

# 目标形状表达算法综述

吴 刚 徐利敏

(南京财经大学电子商务系 南京 210003)

**摘 要** 形状的检索和识别在医疗诊断、目标识别、图像检索和计算机视觉等领域都有重要应用,而高效的形状检索和识别取决于形状表达算法的优劣。给出了形状表达优劣的一般判断标准,对目前的主要形状表达方法进行分类,包括线性组合表达、空间关联关系表达、基于微分和积分的特征表达,以及变形表达等方法。对每一类方法,从使用的数学模型、多分辨率表达能力、不变量、鲁棒性、形状重构、信号噪声分辨性能等方面进行分析和综合评价,指出这些表达方法的优劣,尤其是就这些算法的数学机理进行分析探讨,并给出了未来需要进一步研究的方向和思路。

**关键词** 形状表达,形状分析,目标识别,形状检索,图像分析

中图分类号 TP391 文献标识码 A DOI 10.11896/j.issn.1002-137X.2019.07.005

## Review of Shape Representation for Objects

WU Gang XU Li-min

(Department of Electronic Business, Nanjing University of Finance and Economics, Nanjing 210003, China)

**Abstract** Shape retrieve and objection are widely applied into medical diagnostics, target recognition, image retrieve and computer vision, etc. The efficient retrieve and objection of shapes completely depend on an excellent shape representation algorithm. This paper proposed the assessment criterion for shape representation. Then, according to the criterion, the existing shape representations were categorized into linear combination representations, spatial association relationship, feature representation based on differential and integral property of shapes and deformation representations. Each of these methods was analyzed and accessed in terms of mathematical principle, the ability of multiscale representation, variants, robust, reconstruction of original shapes, identification of signal and noise, etc. Furthermore, the advantages and disadvantages of each algorithm were discussed, especially, explored from the principle of mathematics. Finally, the suggestions for the future research were also given.

**Keywords** Shape representation, Shape analysis, Object recognition, Shape retrieve, Image analysis

## 1 引言

形状表达就是抽取形状的一些少量特征参数,并选用这些参数描述形状。这些特征参数需要具备以下几条性质:1)能够将一个物体形状和其他物体形状区分开来;2)能够将同一个物体在不同角度下获得的形状识别为同一个物体;3)对噪声以及轻微变形不敏感;4)能够识别非刚性变形的形状;5)特征参数尽量少;6)特征参数有利于形状之间的相似性计算;7)能够根据特征参数重构原始形状。目前,许多综述文献<sup>[1-4]</sup>大多聚焦于以形状检索为目的的形状表达问题,较少以这 7 条性质作为标准对形状表达方法进行综述评价。形状检索是指给定一个形状,从数据库中检索与该形状相似的一些形状,并依据相似度对检索到的形状进行排序。例如,给定一片树叶,从数据库中检索所有与这片树叶相似的树叶,这在计

算机自动辨识植物品类方面得到广泛应用。一般情况下,以检索为目的的形状表达只要能够满足旋转、平移和同尺度比例缩放即可。除此之外,还有另外两类形状表达的应用,一类在计算机视觉领域,也就是同一个物体在不同视角下经过透射变换获得形状。这种应用要求形状的表达具有透射不变性,即形状虽然不同,但是表达的是同一个物体,也认为是相同的形状,亦即发生仿射变换时形状的表达应该具有仿射不变性,发生射影变换时形状的表达应该具有射影不变性。另一类表达应用在目标跟踪等领域,一个人在行走时,其人体姿态发生了非透视变换变形,也就是非刚性变形,如人体的四肢发生了运动,这显然是非刚性变形,这种形状的表达也被称为基于变形的形状表达<sup>[5]</sup>。本文综述这 3 个领域的形状表达算法,并从形状表达的 7 条性质出发进行评价,注重从数学角度对这些算法进行分析,试图给出其背后的数学因素。

到稿日期:2018-08-11 返修日期:2018-11-16 本文受国家自然科学基金(61372158)资助。

吴 刚(1969—),男,博士,教授,CCF 会员,主要研究方向为图像处理和计算机视觉,E-mail:gangwu@nufe.edu.cn(通信作者);徐利敏(1980—),女,副教授,主要研究方向为模式识别与计算机视觉。

## 2 线性组合方法

线性组合方法就是将形状看作一个线性空间的元素,然后寻找一组基,将此形状线性表达出来。一般情况下,可以将形状视为线性空间的一个连续函数,那么该空间可以有无数个基函数,形状即是这些基函数的线性组合,而该线性空间中的加法和数乘法运算就是一般意义下的函数加法和复数域上的数和函数乘法运算。假设一组基函数是  $\phi_1(x, y), \phi_2(x, y), \dots, \phi_i(x, y), \dots, \phi_n(x, y)$ , 那么形状线性空间的任一形状函数  $f(x, y)$  能够被这组基函数线性表示出来,即:

$$f(x, y) = \sum_{i=1}^n c_i \phi_i(x, y) \quad (1)$$

其中,  $c_i (i=1, 2, \dots, n)$  是一个复数数列,我们称之为基函数对应的系数,也可以称为基函数对应的坐标,有时根据基函数的选取情况称为矩,比如基函数  $\phi_i(x, y)$  对应的系数  $c_i$  称为第  $i$  阶矩。如此,形状表达问题就转化为如何确定一组基函数和对应的系数的问题,因此确定在这组基下的坐标就成了研究的主要内容。目前,应用到形状表达的基函数主要是一些正交基函数和多项式基函数、高斯基函数以及径向基函数等。选定这些基函数之后,需要确定对应的坐标,也就是系数。如果选定一组基函数是正交的,文献[6]指出系数的计算公式如下:

$$c_i = \frac{\int_R w(x, y) f(x, y) \phi_i(x, y) dx dy}{\int_R w(x, y) [\phi_i(x, y)]^2 dx dy} \quad (2)$$

其中,  $R$  为函数  $f(x, y)$  的定义域,  $w(x, y)$  为权函数。事实上,对式(1)两边同时乘以  $\phi_i(x, y)$  并在  $R$  上取重积分,即得式(2)。一般在图像处理中,给定的形状常常是二值图像,因此上面的积分需要使用离散点求和代替。而对于非正交基函数,无法使用式(2)计算系数,于是许多文献提出使用线性拟合的方法来近似获得系数,此时线性拟合只能实现对形状轮廓的拟合,这是因为形状常常是二值图像,也就是根据形状边界点坐标拟合一个隐式曲线  $f(x, y) = 0$  来描述形状曲线。例如,文献[7-8]提出和使用 Min-Max 方法进行拟合,从而获得系数,具体方法如下。

假设形状曲线有  $m$  个点,  $Z_k, N_k$  和  $T_k$  分别是曲线形状在点  $(x_k, y_k) (k=1, 2, \dots, m)$  处的梯度、单位法向量和切向量,将点  $(x_k, y_k)$  代入式(1),并记  $\phi(x_k, y_k) = [\phi_1(x_k, y_k), \phi_2(x_k, y_k), \dots, \phi_n(x_k, y_k)]^T$ , 于是得到矩阵  $M_0 = [\phi(x_1, y_1), \phi(x_1, y_1) \dots \phi(x_m, y_m)]$ 。根据梯度和单位法向量内积为 1, 而与切向量内积为 0, 得到下面的公式:

$$M_1 = [Z_1 N_1 \quad Z_2 N_2 \quad \dots \quad Z_m N_m]$$

$$M_2 = [Z_1 T_1 \quad Z_2 T_2 \quad \dots \quad Z_m T_m]$$

令  $b = [0_{m \times 1} \quad 1_{m \times 1} \quad 0_{m \times 1}]^T$ , 同时定义一个总的矩阵  $M = [M_0 \quad M_1 \quad M_2]^T$ , 并令  $C = [c_1, c_2, \dots, c_n]^T$ , 即得到一个线性方程组矩阵  $MC = b$ 。使用最小二乘法很容易求解该方程组, 即得:

$$C = (M^T M)^{-1} M^T b \quad (3)$$

以上是一般线性组合方法获得形状表达的基本思想。这种线性表达有两种基本方法:1)直接描述形状边界,也就是基

于边界轮廓的方法;2)描述形状二值图像,也就是基于区域的方法。下面对各种线性表达方法进行具体讨论。

### 2.1 Fourier 描述子

Fourier 描述子即选用  $e^{jkx}$  ( $k=1, 2, \dots, n$ ) 作为基函数组,具体应用有两类。一类是将二维形状边界轮廓重新参数化成一维数据集,也就是将二维形状使用一维形状进行签名表达<sup>[9]</sup>,然后应用 Fourier 基函数对一维签名获得的数据进行离散 Fourier 变换,获得的系数就是 Fourier 描述子。显然,形状签名算法是一维 Fourier 作为基函数进行线性组合表达的基础。文献[10]综述了各种形状签名算法,具体包括中心距离、切角、累积角、曲率、面积、弦长,也就是将二维形状转化为一维形状签名后,再使用一维 Fourier 基函数进行线性表达,从而获得 Fourier 描述子。文献[11]通过实验表明:使用中心距离签名获得的 Fourier 描述子更好地表达了形状特征,有较高的形状检索和识别率。以上形状 Fourier 描述子都是从一维形状签名获得,这种描述子表达形状时不够稳定,而且难以根据描述子重构原始形状曲线;而椭圆 Fourier 描述子能较好地克服这种问题,它将形状边界的每一个点  $(x, y)$  写成参数形式  $(x(t), y(t))$ , 其中  $t$  可以看作弧长参数,然后分别对  $x(t)$  和  $y(t)$  进行 Fourier 变换,进而获得 Fourier 描述子。这种描述子容易计算仿射不变量,有明确的几何意义,但只在封闭的形状曲线中有较好的效果,用在开曲线中时就会在曲线端点处产生振荡,如图 1(b)所示。如果将开曲线参数化为从开始点到结束点,再从结束点到开始点的集合,那么这条开曲线就变换成了一条封闭曲线,从而可以使用椭圆描述子进行表达;但是这使椭圆描述子的效率低下,而且不容易重构原始形状。



图 1 Fourier 描述子与 Chebyshev 描述子的比较

Fig. 1 Comparison of Fourier descriptors and Chebyshev descriptors

Fourier 描述子的优点是方便计算,有直接的物理含义,低阶描述子表达形状曲线的主要轮廓特征,高阶描述子表达形状曲线的细节特征,这有利于对形状曲线进行多尺度分析;另外,Fourier 描述子对形状轮廓噪声鲁棒,而且很容易计算出欧氏不变量和仿射不变量,这在形状检索和识别中具有较好的应用价值。其缺点是无法确定高频信号是噪声信号还是形状曲线本身的特征;另外,由于形状曲线是数字化的,因此做 Fourier 变换时使用离散点求和代替积分,引入了量化误差,文献[11]等使用 Chebyshev 描述子来克服这个问题。

### 2.2 Chebyshev 描述子

文献[12-15]等提出使用 Chebyshev 函数作为基函数对形状进行表达。其中,文献[12]提出的 Chebyshev 描述子能够对开曲线进行很好的描述,而且可以克服 Fourier 描述子重构的曲线在尾端振荡的问题,如图 1(c)所示,重构的曲线很好地保持了端点的走向。该描述子选择 Chebyshev 多项式作

为基函数,具体形式是  $\cos(i * \arccos(x))$ ,  $i=1,2,\dots,n$ 。文献[6]指出该组正交基函数不仅仅在区间 $[-1,1]$ 内正交,而且关于 Chebyshev 节点离散正交,因此 Chebyshev 描述子表达形状曲线有更好的精度和性能。文献[13-14]提出使用二维 Chebyshev 多项式函数作为基函数,对形状二值图像进行表达,并给出了各种欧氏和仿射不变量,研究表明它们具有较好的形状识别和检索性能。事实上, Chebyshev 描述子不仅具有 Fourier 描述子的所有优点,而且由于它们离散正交,因此误差更小;另外, Fourier 描述子是复数,而 Chebyshev 描述子是实数,因此更容易对轮廓形状重构,而且参数更少,性能却更高。 Fourier 描述子和 Chebyshev 描述子有一个共同的缺点:对于形状细节的表达,无法将噪声和形状高频信息区别开。可以通过分析基函数研究这种形状表达的特性:基函数序列为  $\cos(i * \arccos(x))$ ,令  $t = \arccos(x)$ ,则基函数的形式为  $\cos(it)$ ,显然,随着  $i$  的增加,基函数频率越来越高,而一个形状的高频信息总是远远少于低频信息的,故计算的描述子低阶矩对应的系数较大,高阶矩对应的系数较小。观察式(2)的分子可以发现,其实质是计算两个函数向量的内积,也就是计算形状曲线  $f(x,y)$  和余弦曲线  $\cos(it)$  的内积作为它们的相似度。显然,一个形状曲线的高频部分的信息一般少于低频部分的信息,因此形状曲线和余弦函数的相似度随着余弦函数频率的增大而变小,即对应的系数变小,这是三角函数作为基函数能够多尺度表达形状曲线的数学背景。

Chebyshev 描述子作为一类参数化方法,存在一个与起始点相关的主要问题:因为它的基函数是余弦函数,因此对于非封闭的形状表达效果较好;但是对于封闭形状,计算的不变量依赖于起始点的选择,另外分析 Chebyshev 多项式可以发现,其零点在形状两端分布密集,而在形状中间分布较为稀疏,这导致其形状表达能力上的不一致。

### 2.3 Zernike 矩

Zernike 矩表达<sup>[10,16]</sup>选用的一组正交基函数是  $V_{pq}(r,\theta) = R_{pq}(r)e^{iq\theta}$ ,其中  $p,q$  取整数,且满足  $0 \leq q \leq |p|$ ;  $R_{pq}(r)$  是关于  $r$  的实值多项式函数。 $p,q$  取不同的整数值,则得到不同的基函数,而这些基函数在单位圆上正交,因此将其应用到二值形状图像时需要将分析的形状映射到单位圆,这样计算得到的 Zernike 矩具有旋转不变性。显然, Zernike 矩具有与 Fourier 描述子以及 Chebyshev 描述子相似的性质,即低阶矩描述形状的主要轮廓特征,高阶矩描述形状的细节特征,故具有较强的噪声鲁棒性,也能够相对容易地根据 Zernike 矩重构原始形状。但是与前两种描述子相比,其计算量较大,需要复数计算,而且需要将图像重新映射到单位圆,因而会引入量化误差;另一方面,在 Zernike 矩中,径向特征属于空域,而圆度特征属于频域,因此在径向方向上不具有多尺度分析性质。文献[14]等指出在形状的检索和识别性能方面,该描述子低于 Fourier 描述子和 Chebyshev 描述子。

### 2.4 几何矩

几何矩最先由文献[17]提出,并在文献[18]等研究中得到广泛应用,主要用来实现对形状的检索。几何矩选择的基函数是  $x^p y^q$ ,  $p,q=0,1,2,\dots$ ,很显然这组基函数不是正交函数,无法使用式(2)计算基函数对应的矩即坐标,但是仍然可

用该公式计算其伪坐标。从数学角度看,几何矩实质上是一种伪矩,也就是坐标在基函数上的近似。在计算几何矩时,一般需要将形状数据进行 Z-Score 标准化,使形状点的坐标都落在 0 和 1 之间。根据式(2)容易发现,低阶几何矩逼近基函数的坐标值,而高阶矩误差较大,因此一般只用低阶几何矩检索形状。从另外一个角度分析,低阶矩还有其他物理解释,比如可将其看作形状距离中心点的方差,而高阶矩没有其他物理意义。相关研究表明,几何矩对基于区域发生仿射变换的形状进行表达时性能较差,对噪声敏感,因此逐渐被其他方法代替。

### 2.5 隐式多项式曲线

几何矩的基函数是二元多项式函数项,使用伪正交基函数的方法来计算基函数对应的坐标,这种计算只是一种近似计算,存在很大的局限。因此,文献[7]等提出不再使用基函数的正交性计算对应的矩,而是将形状边界曲线看作连续函数空间的一个子空间,也就是满足连续函数为 0 的空间,在此空间中使用隐式多项式函数拟合的方法实现对应矩的计算,这种拟合方法包括线性方法和非线性方法两类。线性方法可以根据式(3)计算出来,非线性方法如文献[19]等借助优化的方法给出。一般情况下,线性方法计算简单,但是逼近精度不及非线性方法。隐式多项式曲线对描述复杂曲线有局限性,于是文献[8]提出用有理隐式多项式曲线来表达形状。与前几种表达方法相比,隐式多项式曲线表达形状具有自动修补形状边界缺失、容易计算代数和几何不变量等性质,因此常常用在形状识别中;但是由于对复杂形状描述的效率低,因此很少用在形状检索中。

隐式多项式曲线是对形状的隐式表达,而基于正交基函数的线性组合方法是显式的参数化表达,因而更为稳定。在某些应用中,隐式表达比参数表达更有效,例如判断一点是否在形状内,只需要将该点坐标代入隐式方程即可;但是显式的参数化表达很难实现。隐式表达最为困难的问题是如何根据边界形状数据拟合出稳定而且较为精确的隐式方程。这个问题的实质是如何根据边界的约束条件计算出最为高效(也就是阶数最低)的隐式多项式方程,这需要从数学理论上给以突破。我们可以将二维平面视为一个赋范线性空间,对于给定的形状边界( $L$ ),所谓隐式多项式曲线表达就是寻找一个连续线性泛函  $f$ ,使得:(1)  $x \in L, f(x) = 0$ ; (2)  $x_0 \notin L, f(x_0) = d^n$ ; (3)  $\|f\| \leq 1$ 。其中,  $d$  为  $x_0$  到数据集  $L$  的距离。根据连续线性泛函延拓定理,当  $n=1$  时,这样的线性泛函是存在的,其实就是距离函数,但是这种函数不光滑,因此从理论上确定合适的次数  $n$  是解决问题的关键。

### 2.6 其他基函数表达以及问题讨论

基于上面基函数表达的思想,只要给定一组连续函数的基函数,就可以得到一种形状表达方法,它们或者基于形状二值图像,或者基于形状边界轮廓,最终都可以给出一组描述形状的描述子,即称为矩。例如,给定一组 Legendre 多项式作为基函数,就可以给出基于 Legendre 矩的描述子,文献[20-21]提出了这种描述子。研究发现,这种描述子对噪声敏感,因此较少应用于形状检索和识别中。文献[22]使用 B 样条作

为基函数对形状进行表达,这种表达精度高,但是较难计算形状的仿射不变量等,因此也很难获得应用。

选择基函数计算出矩以后,常常需要确定在多少阶矩之后反映的是噪声信号而在此之前反映的是形状的实际特征。一般可以使用信噪比来解决这个问题,比如信噪比是 30,则 Zernike 矩选择前 10 个就可以反映形状信息,其之后的矩反映的是噪声信息。对于不同的基函数,如何选择具体的矩阶数,目前还没有相应的文献进行研究。另外,该类算法还存在如下一些主要问题:计算的不变量常常是从形状的纯几何变换得到,而现实物体很少严格遵守几何变换,特别是非刚性变形;不变量对边界噪声敏感;新的不变量很难寻找;基于不变

量的形状匹配困难。还有其他一些问题,包括基函数线性组合描述方法无法区分噪声和形状高频信息,在过滤噪声信号的同时将有用的形状细节信息也一同过滤掉。比如对于具有相似形状的不同树叶,它们的区别仅仅是在树叶边缘有微小的锯齿状变化,而基函数线性组合表达常常将这些信息作为噪声过滤掉,导致树叶形状检索的性能下降。产生这些问题的根本原因是:基函数线性组合表达属于整体形状表达,会丢失或者无法表达空间点之间的关联关系。因此,许多文献开始研究空间关联关系表达方法。

表 1 主要从变换性质、优缺点以及适用范围等方面对线性组合类方法做了较为详细的比较说明。

表 1 线性组合表达方法的比较

Table 1 Comparison of linear combination representation algorithms

算法	变换性质	优点	缺点	适用范围
Fourier 描述子	连续正交变换	多分辨率表达,能够计算不变量,对噪声鲁棒	数字化误差大,形状数据集需要重新参数化	封闭边界形状描述
Chebyshev 描述子	离散正交变换	多分辨率表达,效率高,不变量计算容易,对量化误差鲁棒	封闭形状表达起始点敏感,形状数据集需重新参数化	不封闭边界形状描述
Zernike 矩	空域和频域混合变换	多分辨率表达,频域和空域同时对形状表达	需要形状做单位圆映射,不变量性能差,对噪声敏感	形状区域的表达
几何矩	伪正交变换	不变量计算容易,形状表达简单	性能差,对噪声敏感,高阶矩缺乏直接的物理解释	形状区域的表达
隐式多项式	多项式最小二乘拟合	对噪声鲁棒,容易计算不变量,能够修补形状边界的缺失	复杂形状描述效率低,拟合算法寻找困难	各种边界形状的表达

### 3 空间关联关系表达

#### 3.1 几何关系表达

空间关联关系表达在很多文献<sup>[1,3,23-26]</sup>中被称为结构化方法或者空间剖分方法。该方法主要考虑形状边界点在一定领域范围内,关联关系以及结构的特点。显然,关联关系表达得越充分,则形状检索和识别的性能越高。从这个角度出发,可以计算形状的主轴、对称轴、圆度、偏心率和面积,但是这些形状表达都较为粗糙,应用的范围小,性能较差,于是许多文献开始提出关联性越来越强的算法。文献[27]提出了三角形面积表达的两个新描述子,一个是面向三角形的角表达(TOA),一个是三角形边长和角度表达(TSLA),这两个描述子提供了比仅仅使用三角形面积表达更强的形状局部分辨信息。文献[28]提出多尺度面积表达(MTAR)方法,多尺度面积即在不同尺度下的每一点和边界上其他点所围成的三角形面积。该方法直接在空域范围内提供从粗粒度到细粒度的形状表达,但是这两种表达方法通过计算边界点所围成的三角形面积来表达形状点的关联关系,所使用的信息不充分,因此表达的性能不是十分理想。文献[29]从边界点对的空间关系来捕捉几何特征,提出了多尺度矩阵。该方法通过计算所有边界点对之间的距离生成多尺度距离矩阵,尺度就是弧长,矩阵的每一行捕捉某种范围的几何特征,然后通过移动和排序来创建不变量矩阵,以此进行形状表达。文献[30]提出边界点组成的角以及它到两个等距离相邻点组成一个角度模式(AP),然后将相邻 AP 的关系编码成二值字符串,称为二值

角度模式(BAP),最后通过计算不同尺度下的 AP 和 BAP 来实现形状轮廓的表达。这两种方法都是在给定尺度下计算形状边界上两个点之间的关系,然后通过计算多种尺度下的这种关系来寻求所有点之间的关联关系,显然无法真正、有效地表达出边界点之间的关联特点;而且这种两个点之间的计算受噪声和形状部分缺少的的影响很大,从广义角度看,以上方法都使用三角形来近似表示曲线的曲率,因此存在微分方法的缺点。为此,文献[31]提出多尺度弦切方法,也就是通过连接形状边界上的两个点的直线将形状分割成多个部分,计算形状和这条弦所围面积的最大值、最小值、分割的曲线长度以及长度与弦长的比这 4 个参数来实现形状的表达。显然,面积的计算类似于积分的方法,因此对噪声的干扰性较强。为了避免两个点的关联关系过少,通过两个点的距离大小来生成多尺度的 4 个参数,从而实现关联关系的充分使用。该方法只是对以上微分方法的改进,在方法上没有实质上的变革,一样存在形状边界点的关联分析没能统一使用和分析的问题;同时还存在一个较大的缺点,即该方法不仅没有将形状维数降低,还大大提高了维数,增加了计算开销。

#### 3.2 统计关系表达

几何关系表达主要是计算形状边界点之间的几何关系,例如计算直线距离、曲线和直线所围面积等,较少考虑形状边界点存在随机噪声的情况,因此有的文献基于边界点的统计特征给出相应的表达方法<sup>[32]</sup>。文献[33-35]提出形状的上下文信息表示,不是直接计算形状边界点上的几何信息,而是利

用其统计特征。具体方法是:在形状边界点上均匀地取一些控制点,对于每一个控制点,以它为中心进行圆形网格的剖分,计算落在每一个网格中的形状边界点的数目,而落在同一个网格的点有确定的位置,即通过圆形网格距离控制点的距离以及与水平直线的夹角来确定,因此可以根据控制点信息、每一个网格坐标信息以及落在网格坐标的边界点数目来绘制位直方图,并以此作为形状表达。该方法具有两个优点:1)考虑了给定控制点相邻边界点的特征,即距控制点远的边界点自然落在外层网格中,距离近的点落在内层网格中,体现了边界点之间的关系;2)避免了噪声信号的干扰,因为只是统计了边界点的数目,并不计算边界点的几何关系,所以具有较强的鲁棒性。但是,这种克服噪声的方法也不可避免地丢弃了落在同一个网格内的点之间的关联信息,在某一类形状的检索和识别中会降低其性能。

### 3.3 存在的问题讨论

空间关联分析方法直接在形状边界点之间探求关联关系,不需要进行频域变换,因此理论上能够确保大多数形状信

息得到分析利用,尤其是形状的局部细节信息,因此在处理部分物体遮挡以及部分匹配时有优势;对于需要识别类似树叶这种需要精确表达边缘细节特征的应用领域,空间关联关系方法表现的性能较好。但是,其也存在一些缺点:1)常常需要寻找一些控制点,而这需要先验知识;2)在形状检索和识别时,匹配困难,计算复杂,很多方法使用动态规划方法进行匹配,这很难在线实时应用;3)几何关系的表达算法对噪声敏感,直接计算两个边界点之间的关系时不可避免地会引入噪声信息;4)缺乏数学理论基础,无论是空间几何关系方法还是统计关系方法,都缺乏严密的数学理论基础,是一种试探性的实验验证,即计算一些几何关系,或者做些统计,很难从理论上对形状描述的精度进行分析,都是通过对一些数据集的验证来评判算法的优劣,该类算法还处于探索阶段;5)很难从空间结构关系表达中重构原始形状,这也是难以从理论上对该类方法进行评判的主要原因。表2主要从空间关联性质、优缺点以及适用范围等方面对几何关系类和统计关系类表达方法做了较为详细的比较和总结。

表2 几何与统计关系表达方法的比较

Table 2 Comparison of geometrical and statistical relation representation algorithms

算法	空间关联性质	优点	缺点	适用范围
几何关系	形状边界点的几何特性	描述精确,可用于局部匹配和局部数据点关联分析	对噪声敏感,计算复杂,无法重构形状,无法计算仿射不变量	形状边界表达
统计关系	形状区域的统计性质	表达稳定,对噪声鲁棒,将局部关联分析和局部统计分析相结合,适用范围广	丢失了局部数据点的关联关系,难以表达形状的高频信息,无法计算仿射不变量	形状区域表达和形状的边界表达

## 4 基于微分和积分的特征表达方法

### 4.1 微分表达

特征点是形状边界轮廓中的重要特征,在形状的检索和识别中获得广泛应用<sup>[36-37]</sup>。显然,角点是形状边界轮廓的重要特征点,在它们的提取过程中需要计算曲率,因此有许多基于形状边界点的曲率特征点表达方法,其中曲率尺度空间(CSS)描述子<sup>[38-39]</sup>就是一类典型的基于特征点的形状表达方法。该算法的具体步骤是:首先对给定形状边界上的一点使用较小尺度 $\alpha$ 的高斯函数对包含该点的形状边界进行平滑;然后计算形状边界在该点的曲率;以此方法计算出形状边界上每一点的曲率后,将曲率过零点作为特征点,因此给定一个尺度 $\sigma$ 就可以获得形状边界上的所有曲率过零点,由此生成关于曲率尺度空间的二维图像,该二维图像就是形状的表达即CSS图像。对于不同的形状,可以得到不同的CSS图像,通过CSS匹配实现形状的检索和识别。该方法虽然可以得到形状较好的表达,但是由于曲率的估计涉及到导数的计算,因此受噪声干扰大。对此,文献[40]等提出使用给定长度的三角形夹角代替曲率的估算,这种方法实质上是第3节提到的形状边界点关联关系方法,因此仍然存在鲁棒性差等缺点,为此有一些研究者提出了基于积分方法的形状表达算法。

### 4.2 积分表达

由于积分方法能够消除噪声的干扰,且鲁棒性较强,具有微分法不具有的特点,因此使用积分方法研究形状的表达成为另外一个研究方向。文献[41-42]中的相关方法就是一类

基于积分形状的描述方法,其具体思想是:对于给定区域 $D$ (它可以是简单或者连通区域),引入形状特征函数 $F_D^\sigma$ ,在给定尺度 $\sigma$ 使用核 $G_\sigma$ 来描述区域 $D$ 的局部几何特征。具体公式如下:

$$F_D^\sigma = \chi_D(x) \cdot (1 - G_\sigma * \chi_D(x)) \quad (4)$$

其中, $G_\sigma$ 是一个标准正态高斯核,其二维标偏 $\sigma > 0$ 。

$$G_\sigma(x) = \frac{1}{2\pi\sigma^2} e^{-\frac{\|x\|^2}{2\sigma^2}} \quad (5)$$

注意,式(4)中高斯核和特征函数是卷积。这说明形状内每一点处的特征值 $F_D^\sigma$ 可以解释为经过高斯核加权后的形状区域的面积。于是,提出一种新的形状描述子——通过在一个尺度范围内计算形状特征的积分不变量来构造一个尺度的函数。下面定义在尺度 $\sigma$ 下形状 $D$ 的描述函数 $S_D(\sigma)$ :

$$S_D(\sigma) = \frac{\int_\Omega F_D^\sigma(x) dx}{\int_\Omega dx} = \frac{\int_\Omega \chi_D(x) \cdot (1 - G_\sigma * \chi_D(x)) dx}{\int_\Omega \chi_D(x) dx} \quad (6)$$

其中, $\rho = \sigma \sqrt{\int_\Omega \chi_D(x) dx}$ 。可以看出,在形状区域计算特征值积分即获得形状表达,同时进行标准化以获得尺度不变量。比较这种积分方法与微分方法可以发现:微分方法使用导数求形状的特征点;而积分方法是通过求积分计算形状内外所占的面积比,因此更鲁棒,但是对于边界形状十分细微的信息其也很难刻画出来。表3主要从特征点提取方法、优缺点以及适用范围等方面对微分表达方法和积分表达方法做了较为详细的比较和总结。

表 3 微分与积分表达方法的比较

Table 3 Comparison of differential and integral representation algorithms

算法	特征点提取方法	优点	缺点	适用范围
微分表达	形状边界点的微分性质	多尺度表达,能够重建形状,表达效率高,较好地应用形状匹配	对噪声敏感,差分估值微分误差大,无法计算仿射不变量	形状边界表达
积分表达	形状区域的积分性质	表达稳定,对噪声鲁棒	表达精度差,无法计算仿射不变量	形状区域表达和形状边界表达

## 5 变形表达方法

一般形状是物体在空间发生平移、旋转或者缩放变化所获得的图像,有的是物体在空间发生仿射和射影几何变换后所获得的图像。一般的形状表达方法都是针对这两类形状,但是对形状发生非刚性变形进行检索和识别就困难得多。例如,人行走时,上半身形状基本保持不变,但是手臂和双腿就会不停地移动,而且这种移动不满足任何一个简单的几何变换公式;还有云朵一类的形状也可以毫无规律地发生局部变化。这些都是非刚性变形,对这种形状进行表达,并基于表达寻找不变量进行检索和识别,是比较困难的。目前研究的基本思想就是将形状看作参数函数,而将形状空间看作希尔伯特空间<sup>[43]</sup>,这样的空间就是定义度量并可以处理为黎曼流行的非线性和无限维形状空间,建立黎曼流行结构以后,对于不同的形状,可以定义度量来计算两个变形形状的相似度。文献[44-46]等在变形空间中定义度量,而将空间看作一个群,通过最小化被称为效率函数的目标函数来计算变形形状的相似性,因此对于不同的变形空间需要定义一个变形模板,几何不变量的计算都要参考这个变形模板。文献[5]提出另外一种方法,即不使用变形模板,而是将形状表达为有限维李群矩阵的一个元素,这样尺度和位置的改变都可以在预处理计算

中过滤掉,各种变形通过求解一个显式测地方程即可得到。实验表明,该方法对非刚性变形的检索和识别率较好。该方法并不要求先验知识和其他变形方法相比,限制条件较少,对局部移动、扰动和缺失鲁棒性强,同时不要求拓扑变形必须连续;但是该方法无法处理拓扑改变的情况,同时计算复杂,需要进一步减小计算量。

表 4 分别从实现的数学计算方法、多分辨率表达能力、欧氏不变量、仿射不变量、检索性能、识别性能、鲁棒性、噪声和 高频信号分辨性能、适用范围以及能否重构形状等方面对本文讨论的 5 种形状表达方法进行比较。对于多分辨率表达能力、检索性能、识别性能、鲁棒性、噪声和 高频信号的分辨性能这几个指标,将其划分为 4 个等级(分别为强、较强、一般和差)进行比较。通过比较可以观察到只有线性组合类方法的性能较为全面,事实上,这类方法在目前应用得最为广泛,其他表达方法在某些性能方面表现较差。从不变量计算和能否重构形状等方面也可以观察到,线性组合方法表现最好,这与该类方法有严密的数学理论基础直接相关;而其他类方法只是用在一些特殊的领域,比如在某些特殊的树叶类精确检索中空间关系表达表现得更为有效,而变形表达方法在人体肢体运动方面的效果更好。

表 4 各类形状表达方法的比较

Table 4 Comparison of existing representation algorithms

表达类型	数学计算方法	多分辨率表达能力	不变量 (欧氏)	不变量 (仿射)	检索性能	识别性能	鲁棒性	噪声和 高频信号 的分辨性能	适用范围	能否重构形状
线性组合表达	正交变换	强	有	有	较强	较强	强	一般	形状边界和区域	能
空间关系表达	空域关联分析	较强	有	无	强	一般	一般	较强	形状边界	不能
微分表达	基于微分的特征点检测	无	无	无	较强	一般	一般	差	形状边界	近似构建
积分表达	基于积分的特征点检测	无	无	无	较强	一般	强	差	形状边界和区域	不能
变形表达	变形最小能量	无	有	有	一般	较强	强	差	非刚性变形	不能

## 6 形状表达研究的展望

### 6.1 线性组合的测度定义

对于线性组合方法,尤其是正交基函数的线性组合,形状描述子的实质是线性空间在这组正交基下的坐标,而坐标的计算是通过形状曲线函数和正交基的内积运算获得的,这种运算的结果会造成形状细节的丢失。尽管可以通过高阶基函数获得细节信息,但是又无法区分噪声和形状细节,因此陷入了矛盾;若要提高形状细节描述,就要提高描述子的阶数,这势必要引进噪声;若要降低噪声的影响,就要降低描述子的阶数,这又将造成形状细节信息无法充分描述的问题。这是该类线性组合方法,也就是矩变换方法普遍存在的问题。从另外一个角度观察式(2),其分母由于基函数正交,因此为固定值,假设为 $\omega$ ,而基函数可以作为某种广义测度,于是该式可以改写为:

$$c_i = \frac{1}{\omega} \int_R \omega(x, y) f(x, y) du(x, y) \quad (7)$$

其中, $u(x, y)$ 是定义的测度,而该测度实质上就是定义在集 $R$ 上 $f(x, y)$ 的权值分布情况,这恰好反映了噪声信息和形状细节信息。因此研究新的测度定义,使得形状的噪声信号和形状细节信息能够区分开来,应该是线性组合类表达的关键。研究出这种测度以后,关于该测度的积分就相当于重新定义描述子的算法,计算的描述子就可以弥补对形状细节描述带来的缺陷。

### 6.2 空间关联关系的数学机理

空间关联关系分析方法,实质是计算局部空间点之间的关系。目前大多数算法都是试探性地定义一些参数,比如三角形面积、局部形状曲线到弦的距离等等,然后对一些数据集进行测试,以验证算法的优劣。但是,这些算法在数学上对形状检索和识别的具体机理则缺乏深入研究,这就使得该类表达方法处于数据验证再改进的迭代探索阶段,因此该类算法

对一些数据集可能有效,而对另外一些数据集,其性能可能表现得较差。因此,研究空间关联关系算法的数学机理是这类方法进一步发展的重要方向。

### 6.3 形状表达其他研究方法的思考

从线性组合类方法中关于描述子的计算中可以发现,线性组合方法本质上是积分形状表达方法的一种特例,因此线性组合方法具有对噪声鲁棒的特性。同样考查空间关联关系形状表达方法,也很容易发现其中很多算法是微分形状表达方法的特例,比如使用三角形夹角形状表达的实质是估算形状边界曲线的曲率。因此对任一类表达方法的研究,对其他类表达方法都有很大的改进作用。目前,深度学习算法也被引进到形状表达的研究中,如文献[37,47-50]利用深度学习技术,将一些已知类型的形状数据集输入卷积神经网络进行学习,以此应用到形状的检索和识别中,获得了较好的效果。进一步研究发现,底层的学习卷积核类似于边缘检测算子,而且随着神经网络层的增加,捕捉复杂特征,也就是说:第二层节点对应角点和其他边缘连接,第三层节点捕捉形状的相似问题,而第四层节点捕捉诸如鸟的腿或者狗的脸等特定特征。利用深度学习技术训练神经网络来检索和识别形状,其信息存放在神经网络节点的权值,那么这些权值究竟是什么仍然无法从物理上或者几何意义上进行解释,同时也无法基于这些权值重构形状曲线,自然也就无法从理论上对形状表达进行评估,这也是基于深度学习技术的神经网络表达所遇到的一般问题。

### 参考文献

- [1] KURNIANGGORO L, WAHYON O, JO K H. A survey of 2d shape representation: Methods, evaluations, and future research directions[J]. *Neurocomputing*, 2018, 300: 1-16.
- [2] ZHANG D S, LU G J. Review of shape representation and description techniques[J]. *Pattern Recognition*, 2004, 37(1): 1-19.
- [3] KAZMI I K, YOU L, ZHANG J J. A survey of 2d and 3d shape descriptors[C]//2013 10th International Conference Computer Graphics, Imaging and Visualization. Macau: IEEE Press, 2013: 1-10.
- [4] DUAN L J. A Survey of Shape Feature[J]. *Computer Science*, 2007, 34(8): 215-218. (in Chinese)  
段立娟. 形状特征的编码描述研究综述[J]. *计算机科学*, 2007, 34(8): 215-218.
- [5] DEMISSE G G, AOUADA D, OTTERSTEN B. Deformation based curved shape representation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(6): 1338-1351.
- [6] MASON J C, HANDSCOMB D. Chebyshev Polynomials[M]. Berlin: Chapman & Hall/CRC, 2003: 20-30.
- [7] HEIZER A, BARZOHAR M, MALAH D. Stable fitting of 2d curves and 3d surfaces by implicit polynomials[J]. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 2004, 26(10): 1283-1294.
- [8] WU G, ZHANG Y C. A novel fractional implicit polynomial approach for stable representation of complex shapes[J]. *Journal of Mathematical Imaging and Vision*, 2016, 55(1): 89-104.
- [9] OLIVEIRA A B D, SILVA P, DA C R, et al. A novel 2d shape signature method based on complex network spectrum[J]. *Pattern Recognition Letters*, 2015, 63: 43-49.
- [10] NIXON M, AGUADO A S. Feature Extraction and Image Processing for Computer Vision[M]. London: Academic Press, 2012: 100-121.
- [11] ZHANG G, MA Z M, NIU L Q, et al. Modified fourier descriptor for shape feature extraction[J]. *Journal of Central South University*, 2012, 19(2): 488-495.
- [12] WU G Y, ZHANG Y C. A new chebyshev polynomials descriptor applicable to open curves[J]. *Pattern Recognition Letters*, 2015, 62: 41-48.
- [13] WU H Y, YAN S L. Computing invariants of tchebichef moments for shape based image retrieval[J]. *Neurocomputing*, 2016, 215(26): 110-117.
- [14] Mukundan R, Ong S H, Lee P A. Image analysis by tchebichef moments[J]. *IEEE Transactions on Image Processing*, 2001, 10(9): 1357-1364.
- [15] PEE C Y, ONG S H, RAVEENDRAN P. Numerically efficient algorithms for anisotropic scale and translation tchebichef moment invariants[J]. *Pattern Recognition Letters*, 2017, 92: 68-74.
- [16] CHEN Z, SUN S K. A zernike moment phase-based descriptor for local image representation and matching[J]. *IEEE Transactions on Image Processing*, 2010, 19(1): 205-219.
- [17] HU M K. Visual Pattern Recognition by Moment Invariants[J]. *IRE Transactions on Information Theory*, 1962, 8(2): 179-187.
- [18] SAJJANHAR A. A technique for similarity retrieval of shapes[D]. Melbourne: Monash University, 1997: 80-90.
- [19] ROUHANI M, SAPPA A D. Implicit polynomial representation through a fast fitting error estimation[J]. *IEEE Transactions on Image Processing*, 2012, 21(4): 2089-2098.
- [20] TEAGUE M R. Image analysis via the general theory of moments[J]. *Journal of the Optical Society of America*, 1980, 70(8): 920-930.
- [21] WANG K J, ZHANG H G, CHAI L S, et al. A comparative study of moment-based shape descriptors for product image retrieval[C]//2011 International Conference on Image Analysis and Signal Processing. Hubei: IEEE press, 2011: 355-359.
- [22] ROUHANI M, SAPPA A D, Boyer E. Implicit b-spline surface reconstruction[J]. *IEEE Transactions on Image Processing*, 2015, 24(1): 22-32.
- [23] EL-GHAZAL A, BASIR O, BELKASIM S. Farthest point distance: A new shape signature for fourier descriptors[J]. *Signal Processing, Image Communication*, 2009, 24(7): 572-586.
- [24] BAI X, RAO C, WANG X. Shape vocabulary: A robust and efficient shape representation for shape matching[J]. *IEEE Transactions on Image Processing*, 2014, 23(9): 3935-3949.
- [25] DALIRI M R, TORRE V. Robust symbolic representation for shape recognition and retrieval[J]. *Pattern Recognition*, 2008, 41(5): 1782-1798.

- [26] LIN C, PUN C M, VONG C M, et al. Efficient shape classification using region descriptors[J]. *Multimedia Tools and Applications*, 2017, 76(1): 83-102.
- [27] MOUINE S, YAHIAOUI I, VERROUST-BLONDET A. A shape-based approach for leaf classification using multiscale triangular representation[C]// *Proceedings of the 3rd ACM Conference on Multimedia Retrieval*. Dallas: ACM press, 2013: 127-134.
- [28] ALAJLAN N, RUBE I E, KAMEL M S, et al. Shape retrieval using triangle-area representation and dynamic space warping [J]. *Pattern Recognition*, 2007, 40(7): 1911-1920.
- [29] HU R, JIA W, LING H, et al. Multiscale distance matrix for fast plant leaf recognition[J]. *IEEE Transactions on Image Processing*, 2012, 21(11): 4667-4672.
- [30] HU R, JIA W, LING H, et al. Angular pattern and binary angular pattern for shape retrieval[J]. *IEEE Transactions on Image Processing*, 2014, 23(3): 1118-1127.
- [31] WANG B, GAO Y. Hierarchical string cuts: A translation, rotation, scale, and mirror invariant descriptor for fast shape retrieval[J]. *IEEE Transactions on Image Processing*, 2014, 23(9): 4101-4111.
- [32] LI M, YUAN B Z. 2d-lda: A statistical linear discriminant analysis for image matrix [J]. *Pattern Recognition Letters*, 2005, 26(5): 527-532.
- [33] BELONGIE S, MALIK J, PUZICHA J. Shape matching and object recognition using shape contexts[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, 24(4): 509-522.
- [34] ZHAO L, PENG Q Q, HUANG B Q. Shape matching algorithm based on shape contexts[J]. *IET Computer Vision*, 2015, 9(5): 681-690.
- [35] AN G, YU W. Captcha recognition algorithm based on the relative shape context and point pattern matching[C]// *2017 9th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)*. Changsha: IEEE press, 2017: 168-172.
- [36] BOUAGAR S, LARABI S. Discriminative outlines parts for shape retrieval[J]. *Journal of Visual Communication and Image Representation*, 2015, 33: 149-164.
- [37] ZHU Z T, WANG X G, BAI S, et al. Deep learning representation using autoencoder for 3d shape retrieval[J]. *Neurocomputing*, 2016, 204: 41-50.
- [38] ASADA H, BRADY M. The curvature primal sketch[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1986, 8(1): 2-14.
- [39] BERRADA F, ABOUTAJDINE D, OUATIK S E, et al. Review of 2d shape descriptors based on the curvature scale space approach[C]// *2011 International Conference on Multimedia Computing and Systems*. Ouarzazate: IEEE press, 2011: 1-6.
- [40] ADAMEK T, O'CONNOR N E. A multiscale representation method for nonrigid shapes with a single closed contour [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2004, 14(5): 742-753.
- [41] HONG B, SOATTO S. Shape matching using multiscale integral invariants[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(1): 151-160.
- [42] HONG B W, PRADOS E, SOATTO S, et al. Shape representation based on integral kernels: Application to image matching and segmentation[C]// *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. New York: IEEE press, 2006: 833-840.
- [43] ANDREA C G. *Metrics of Curves in Shape Optimization and Analysis*[M]. Cham: Springer International Publishing, 2013: 205-319.
- [44] Younes L. Computable elastic distances between shapes[J]. *SIAM Journal on Applied Mathematics*, 1998, 58(2): 565-586.
- [45] YOUNES L. *Parametrized Plane Curves*[M]. Berlin: Springer Berlin Heidelberg, 2010: 1-42.
- [46] YOUNES L. Optimal matching between shapes via elastic deformations[J]. *Image and Vision Computing*, 1999, 17(5): 381-389.
- [47] GEMMEKE J F, ELLIS D P W, FREEDMAN D, et al. Audio set: An ontology and human-labeled dataset for audio events[C]// *2017 IEEE International Conference on Acoustics, Speech and Signal Processing*. LA: IEEE press, 2017: 776-780.
- [48] ZEILER M D, FERGUS R. Visualizing and understanding convolutional networks[M]// *Computer Vision ECCV 2014*. Berlin: Springer, 2014: 818-833.
- [49] LECUN Y, BENGIO Y S. Deep learning[J]. *Nature*, 2015, 521(7553): 436-444.
- [50] NASCIMENTO J C, CARNEIRO G. Deep learning on sparse manifolds for faster object segmentation[J]. *IEEE Transactions on Image Processing*, 2017, 26(10): 4978-4990.