

基于稀疏低秩描述的图像检索方法

陈刚¹ 岳晓冬² 陈宇飞¹

(同济大学企业数字化技术教育部工程研究中心 上海 200092)¹

(上海大学计算机工程与科学学院 上海 200444)²

摘要 使用颜色、形状、纹理等特征的基于内容的图像检索技术,将图像看作向量空间中的点,通过计算两点之间的某种距离来衡量图像间的相似度,然而在提取图像特征时相同类型的图像会出现不一致的特征,极大地影响了检索算法的准确率。针对该问题,提出一种稀疏低秩描述的多特征图像检索方法。通过对图像集的稀疏低秩描述,保持了相同类别特征的全局结构,同时也降低了对局部噪声的敏感度,增强了检索算法的鲁棒性。在 Corel 图像集上的检索实验结果表明,该方法较已有的基于内容的图像检索方法有更好的检索效果。

关键词 基于内容的图像检索,稀疏低秩描述,特征提取

中图分类号 TP391 文献标识码 A

Image Retrieval Method Based on Sparse Low-rank Representation

CHEN Gang¹ YUE Xiao-dong² CHEN Yu-fei¹

(Engineering Research Center for Enterprise Digital Technology, Ministry of Education, Tongji University, Shanghai 200092, China)¹

(School of Computer Engineering and Science, Shanghai University, Shanghai 200444, China)²

Abstract The content based image retrieval method extracts the color, textural, shape features of images, which can be represented in the feature space, with similarities among them obtained by some distance between feature vectors. Its accuracy critically depends on the feature vectors. However, images in same class will have different features. This paper presented an image retrieval method based on sparse low-rank representation. After the low-rank components of each set was recovered, both the global mixture of subspaces structure and the locally linear structure of the features were captured. The experimental results show that the method not only has a strong robustness to the unstable features, but also has a good retrieval performance.

Keywords Content based imageretrieval, Sparse low-rank representation, Feature extraction

1 引言

随着数字图像技术与互联网技术的快速发展,每天都有大量的图片产生,如何从图像数据库中快速、准确地找出所需图像成为近年来计算机视觉领域内研究的热点。基于文本的图像检索^[1] (textbased image retrieval) 技术是对图像的关键字进行检索,而关键字的标注需要大量的人力、物力,关键字不能完全描绘出图像中的全部内容,所以没法适应海量图像检索。

计算机视觉领域常用的视觉特征有颜色、纹理、形状等:颜色是最简单的图像特征,对图像中物体的大小、方向、旋转等的依赖性较小;纹理是指具有同源属性,但并非由单一颜色或亮度表示的视觉形态;形状特征可以将不同图像中的目标区分开。基于内容的图像检索^[2] (contentbased image retrieval) 技术就是依据图像的视觉特征信息来描述图像内容,将特征信息转换为特征向量,然后采用相关方法在图像数据中查找符合条件的图片。由于该方法是对“关键词找图”技术的突

破,能有效地对图像进行智能化管理和查询,因此针对它的研究最多。文献[3]为了保持图像中像素的结构信息,避免不同颜色间的相互影响,使用了一种基于模糊颜色直方图的方法。文献[4]对提取出来的特征,采用行聚类学习方式进行图像检索。文献[5]使用图像的颜色、纹理等多种底层特征进行检索。

从相同类型图像中提取出来的特征里经常会出现不稳定状况,这些不稳定特征会降低检索的准确率。因此本文提出一种基于稀疏低秩描述的图像检索方法,该方法能够去除图像特征中的噪声,使检索结果更准确。与传统图像检索方法直接使用图像特征的方式不同,稀疏低秩描述方法使用核范数最小化和 L1、L21 范数最小化,将特征修复问题转化为低秩矩阵去噪,并通过增广拉格朗日乘法 (Augmented Lagrange Multiplier, ALM) 得到有效求解。实验结果表明该方法具有较好的检索性能。

本文第 2 节介绍图像特征提取;第 3 节讨论稀疏低秩描述法进行去噪;第 4 节介绍本文图像检索方法;第 5 节是实验结果和讨论;最后进行总结。

到稿日期:2013-05-13 返修日期:2013-08-15 本文受国家自然科学基金(61103070),国家科技支撑计划课题(2012BAF10B12)资助。

陈刚(1981-),男,博士生,主要研究方向为图像处理、计算机视觉;岳晓冬(1980-),男,讲师,主要研究方向为数据挖掘;陈宇飞(1982-),女,讲师,主要研究方向为数字图像处理。

2 特征提取

2.1 颜色特征

大多数数字图像采用的是 RGB 三色存储,但 RGB 颜色空间可分辨性较差,不符合人眼对颜色理解。人眼对于 HSV 颜色空间中的 H 分量(色调)较为敏感, H 、 S 、 V 3 个分量相互独立且与人眼能够感知的颜色特性相对应,HSV 颜色空间更适合用来表达不同色彩间的差异^[6]。

颜色特征通常包括颜色空间、颜色深度、主色调、直方图。其中颜色直方图表示图像中颜色的联合分布,它具有计算效率高、对镜头位置的变化不敏感、表示形式紧凑和计算复杂度低等优点,在颜色相似性测量中是一种最简单、最常用的特征。因此,本文采用 HSV 颜色直方图来描述图像中的颜色特征。

图像的颜色特征提取首先把图像从 RGB 格式转化为 HSV 格式,接着使用非等间隔量化方法对图像 HSV 颜色空间进行量化,将色调 h 分成 16 份、饱和度 s 和亮度 v 分别分成 4 份。量化之后将 3 个分量合成一个 $I=16H+4S+V$, I 的取值范围为 $[0, 255]$ 。最后按照式子: $V_{color}(i) = n_i/N, i = 0, 1, \dots, L-1$, 对图像的颜色做直方图统计,得到一个 256 维的特征向量。其中 i 表示灰度级, L 表示灰度级种类数, n_i 表示图像中具有灰度级 i 的像素总数, N 表示图像总的像素数。

2.2 纹理特征

局部二值模式(Local binary pattern, LBP)^[7]用于描述图像中每个像素值和其邻域像素值的关系,它由于具有旋转不变性和灰度不变性等显著优点,可以作为一种简单有效的纹理描述方法。

LBP 的基本原理是将图像上的像素与其为中心的周围 8 邻域像素灰度进行比较,如果该中心像素灰度值小于周围像素,则该点值为 1, 否则为 0。按顺时针方向排列得到一个 8 位二进制数,将二进制码转换成十进制数即得中心像素的 LBP 值。如图 1 所示,图 1 中二进制码为 10101100, LBP 值为 172。LBP 具体计算方法可以用式(1)表示:

$$LBP(x, y) = \sum_{i=0}^7 s(p_i - p_c) \times 2^i \quad (1)$$

$$s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

式中, (x, y) 表示像素坐标, p_c 表示中心像素灰度, p_i 表示其邻域像素的灰度。图像中各像素的 LBP 值显示了其邻域的灰度分布情况,统计图像中不同 LBP 值出现的概率,就可以得到纹理直方图来描述该区域的纹理结构(即 LBP 描述符)。

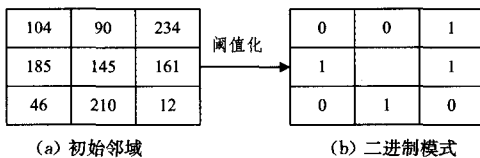


图 1 基本的 LBP 算子示例

2.3 形状特征

形状是比颜色、纹理更高级的特征,通常与图像中的目标有关,利用形状特征可以进一步地提高检索的准确性。Zernike 矩是基于 Zernike 多项式的正交化函数,具有旋转、比例和平移不变、冗余性小等优点,是一种广泛使用的形状特征描述符^[8]。在计算 Zernike 矩的时候不需要计算图像的边缘,可以很容易地构造出图像的任意高阶矩。

Zernike 多项式在极坐标下单位圆($x^2 + y^2 < 1$)内的定义如下:

$$V_{nm}(x, y) = R_{nm}(r) \cdot \exp(jm\theta) \quad (3)$$

式中, $R_{nm}(r)$ 是在单位圆内定义的一组正交多项式,即 Zernike 矩的径向多项式,具体定义如下:

$$R_{nm}(r) = \sum_s^{\frac{(n-|m|)}{2}} \frac{(-1)^s (n-s)!}{s! \left(\frac{n+|m|}{2} - s\right)! \left(\frac{n-|m|}{2} - s\right)!} r^{n-2s} \quad (4)$$

式中, n, m 为 Zernike 多项式的阶数, n 为非负整数, $n - |m|$ 为偶数,且 $n \geq |m|$ 。对一幅阶数为 n 、重复率为 m 的图像 $f(x, y)$, 其 Zernike 矩计算方式如下:

$$Z_{nm} = \frac{n+1}{\pi} \sum_x \sum_y f(x, y) \cdot V_{nm}(x, y) \quad (5)$$

本文对每幅图像计算其 10 阶 Zernike 矩^[9]得到 36 维向量,将其作为图像的形状特征描述符。

3 特征的稀疏低秩表示

提取出图像的颜色、纹理、形状特征,用向量形式表示,每张图像就能看作是向量空间中的点,图像集就可以用所有图像特征组成的子空间来描述。相同图像集内的图像之间虽然具有共同特性,但差别仍然比较大,其中的数据并非全部属于同一个特征子空间,而是由多个子空间共同生成^[10]。用一个子空间来描述所有特征数据会导致数据描述不准确,且数据经常会遇到噪声的干扰,需要将数据所属的多个子空间结构提取出来,同时修复被污染的数据。在图像集中虽然同一类图像之间差异通常较大,但各张图像仍然具有表示该类的共性特征。为此文献^[10]使用低秩描述(low-rank representation, LRR)^[11]来抽取图像集中的共性特征,剔除不稳定特征:

$$\begin{aligned} \min_{Z, E} \|Z\|_* + \lambda \|E\|_{2,1} \\ \text{s. t. } D = AZ + E \end{aligned} \quad (6)$$

式中, D 为图像集的特征向量集合, A 是能够张成 D 的过完备特征字典, λ 为正则项系数,用于平衡低秩部分与噪声部分。矩阵 Z 的核范数 $\|Z\|_* = \sum_{k=1}^m \sigma_k(Z)$ 就是它的奇异值之和。 $l_{2,1}$ 范数是计算矩阵 E 中每列的 l_2 范数之和: $\|E\|_{2,1} = \sum_i \sqrt{\sum_j (E(j, i))^2}$ ($E(j, i)$ 是矩阵 E 的第 (j, i) 个元素)。

LRR 的主要目的是捕获数据向量的全局子空间最小结构,同时降低全局噪声的影响,但 LRR 对于局部噪声就很敏感^[12]。实际上图像集内所有图像的特征在具有低秩性的同时,同一类图像的特征应具有局部线性结构。这样既能确保相同子空间里数据点的邻居最少,又能对局部噪声具有鲁棒性。本文使用稀疏低秩描述(Sparse low-rank Representation)来构造图像特征的数据结构:

$$\begin{aligned} \min_{Z, E} \|Z\|_* + \eta \|Z\|_1 + \lambda \|E\|_{2,1} \\ \text{s. t. } D = AZ + E \end{aligned} \quad (7)$$

式中, $\|Z\|_1$ 用于描述相同类别图像的特征对应的局部线性结构, l_1 范数是计算矩阵 Z 中所有元素绝对值之和: $\|Z\|_1 = \sum |Z(j, i)|$, 参数 η 用于稀疏部分对结果的影响。取数据 D 本身作为特征字典,问题(7)可转为求解如下凸优化问题:

$$\begin{aligned} \min_{J, L, E} \|J\|_* + \eta \|L\|_1 + \lambda \|E\|_{2,1} \\ \text{s. t. } D = DJ + E \\ Z = J, Z = L \end{aligned} \quad (8)$$

考虑到问题(8)实际上是一个凸优化问题,可以使用增广拉格朗日乘子法^[13]来求解,而且文献^[13]证实如果 μ 值选择得合适,问题(8)的解与下面最小化问题的解相同:

$$\min_{J,L,E,Z,Y_1,Y_2,Y_3} \|J\|_* + \eta \|L\|_1 + \lambda \|E\|_{2,1} + \text{tr}(Y_1^T(D - DZ - E)) + \text{tr}(Y_2^T(Z - J)) + \text{tr}(Y_3^T(Z - L)) + \frac{\mu}{2} (\|D - DZ - E\|_F^2 + \|Z - J\|_F^2 + \|Z - L\|_F^2) \quad (9)$$

式中, Y_1 、 Y_2 和 Y_3 是拉格朗日乘子, μ 是调控参数。整个增广拉格朗日乘子法的计算过程可以概况为算法1。虽然算法1中步骤1、步骤2和步骤4都是凸优化计算,但是它们都有各自不同的求解方式,步骤1采用奇异值阈值运算符(singular value thresholding operator)求解^[14],步骤2通过文献^[15]中的软阈值运算符(soft-thresholding operator)来求解,步骤4通过文献^[11]中的定理3.2来求解。求得式(8)的最优解(Z^* , E^*)后,得到图像集稀疏低秩约束条件下的特征描述: $V = D \times Z^*$,消除了不稳定特征中的噪声,有利于后面图像检索中相似性的准确度量。

算法1 ALM算法求解问题(8)

输入:特征集合D,调控系数 λ, η 。

初始化: $Z = J = L = 0; E = 0; Y_1 = Y_2 = Y_3 = 0; \mu = 10^{-6}; \max_{\mu} = 10^{10}; \rho = 1.1; \epsilon = 10^{-8}$

While 未收敛 do

1. 固定其他变量,更新J:

$$J = \arg \min \frac{1}{\mu} \|J\|_* + \frac{1}{2} \|J - (Z + \frac{Y_1}{\mu})\|_F^2$$

2. 固定其他变量,更新L:

$$L = \arg \min \frac{\eta}{\mu} \|L\|_1 + \frac{1}{2} \|L - (Z + \frac{Y_3}{\mu})\|_F^2$$

3. 固定其他变量,更新Z:

$$Z = (2I + D^T D)^{-1} (J + L + D^T (D - E) + \frac{D^T Y_1 - Y_2 - Y_3}{\mu})$$

4. 固定其他变量,更新E:

$$E = \arg \min \frac{\lambda}{\mu} \|E\|_{2,1} + \frac{1}{2} \|E - (D - DZ + \frac{Y_1}{\mu})\|_F^2$$

5. 更新拉格朗日乘子:

$$Y_1 = Y_1 + \mu(D - DZ - E)$$

$$Y_2 = Y_2 + \mu(Z - J)$$

$$Y_3 = Y_3 + \mu(Z - L)$$

6. 更新参数:

$$\mu = \min(\rho\mu, \max_{\mu})$$

7. 判断收敛条件:

$$\|D - DZ - E\|_{\infty} < \epsilon \& \& \|Z - J\|_{\infty} < \epsilon \& \& \|Z - L\|_{\infty} < \epsilon$$

End while

输出: $Z^* = Z, E^* = E$

4 基于内容的图像检索

本文采用颜色直方图来表示图像颜色特征,采用局部二值模式来表示图像纹理特征,用Zernike矩来表示图像的形状特征,接着使用稀疏低秩描述法剔除特征向量里的噪声,然后通过比较颜色、纹理、形状向量之间的距离,来判断两幅图像的相似程度,并给出与所查图像最相似的若干图像。

假设查询图像为Q,其颜色特征为 V_c^Q ,纹理特征为 V_t^Q ,形状特征为 V_s^Q ;图像库中的图像为P,颜色特征为 V_c^P ,纹理特征为 V_t^P ,形状特征为 V_s^P 。图像特征用向量形式来表示,这样就能把图像看作是向量空间中的点,通过计算两点之间的某种距离来衡量图像间的相似度^[16]。本文图像相似度采用

欧氏距离(Euclidean)^[16]来度量特征向量之间的距离,并根据不同向量设定不同的权重值。查询图像Q与图库中第j张图像之间的欧氏距离为:

$$d(P_j, Q) = \sqrt{\sum_{i=1}^n (V_{pi} - V_{qi})^2} \quad (10)$$

式中,n为特征向量维数。利用式(10)计算两幅图像颜色、纹理和形状特征相似度,分别为 d_c, d_t 和 d_s ,则任意两幅图像的相似度为:

$$S(P, Q) = W_c d_c + W_t d_t + W_s d_s \quad (11)$$

W_c, W_t 和 W_s 分别为颜色、纹理和形状特征距离的加权值, $W_c + W_t + W_s = 1$ 。在一次检索结束后,可利用反馈环节对3个权重进行调整,以达到最好的检索效果,最后根据相似性的大小,将最相似的图像作为结果输出。

5 实验结果与分析

本文的所有实验都是在CPU为P4 Dual-Core 2.5GHz,内存为2G,操作系统为Windows7的计算机上进行,实验环境为Matlab2009b。实验用的数据集是从http://wang.ist.psu.edu/docs/related/下载的SIMPLIcity系统的标准图像集^[17]。这些图像都是384 * 256和256 * 384的JPG格式,包括汽车、恐龙、马、鲜花、建筑等共10类,每类含有100幅,共有1000幅图像。

图2是不同方法对非洲土著居民进行检索的前20幅图像,每幅图中左上角的图像为查询图像,其余图像是查询结果。从左到右、从上到下,相似度依次减小。其中图2(a)是只使用颜色特征来检索的结果;图2(b)是使用颜色、纹理、形状3种特征,但不特征中的噪声进行处理的检索结果;图2(c)是使用LRR剔除特征向量里的噪声的检索结果;图2(d)是本文方法的检索结果。

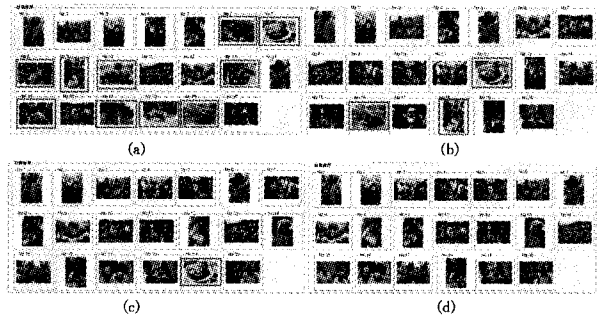


图2 不同方法对土著居民检索的结果

从图中可以看出,本文方法检索的20幅图像都是土著居民,准确率为100%,如图2(d)所示;颜色直方图的检索结果有10幅图像不正确,检索准确率为50%,如图2(a)所示;使用3种特征检索结果中有3幅图像不正确,检索准确率为85%,如图2(b)所示;LRR剔除噪声后的检索结果有1幅图像不正确,检索准确率为95%,如图2(c)所示。证明本文算法能将颜色直方图、局部二值模式、Zernike矩阵有效结合,并有效地去除了特征向量里的噪声,获得了更高的准确率。

为了进一步说明本方法的性能,还将本文算法与文献^[4, 5]中的算法做了实验对比,采用平均查准率来评价检索结果。具体步骤为:从图像库中每一类图像中随机抽取10幅图像,并对每张图像进行查询,然后分别求出每类图像的平均查准率。查准率的定义为:

$$\text{precision} = \frac{s}{s + u} \quad (12)$$

式中, s 为一次查询中检索到的相关图像的数目, u 为一次检索过程中检索到的不相关的图像数目。图 3 为平均查准率直方图, 从图中可以看到本文所提出的方法较文献[4, 5]中的方法有更好的检索效果。另一种衡量查询结果的方法是查准率-查全率曲线。查全率的定义为:

$$recall = \frac{s}{s+v} \quad (13)$$

v 为图像库中和检索图像相关但未被检索到的图像数目。查准率和查全率这两个指标越高, 说明检索性能越好, 但从这两个指标的定义可以看出, 它们之间相互矛盾: 查准率越高则查全率越低, 查全率越高则查准率越低。在相同查全率条件下, 查准率越高性能越好。图 4 显示了所有查询图像平均查准率和平均查全率绘制的曲线, 从图中的曲线可以看出, 本文的算法比其他两种方法效果要好一些。

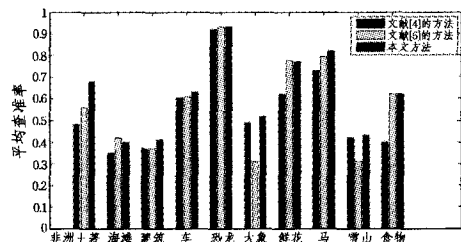


图 3 平均查准率对比

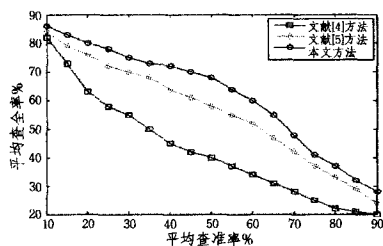


图 4 平均查准率-查全率曲线

结束语 针对基于内容的图像检索方法中提取出来的特征向量不稳定以及对局部噪声敏感的问题, 本文提出一种用稀疏低秩描述图像集特征的图像检索方法。对图像集进行检索时, 首先提取图像的颜色、纹理、形状特征, 然后通过稀疏低秩描述的方法获得稳定特征, 在此图像集描述的基础上进行图像检索。实验表明本文方法具有较好的检索性能, 可将原来不稳定的特征转换为稳定的特征, 并且在保持了全局子空间最小结构的同时, 降低了局部噪声对于检索结果的影响。

参考文献

[1] Datta R, Joshi D, Li J, et al. Image retrieval: ideas, influences, and trends of the new age[J]. ACM Computing Surveys, 2008, 40(2): 1-60
 [2] Han C H, SimKwee Bo. Real-time face detection using AdaBoost algorithm[C] // Control, Automation and Systems ICCAS, 2008

International Conference on Seoul, Korea, 2008; 1892-1895
 [3] Konstantinidis K, Gasteratos A, Andreadis I. Image retrieval based on fuzzy color histogram processing[J]. Optics Communications, 2005, 248(15): 375-386
 [4] Zakariya S M, Ali R, Ahmad N. Combining visual features of an image at different precision value of unsupervised content based image retrieval[C] // Computational Intelligence and Computing Research, 2010 IEEE International Conference. 2010; 1-4
 [5] Hiremath P S, Pujari J. Content based image retrieval using color, texture and shape features[C] // 15th International Conference on Advanced Computing and Communications, IEEE Computer Society, 2007; 780-784
 [6] Chen G, Peng R, et al. Pallet Recognition and Localization Method for Vision Guided Forklift[C] // Wireless Communications, Networking and Mobile Computing, 2012 8th International Conference on Shanghai, China, 2012
 [7] Ojala T, Pietikäinen M, Harwood D. A comparative study of texture measures with classification based on feature distribution [J]. Pattern Recognition, 1996, 29(1): 51-59
 [8] 王向阳, 李东明, 杨红颖. 基于 Zernike 色度分布矩的彩色图像检索算法[J]. 模式识别与人工智能, 2012, 25(2): 313-317
 [9] Wee C H, Paramesran R. On the computational aspects of Zernike moments[J]. Image and Vision Computing, 2007, 25(6): 967-980
 [10] Lv X, Chen G, Wang Z C, et al. Grassmannian Manifolds Discriminant Analysis Based On Low-Rank Representation for Image Set Matching[J]. Chinese Conference on Pattern Recognition, 2012(321): 17-24
 [11] Liu G, Lin Z, Yu Y. Robust subspace segmentation by low-rank representation[C] // 27th International Conference on Machine Learning, Haifa; National Science Foundation, 2010; 663-670
 [12] Zhuang L S, Gao H Y, Lin Z C, et al. Non-negative low rank and sparse graph for semi-supervised learning[C] // Computer Vision and Pattern Recognition, IEEE Computer Society, 2012
 [13] Lin Z, Chen M, Wu L, et al. The augmented Lagrange multiplier method for exact recovery of a corrupted low-rank matrices[J]. Mathematical Programming, 2009
 [14] Cai J F, Candes E J, Shen Z. A singular value thresholding algorithm for matrix completion[J]. SIAM Journal on Optimization, 2010, 20(4): 1956-1982
 [15] Hale E T, Yin W, Zhang Y. Fixed-point continuation for l1-minimization: methodology and convergence [J]. SIAM Journal on Optimization, 2008(19): 1107-1130
 [16] Nguyen G P, Worring M, Smeulders A W M. Similarity learning via dissimilarity space in CBIR[C] // 8th ACM International Workshop on Multimedia Information Retrieval. 2006; 107-115
 [17] Li J, Wang J Z. Real-time computerized annotation of pictures [C] // ACM Multimedia Conference. 2006; 911-920
 [18] 黄丽雯, 汪鑫, 王涛. 一种基于形状特征的颅颌面 X 片图像检索方法[J]. 重庆理工大学学报: 自然科学版, 2013, 27(6): 72-75

(上接第 278 页)

[3] Hautamaki V, Karkkainen I. Outlier detection using k-nearest neighbor graph[C] // Proc. 17th IEEE Int. Conf. on Pattern Recognition, 2004, 3: 430-433
 [4] Angiulli, F, Palopoli L. Detecting outlying properties of exceptional objects [J]. ACM Transaction on Database Systems, 2009, 34(1): 62-74
 [5] Richard J, Chris C. Fuzzy-rough nearest neighbor classification and prediction [J]. Theoretical Computer Science, 2011, 412

(42): 5871-5884
 [6] Pandya D H, Upadhyay S H, Harsha S P. Fault diagnosis of rolling element bearing with intrinsic mode function of acoustic emission data using APF-kNN[J]. Expert Systems with Applications, 2013, 40(10): 4137-4145
 [7] Xu Yong, Zhu Qi, et al. Coarse to fine K nearest neighbor classifier [J]. Pattern Recognition Letters, 2013, 34(9): 980-986
 [8] 符永铨, 王意浩. DKNNS: 面向延迟敏感型应用的可扩展精确分布式 K 近邻搜索算法研究[J]. 中国科学(信息科学), 2012, 42(5): 561-577