image dataset. (a) Samples from the dataset; (b) 21 samples from a class in the dataset, where the shape images shown from row 2 to row 5 are the rotated, scaled and perspective transformed versions of the sample shown in the first row, respectively

4.2 MPEG-7 CE-2 区域形状测试集

用于评估 HCMD 方法性能的第 2 组实验是 MPEG-7 CE-2 区域形状测试. 在 CE-2 测试集中, 总共有 3621 个区域形状图像 (部分样本见图 8(a)), 其中的 651 个样本被分成 31 组, 每组 21 个样本, 属于同一组的是同类形状 (图 8(b) 给出了其中一组的 21 个样本). 从图 8 可以看出, 许多形状具有非常复杂的内部结构, 而且一些形状是由彼此分开的几个形状区域组成, 即形状区域并不是一个联通的整体, 显然基于曲线的描述方法无法处理该类形状, 因为无法将边界点序化成一条曲线. 本组实验的性能度量仍然采用 “bull’s-eye test” 评估方法. 稍有不同的是, 用作检索样本不是整个测试集, 而是其中被分成 31 组的 651 个样本. 对于这 651 个样本, 将每一个作为检索样本跟测试集中的 3621 个样本进行形状相似性比较, 返回差异度最小的前 $2 \times 21 = 42$ 个样本, 再统计返回的 42 幅图像中, 与查询样本属于同一组的数目, 记为 r , 显然有 $r \leq 21$. 计算 $r/21$ 的值作为该查询样本的检索率, 将 651 个检索样本的检索率的平均值作为整个测试的检索率 (bull’s-eye test score).

因基于曲线的描述子无法处理 CE-2 形状, 所以在本组实验中, 仅用两种基于点集的方法 Shape Contexts [23] 和 UPSD [25], 5 种区域描述子 Zernike Moments [29], GFD [33], RCF [37], AHDH [39] 和 PHTs [31] 作为比较对象. 表 2 中给出了 HCMD 方法与这些方法在 MPEG-7 CE-2 形状测试集上的检索率. 需要说明的是, 在上组实验中, 对于两种基于点集的方法 Distance set [24] 和 Shape L’Âne Rouge [26], 直接引用了文献 [24, 26] 报告的结果. 由于报告这两种方法的文献没有具体给出算法实现的细节, 所以本组实验略去了跟这两种基于点集的方法的比较, 但仍保留了与 Shape Contexts [23] 和 UPSD [25] 两种基于点集描述方法的比较. 而本组实验则继续着重于与同类方法——基于区域的形状描述子的比较 (共有 5 种方法参与了比较).

从表 2 列出的实验结果可以看出, HCMD 方法在 CE-2 测试集上检索率比同类方法 Zernike Moments [29], GFD [33], RCF [37], AHDH [39] 和 PHTs [31] 的检索率分别高出了超过 13%. 而跟两种基于点集的描述子 Shape Contexts [23] 和 UPSD [25] 相比, 则分别高出了 24.42% 和 10.39%. 值得指出的是, 与本文的想法最为接近的是同样采用分层描述的 AHDH [39] 描述子, 但该方法在 CE-2 测试集上, 仅取得了 49.94% 的检索率, 其主要问题是仅从单个方向上描述形状特性, 该方法的描述和匹配依赖于目标形状的方向, 而 HCMD 方法采用了一种旋转分层的描述框架, 从多个方向上描述了形状的分层

表 2 在 MPEG-7 CE-2 形状测试集上的检索精确率
Table 2 Retrieval rates on MPEG-7 CE-2 shape dataset

	Algorithm	Bulls-eye test score (%)
Point-set based	Shape Contexts ^[23]	70.98
	Unordered point-set description (UPSD) ^[25]	84.13
Region based	Zernike Moments ^[29]	80.20
	Generic Fourier descriptors (GFD) ^[33]	81.20
	Radon composite features (RCF) ^[37]	67.40
	Adaptive hierarchical density histogram (AHDH) ^[39]	49.94
	Polar harmonic transforms (PHTs) ^[31]	64.13
	The proposed HCMD	94.52

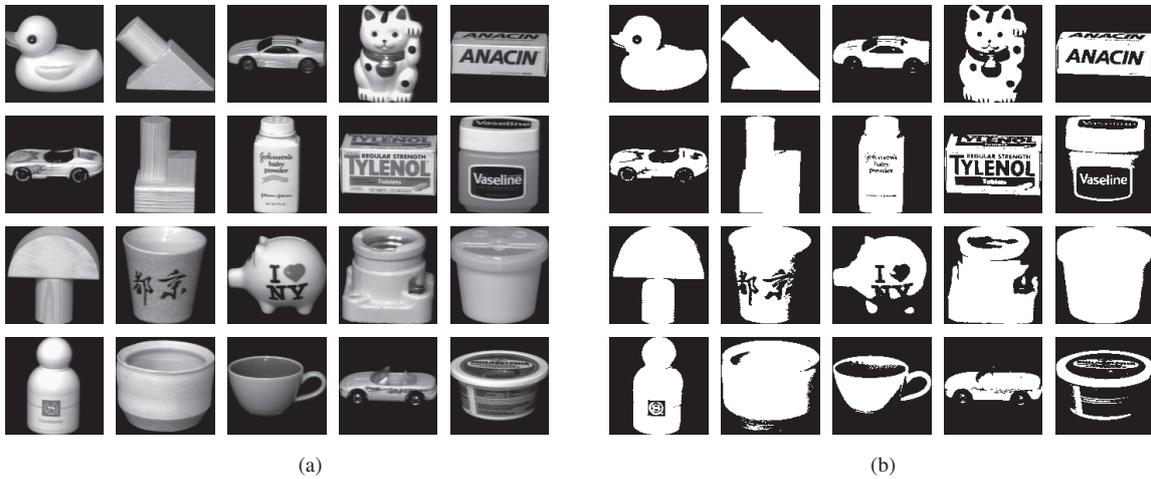


图 9 (a) COIL-20 图像库中的 20 种目标的灰度图像; (b) 图 (a) 中图像的二值化图像

Figure 9 (a) 20 object samples from the COIL-20 database; (b) the binary images derived from the images shown in (a)

复杂性, 且其形状匹配结果不依赖于形状的方向. 该组实验结果表明, HCMD 方法在区域形状测试集 MPEG-7 CE-2 的性能要优于参与比较的同类方法, 也好于参与比较的两种基于点集的形状描述子, 从而证明 HCMD 方法比其他方法更能有效地处理包含复杂内部结构的形状图像.

4.3 COIL-20 三维目标测试集

本文的第 3 组实验是评估算法 HCMD 在 COIL-20^[40] 三维目标测试集上的识别性能. 该测试集取自于美国哥伦比亚图像数据库²⁾, 它含有 20 种家用物品 (见图 9(a)) 的 1440 幅灰度图像 (每种目标 72 幅, 部分样本见图 10(a)). 它们的获取方式是: 将每个目标放置于一个转盘上, 且每转过 5° 角, 对该目标拍摄一幅图像, 转盘旋转一周, 得到该目标的 72 个视角的图像. 在实验中, 采用与文献 [23] 相同的目标识别性能评估方法, 对每一个目标的 72 幅视角图像, 等视角间隔的选取 M 幅图像作为训练样本, 剩下的 $72 - M$ 幅图像作为测试样本, 由此构成一个样本数为 $20 \times M$ 的训练集和一个样本数为

2) <http://www.cs.columbia.edu/CAVE/software/softlib/coil-20.php>.

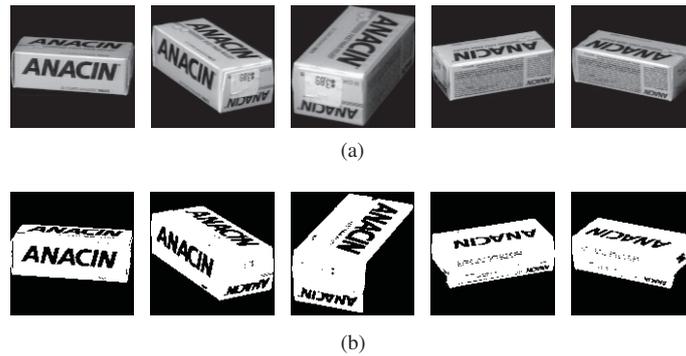


图 10 (a) COIL-20 图像库中一个目标的五个不同视角的图像; (b) 图 (a) 中图像的二值化图像

Figure 10 (a) Example views of one object from the COIL-20 database; (b) the binary images derived from the images shown in (a)

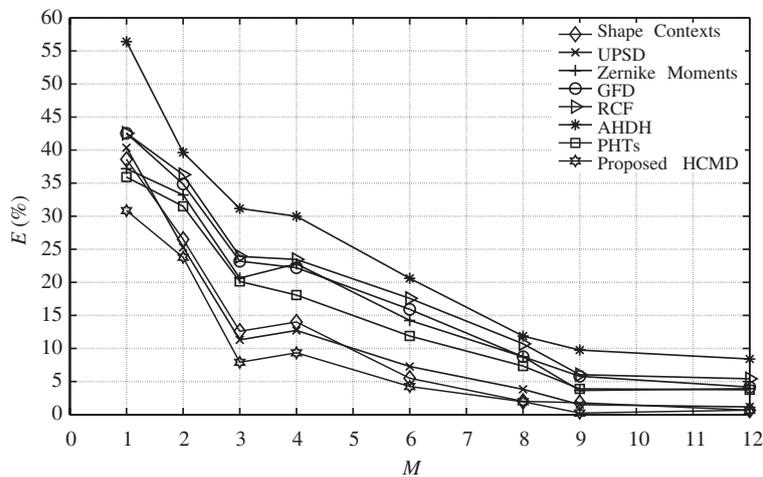


图 11 在 COIL-20 测试集上的目标识别错误率 E 对每个目标等视角间隔选取的训练样本数 M 变化的误差率曲线

Figure 11 The curve of the recognition error rate E vs. the number M of the uniformly selected views per object for training using the COIL-20 database

$20 \times (72 - M)$ 的测试集, 采用“1-NN”分类器给出识别错误率. 同文献 [23], M 分别取 1, 2, 3, 4, 6, 8, 9 和 12, 由此得到 8 组训练集和相应的测试集, 分别进行分类实验, 然后计算每一组实验的分类错误率 E , 将实验结果绘成错误率 E 对每个目标等视角间隔选取的训练样本数 M 变化的误差率曲线.

图 11 给出了 HCMD 方法和其他 7 种方法 (两种基于点集的方法: Shape Contexts [23] 和 UPSD [25], 5 种基于区域的方法: Zernike Moments [29], GFD [33], RCF [37], AHDH [39] 和 PHTs [31]) 的识别误差率曲线. 其中 Shape Contexts [23] 的误差曲线, 直接取自该文献报告的结果. 值得指出的是, 因 HCMD 方法主要利用目标的形状特征, 工作于目标的二值图像模式, 因此将图像库中的原灰度图像进行二值化处理, 图 9(b) 和 10(b) 分别给出了图 9(a) 和 10(a) 中的灰度图像经二值化处理的图像. 由图 11 可以看出, 对误差率曲线的 8 个数据点中的每一个数据点, 即给定相同的训练样本数 M , HCMD 方法的误差率都要低于参与比较的其他 7 种方法. 其中, 在 $M = 1$ 时, 即每个目标只有一个训练样本, 参与比较的其他方法的错误识别率比本文提出的 HCMD 方法的错误识别率高出 5% 以上, 在 $M = 3$



图 12 一个原形状样本 (左图) 和对其加入了不同强度椒盐噪声的形状样本 (右边 4 幅图: 噪声强度分别为 0.2, 0.4, 0.6 和 0.8)

Figure 12 An original shape sample (the left figure) and its versions distorted by salt and pepper noise (the right four figures with noise intensity of 0.2, 0.4, 0.6 and 0.8, respectively)

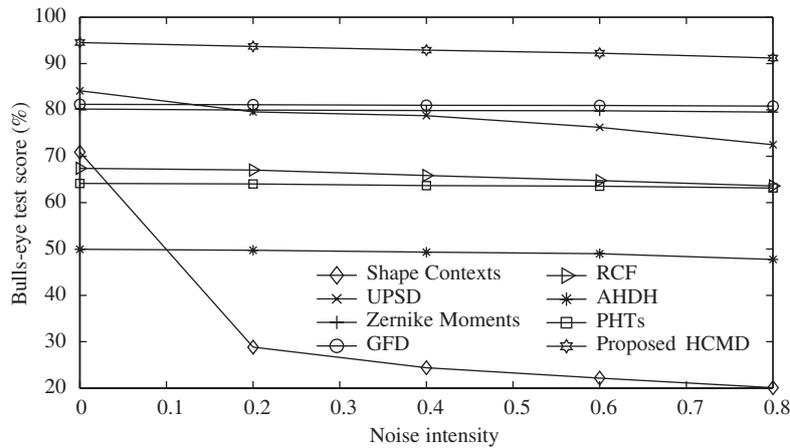


图 13 在 MPEG-7 CE-2 形状测试集上, 各类算法在加入了不同强度的椒盐噪声干扰下的检索率 (噪声强度分别设置为 0, 0.2, 0.4, 0.6 和 0.8)

Figure 13 The retrieval rates of various algorithms on the MPEG-7 CE-2 shape dataset distorted by different levels of salt and pepper noise (the noise intensities are set to 0, 0.2, 0.4, 0.6 and 0.8, respectively)

和 $M = 4$ 时, 参与比较的其他方法的识别错误率比本文提出的方法的识别错误率高出了 3.3% 以上. 图 11 给出的实验结果, 进一步验证了本文提出的 HCMD 方法具有比其他方法更强的形状识别能力.

4.4 对噪声干扰的鲁棒性测试

第 4 组实验用来测试 HCMD 方法对噪声干扰的鲁棒性, 这里采用与文献 [37] 相同的实验方案. 对 MPEG-7 CE-2 形状测试集中的每一个形状样本分别加入噪声强度分别为 0.2, 0.4, 0.6 和 0.8 的椒盐噪声, 由此得到 4 个加入了不同强度噪声干扰的形状测试集. 图 12 给出了一个原形状样本和对其加入不同强度椒盐噪声干扰的形状样本. 在这 4 个形状测试集中, 分别进行形状检索实验, 得到在各级噪声强度下的检索率. 为便于对各类方法在噪声干扰下的检索率进行比较, 将噪声强度作为横轴, 检索率作为纵轴, 每个算法的实验结果绘成一条检索率对噪声强度曲线. 图 13 给出了 HCMD 方法、两种基于点集的方法和 5 种基于区域的方法的噪声实验结果. 从该图可以看出, HCMD 方法随噪声干扰强度增加, 检索精确率下降幅度较小, 即使在噪声强度达到 0.8 时, HCMD 仍然达到了 91.20% 的高检索精确率, 表明了其对噪声干扰的稳定性. 观察参与比较的各方法, 6 种基于区域的方法相较于两种基于点集的方法 Shape Contexts [23] 和 UPSD [25] 对噪声干扰的稳定性更好, 它们的检索率都具有非常小的下降幅度, 其对噪声干扰的鲁棒性得益于描述子产生于形状区域的所有的像素点, 而非仅边界像

素点. 特别指出的是, Shape Contexts^[23] 方法在噪声干扰下, 检索率已不到 30%, 因为该方法在描述和匹配过程中, 需要利用目标边界点的切角信息, 而该信息对噪声干扰特别敏感.

5 结论

本文提出了一种基于分层复杂性度量的描述子 (HCMD), 以应用于一般的形状图像检索任务. 该方法将形状图像从各个方向 ($0^\circ \sim 360^\circ$) 迭代地切割成小的形状区块, 对每一个形状区块度量其方度、平衡性和长宽比等特性, 以描述其几何形态的复杂性, 同一层级各个方向的形状区块复杂性度量产生该层级的形状签名, 组合各个层级经归一化处理的形状签名产生整个形状的分层复杂性度量描述子, 两个形状的差异度由它们的 HCMD 之间的 $L1$ 距离来度量. HCMD 具有由粗到细的形状描述能力, 能进行不依赖于形状的缩放、旋转、平移和镜像变换的有效识别, 而且计算简单, 具有一般的应用能力, 既能有效处理简单的轮廓线形状, 又能处理具有复杂内部结构的区域形状的检索问题. 用 MPEG-7 的两组标准形状图像测试集对 HCMD 方法进行检索性能评估, 该方法无论是在 CE-1 轮廓线形状测试集, 还是在 CE-2 区域形状测试集, 取得的检索率都高出其他同类方法至少 13%, 也高出著名的基于点集的形状上下方法^[23] 至少 9%. 而且特别值得指出的是, 虽然 HCMD 方法不是专门为轮廓线形状设计的, 该方法在特征抽取和形状匹配中, 并没有利用边界点的有序信息, 但该方法仍然在 CE-1 测试集上达到了 86.36% 的检索率, 该结果已经好于许多基于曲线的形状描述子^[6, 7, 9, 10, 20~22]. 而且 HCMD 方法具有较高的计算效率, 比在 CE-1 测试集上取得了高检索率的两种基于曲线的描述方法: 形状树方法^[8] 和局部仿射不变描述子^[14], 检索速度要快 25 倍以上. 在著名的 COIL-20 测试集上的三维目标识别实验结果表明, HCMD 方法取得了比其他参与比较的基于点集的描述子和基于区域的描述子的更低的识别错误率, 进一步验证了其强的形状描述能力.

HCMD 方法的最大贡献在于提出了一种旋转分层的形状描述框架, 其计算简单, 对噪声干扰的鲁棒性强, 不需要对形状进行特别的结构化分解, 而且该方法不同于现有的仅对形状进行整体复杂性度量的方法, 它对形状进行了多个尺度的复杂性度量, 从而使得该方法具有非常强的形状识别能力. 特别值得指出的是, HCMD 方法对形状区块, 仅用了两种基于面积几何量和两种基于长度几何量的复杂性度量, 实质上, HCMD 方法是一个开放的描述框架, 基于该描述框架还可以引入其他形状区域特征, 如各类矩特征、离心率和紧致性等几何量来描述形状区块的复杂性, 以完成各种形状识别应用任务, 这将是今后进一步的研究工作.

参考文献

- 1 Loncaric S. A survey of shape analysis techniques. *Pattern Recog*, 1998, 31: 983-1001
- 2 Torres R D S, Falcão A X, Costa L D F. A graph-based approach for multiscale shape analysis. *Pattern Recog*, 2003, 37: 1163-1174
- 3 Costa L D F, Cesar R M. *Shape Analysis and Classification: Theory and Practice*. Boca Raton: CRC Press LLC, 2001
- 4 Wallace T P, Wintz P A. An efficient three-dimensional aircraft recognition algorithm using Fourier descriptors. *Comput Graph Image Process*, 1980, 13: 99-126
- 5 Persoon E, Fu K. Shape discrimination using Fourier descriptors. *IEEE Trans Pattern Anal Machine Intell*, 1986, 3: 388-397
- 6 Mokhtarian F, Abbasi S, Kittler J. Efficient and robust retrieval by shape content through curvature scale space. In: *Image Databases and Multimedia Search*. Singapore: World Scientific Publication, 1997. 51-58
- 7 Mokhtarian F, Bober M. *Curvature Scale Space Representation: Theory, Application, and MPEG-7 Standardization*. Dordrecht, Boston, and London: Kluwer Academic Publishers, 2003

- 8 Felzenszwalb P F, Schwartz J. Hierarchical matching of deformable shapes. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, 2007. 1–8
- 9 Adamek T, O'Connor N E. A multiscale representation method for nonrigid shapes with a single closed contour. *IEEE Trans Circuits Syst Video Tech*, 2004, 14: 742–753
- 10 Ling H, Jacobs D W. Shape classification using the inner-distance. *IEEE Trans Pattern Anal Machine Intell*, 2007, 29: 286–299
- 11 Alajlan N, Rube I E, Kamel M S, et al. Shape retrieval using triangle-area representation and dynamic space warping. *Pattern Recog*, 2007, 19: 1911–1920
- 12 Xu C, Liu J, Tang X. 2D shape matching by contour flexibility. *IEEE Trans Pattern Anal Machine Intell*, 2009, 31: 180–186
- 13 Xiong G Q, Qi D X, Guo F H. A class of orthonormal complete piecewise polygonal systems and applications thereof. *Sci Sin Inform*, 2012, 42: 70–82 [熊刚强, 齐东旭, 郭芬红. 一类完备的正交分段多项式函数系及其应用. *中国科学: 信息科学*, 2012, 42: 70–82]
- 14 Wang Z, Liang M. Locally affine invariant descriptors for shape matching and retrieval. *IEEE Signal Process Lett*, 2010, 17: 803–806
- 15 Biswas S, Aggarwal G, Chellappa R. An efficient and robust algorithm for shape indexing and retrieval. *IEEE Trans Multimedia*, 2010, 12: 372–384
- 16 Bai X, Yang X, Latecki L J, et al. Learning context-sensitive shape similarity by graph transduction. *IEEE Trans Pattern Anal Machine Intell*, 2010, 32: 861–874
- 17 Wang J, Bai X, You X, et al. Shape matching and classification using height functions. *Pattern Recog Lett*, 2012, 33: 134–143
- 18 Hu R, Jia W, Ling H, et al. Angular pattern and binary angular pattern for shape retrieval. *IEEE Trans Image Process*, 2014, 23: 1118–1127
- 19 Wang B, Gao Y. Hierarchical string cuts: a translation, rotation, scale, and mirror invariant descriptor for fast shape retrieval. *IEEE Trans Image Process*, 2014, 23: 4101–4111
- 20 Daliri M, Torre V. Robust symbolic representation for shape recognition and retrieval. *Pattern Recog*, 2008, 41: 1799–1815
- 21 Attalla E, Siy P. Robust shape similarity retrieval based on contour segmentation polygonal multiresolution and elastic matching. *Pattern Recog*, 2005, 38: 2229–2241
- 22 Super B. Retrieval from shape databases using chance probability functions and fixed correspondence. *Int J Pattern Recogn Artif Intell*, 2006, 20: 1117–1137
- 23 Belongie S, Malik J, Puzicha J. Shape matching and object recognition using shape contexts. *IEEE Trans Pattern Anal Machine Intell*, 2002, 24: 509–522
- 24 Grigorescu C, Petkov N. Distance sets for shape filters and shape recognition. *IEEE Trans Image Process*, 2003, 12: 729–739
- 25 Wang B. Shape recognition using unordered point-set description and matching of object contour. *J Softw*, 2016, 27: 3131–3142 [王斌. 软用于形状识别的目标轮廓无序点集描述与匹配. *软件学报*, 2016, 27: 3131–3142]
- 26 Peter A, Rangarajan A, Ho J. Shape L^1 -Ane Rouge: sliding wavelets for indexing and retrieval. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, 2008. 1–8
- 27 Zhang D, Lu G. Review of shape representation and description techniques. *Pattern Recog*, 2004, 37: 1–19
- 28 Hu M K. Visual pattern recognition by moment invariants. *IRE Trans Inf Theory*, 1962, IT-8: 179–187
- 29 Khotanzad A, Hong Y H. Invariant image recognition in Zernike moments. *IEEE Trans Pattern Anal Machine Intell*, 1990, 12: 250–261
- 30 Raveendran P, Omatu S. Performance of an optimal subset of Zernike features for pattern classification. *Inf Sci*, 1993, 1: 133–147
- 31 Yap P T, Jiang X, Kot A C. Two-dimensional polar harmonic transforms for invariant image representation. *IEEE Trans Pattern Anal Machine Intell*, 2010, 32: 1259–1270
- 32 Persoon E, Fu K. Shape discrimination using Fourier descriptors. *IEEE Trans SMC*, 1977, 7: 170–179
- 33 Zhang D S, Lu G J. Shape based image retrieval using generic Fourier descriptor. *Signal Process Image Commun*, 2002, 17: 825–848
- 34 Deans S R. *The Radon Transform and Some of its Applications*. New York: John Wiley and Sons, 1983.
- 35 Tabbone S, Wendling L, Salmon J P. A new shape descriptor defined on the Radon transform. *Comput Vis Image*

- Und, 2006, 102: 42–51
- 36 Hoang T V, Tabbone S. The generalization of the R-transform for invariant pattern representation. *Pattern Recog*, 2012, 45: 2145–2163
- 37 Chen Y W, Chen Y Q. Invariant description and retrieval of planar shapes using Radon composite features. *IEEE Trans Signal Process*, 2008, 56: 4762–4771
- 38 Yang M, Qiu G, Huang Y, et al. Near-duplicate image recognition and content-based image retrieval using adaptive geometric centroids. In: *Proceedings of the 18th International Conference on Pattern Recognition*, Hong Kong, 2006. 958–961
- 39 Sidiropoulos P, Vrochidis S, Kompatsiaris I. Content-based binary image retrieval using the adaptive hierarchical density histogram. *Pattern Recog*, 2011, 44: 739–750
- 40 Nene S A, Nayar S K, Murase H. *Columbia Object Image Library (COIL-20)*. Department of Computer Science, Technical Report, Columbia University CUCS-005-96. 1996

Hierarchical complexity measures for effective shape-based image retrieval

Bin WANG^{1,2}

1. *School of Information Engineering, Nanjing University of Finance and Economics, Nanjing 210023, China;*
2. *Key Laboratory of Electronic Business, Nanjing University of Finance and Economics, Nanjing 210023, China*
E-mail: wangbin@njue.edu.cn

Abstract A novel descriptor, called HCMD, which is based on hierarchical complexity measures, is proposed for generic shape based image retrieval. HCMD belongs to region-based methods and iteratively partitions a shape into smaller blocks along various directions. The geometrical properties of these smaller blocks, which are derived from each iterative cut, are measured to form a hierarchical description for the shape. The descriptor has the ability to characterize a shape from coarse to fine, and can effectively capture its complex inner structural features. Shape matching based on HCMD is independent of the rotation, scaling, translation, and mirror transform of the shape. It has low computational complexity and can effectively handle both the contour and region shapes. Three standard test sets, namely, the MPEG-7 CE-1 contour shape database, MPEG-7 CE-2 region shape database, and COIL-20 database, are used to evaluate the performance of the proposed HCMD, and extensive comparisons with state-of-the-art approaches, including five region-based descriptors, four point-set based descriptors, and two curve-based descriptors, are conducted. All experimental results indicate that the proposed HCMD outperforms these approaches in terms of their comprehensive performance based on the retrieval rates, retrieval efficiency, and general applications.

Keywords shape description, hierarchical complexity measures, feature extraction, object recognition, image retrieval



Bin WANG received a Ph.D. degree in computer science from Fudan University, Shanghai, China in 2007. He is currently a professor and supervisor of postgraduates at Nanjing University of Finance and Economics, Nanjing, China. His main research interests include computer vision, image processing, and pattern recognition.