

## 基于数据共性的鲁棒特征表示<sup>①</sup>

岳金朋<sup>②\*</sup> 史忠植\* 张伟\*\* 胡宏\* 张建华\*\*

(\* 中国科学院计算技术研究所智能信息处理重点实验室 北京 100190)

(\*\* 中国科学院大学 北京 100049)

(\*\*\* 中国科学院计算技术研究所前瞻研究实验室 北京 100190)

**摘要** 为了提高图像检索的性能,研究了图像变化对视觉特征鲁棒性的影响,提出了一种新的鲁棒特征表示方法。该方法首先挖掘鲁棒特征的数据共性,即在不同图像变化条件下提取的鲁棒特征之间的共性,然后基于数据共性进行特征的二进制码表示。在特征挖掘阶段,根据特征的视觉信息和在向量空间下的相似性来挖掘数据共性。在特征表示阶段,对具有共性的特征进行离线学习,通过局部保持哈希(LPH)将具有共性的数据表示为相似的二进制码。该方法由于将特征提取过程中的潜在信息即数据共性与特征表示技术相融合,因此能够更好地应对复杂图像变化。实验表明,在图像检索应用背景下该方法的精度比现有方法提高 6% 以上。

**关键词** 图像内容分析, 视觉特征提取, 鲁棒特征表示, 数据共性

## 0 引言

在计算机视觉领域,基于内容的图像检索利用视觉特征提取技术,将图像检索问题转化为特征的相似性查询问题,其中,视觉特征的鲁棒性决定了图像检索的性能。在实际应用中,多种因素导致了图像变化,比如拍摄条件(角度、光照)的不同、传输过程中的噪声以及对图像的各种改变(比如拉伸、剪切)等。而现有的视觉特征检测方法无法应对所有的图像变化<sup>[1]</sup>。当图像变化剧烈时,视觉特征的鲁棒性下降,引起检索系统的性能下降。为此学者提出了对视觉特征提取过程进行扩展的方法<sup>[2-4]</sup>,通过提取并保存各种图像变化条件下的特征,来提高特征鲁棒性。比如多采样多视觉词典树<sup>[3]</sup>检测特征点后,对特征点的局部区块在多个尺度、多个位置下采样并构建多棵视觉词典树,保存了特征的互补信息。再如尺度不变特征转换(scale-invariant feature transform, SIFT)特征具有全尺度不变性,但应对仿射变化的能力有限。Morel 提出了仿射性(affine)SIFT(ASIFT)方法<sup>[4]</sup>将 SIFT 特征扩展到全仿射空

间下,即在多种仿射参数对应的模拟图像中提取特征,提高了特征鲁棒性。多分辨率最稳定极值区域(maximally stable extremal region, MSER)<sup>[2]</sup>为解决 MSER 对模糊攻击敏感的问题,将 MSER 扩展到尺度空间下。但是,扩展方法导致特征数量剧增,不仅增加了噪声和冗余,也加剧了空间消耗,因而严重降低了图像检索系统的性能<sup>[5]</sup>。ASIFT 把对模拟图像提取的特征组成集合,没有对特征进行选择。现有的特征挖掘方法,比如 ASC<sup>[5]</sup>和 S-MSER<sup>[6]</sup>,只是简单地对特征进行统计,没有深入挖掘特征的共性。一方面,挖掘特征的共性可减小冗余特征的数量,降低存储成本,另一方面,基于数据共性对特征进行表示,可借助特征提取过程中的视觉信息来提高特征的鲁棒性。基于此研究思路,本文提出了基于数据共性的鲁棒特征表示方法,并在公开数据集上验证本文方法的性能。

## 1 基于数据共性的鲁棒特征挖掘

特征扩展的目的是提高特征应对各种图像变化的鲁棒性,比如仿射性尺度不变特征转换(ASIFT)<sup>[4]</sup>

① 973 计划(2013CB329502),国家自然科学基金(61035003,60933004)和 863 计划(2012AA011003)资助项目。

② 男,1982 年生,博士;研究方向:视觉感知计算,行为检测;联系人,E-mail:yuejp@ics.ict.ac.cn

(收稿日期:2013-06-05)

是为了提高尺度不变特征转换(SIFT)应对视角变化的能力,多分辨率最稳定极值区域(MSER)<sup>[2]</sup>提高了MSER应对模糊攻击的能力。本节首先介绍基于SIFT的特征扩展算法,然后介绍在该扩展空间下的特征挖掘算法。尽管本文只针对SIFT特征扩展提出了挖掘算法,需要指出的是,该算法适用于其他的特征扩展,比如多分辨率MSER。

### 1.1 特征扩展

在计算机视觉领域,仿射变化定义为各向异性 的不同比例拉伸: $u(x,y) \rightarrow u(ax+by+e,cx+dy+f)$ ,具有 $a,b,c,d,e,f$ 这6个自由度。 $e$ 和 $f$ 表示和平移变化相关的两个自由度。文献[5]提出通过 $a,b,c,d$ 这4个自由度的变化,来模拟全仿射空间中不同视角下的图像变化,进而对模拟图像提取特征实现扩展。

将仿射变化用矩阵 $A$ 表示,可得到公式

$$\begin{aligned} A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \lambda \begin{bmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{bmatrix} \begin{bmatrix} t & 0 \\ 0 & 1 \end{bmatrix} \\ \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix} \quad (1) \end{aligned}$$

其中, $\lambda$ 为缩放比例, $\psi$ 表示摄像机围绕其中心光轴的旋转角,参数 $t$ 由光轴的倾斜纬度 $\theta$ 决定, $\varphi$ 则表示其倾斜经度。通过模拟摄像机光轴的倾斜经度和纬度,可以自动生成不同视角下的仿射图像。对于每幅图像,最多可以产生覆盖观察球面的41幅仿射模拟图像<sup>[5]</sup>,称为该特定图像 $I$ 的“全仿射空间”<sup>[4]</sup>。

为了实现SIFT特征到全仿射空间的扩展,ASIFT对每幅仿射模拟图像提取SIFT特征。ASIFT将提取的所有特征组成集合,并没有根据区分力、鲁棒性精简特征。由于扩展后引入了大量冗余的特征,因此下一步对扩展后的特征进行挖掘。

### 1.2 特征挖掘

对式(1)中的参数 $\lambda,\psi$ 以及由倾斜纬度 $\theta$ 决

定的 $t$ 和 $\varphi$ ,分别取不同的参数值<sup>[4,5]</sup>,从而生成仿射模拟图像集合。对该图像集合提取的SIFT特征中,有的只在某一幅模拟图像中出现,而有的则具有局部稳定性,即在多个模拟图像下重复出现并保持稳定。本文的目标是获取具有局部稳定性的特征。具体来说,对两个从不同模拟图像中提取的位置和形状相似的SIFT特征区块,如果其描述子向量也相近,则称两个区域是局部稳定的。局部稳定的SIFT特征区块在视觉上和描述子的向量空间下都具有相似性。形式化描述为:

假设共有 $N$ 幅模拟图像,对模拟图像 $I'$ 提取的特征集合记为 $RS(I') = \{RS_1(I'), RS_2(I'), \dots\}$ ,在图像全仿射空间下提取的所有特征集合记为 $RS = \{RS(I'_0), \dots, RS(I'_N)\}$ 。令特征 $RS_i(I'_j)$ 的描述子向量为 $v_{ij}$ ,特征 $RS_i(I'_{j'})$ 的描述子为 $v_{ij'}$ ,其中 $j' > j$ 。如果满足以下条件:

(1) 位置和形状相似,即:将 $I'_j$ 和 $I'_{j'}$ 映射至原图像 $I$ 后, $RS_i(I'_{j'})$ 与 $RS_i(I'_j)$ 的中心点距离相差不超过5个像素,区块的大小(以包含像素的个数衡量)相差不超过5%

(2)  $d(v_{ij} - v_{ij'}) \leq t_d$ ,其中 $d$ 为距离度量标准,本节采用欧氏距离。

则称 $RS_i(I'_j)$ 是局部稳定的, $RS_i(I'_{j'})$ 是 $RS_i(I'_j)$ 的冗余特征。此时, $RS_i(I'_j)$ 在模拟图像 $I'_{j'}$ 中重复出现,删除 $RS_i(I'_{j'})$ 不影响特征匹配效果。分析图像全仿射空间下特征的局部稳定性并在此基础上获取高稳定性特征的过程,称为特征挖掘。

数据共性的定义比局部稳定性的定义更加严格,只有同时满足在多种不同的图像变化条件下具有局部稳定性的一组特征,才称为具有数据共性。在实现算法时,如果特征向量在超过5幅模拟图像上具有局部稳定性,则称其具有数据共性。图1展示了对图像的SIFT特征进行扩展以及挖掘的过程。

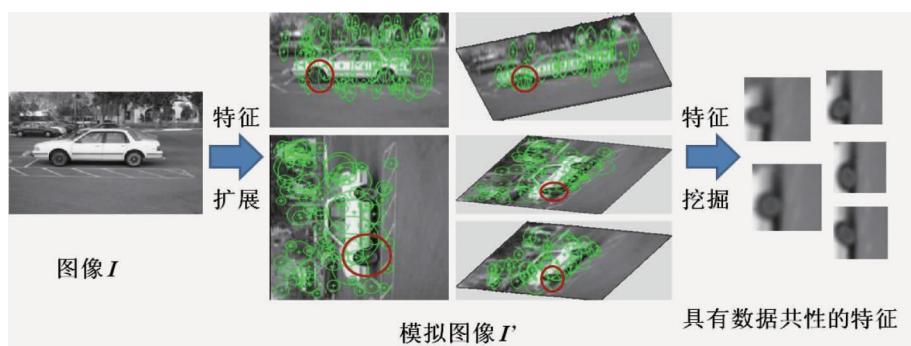


图1 基于数据共性的鲁棒特征挖掘示例

冗余特征在模拟图像中重复出现的次数表征了特征的稳定程度, 出现次数越多特征越稳定。特征挖掘算法在  $RS$  集合中统计冗余特征出现的频率并据此对特征进行排序, 从而获得具有数据共性的特征。

综上所述, 基于数据共性的鲁棒特征挖掘算法步骤如下:(1) 特征扩展, 在全仿射空间下提取模拟图像的特征;(2) 特征挖掘, 在视觉空间以及向量空间下挖掘具有数据共性的特征。下一节将具有数据共性的特征表示为二进制码, 从而利用共性信息来提高特征表示的鲁棒性。

## 2 基于数据共性的二进制码特征表示

二进制码具有占用内存空间少、匹配速度快的优点, 是目前常用的特征表示方法。基于学习的二进制码索引方法可分为两大类: 有监督的学习方法<sup>[7,8]</sup>和无监督的学习方法<sup>[9,10]</sup>。无监督的学习方法比如谱哈希(spectral Hashing, SH)<sup>[9]</sup>, 其目标是保持原始向量之间的欧氏距离, 并没有考虑特征在视觉上的相似性。而视觉上相似的特征在向量空间下可能并不相似。在这种情况下, 欧氏距离不能有效地反映特征之间的相似度。Christoph 提出了有监督的学习方法——线性判别分析哈希(linear discriminant analysis Hash, LDA Hash)<sup>[8]</sup>。该方法用线性判别分析来保持训练集上的相似关系。但是 LDA 考虑的是数据空间的全局结构, 并没有考虑局部结构或者流形结构。当数据分布在流形结构上时, 基于全局结构的学习方法与数据的内在结构并不一致, 导致生成低质量的二进制码。为此, 本文提出了局部保持哈希(locality preserving Hash, LPH)二进制码学习方法, 通过保持数据共性即数据空间的流形结构来生成鲁棒而紧凑的二进制码。

### 2.1 二进制编码模型

局部保持哈希(LPH)将特征之间的数据共性作为监督信息, 通过求解最优化问题, 得到基于数据共性的二进制编码模型。

给定一个训练图像库, 在该库上执行基于数据共性的鲁棒特征挖掘算法, 得到  $N$  组具有数据共性的特征。局部保持哈希将特征向量映射为二进制码, 并在映射过程中保持数据共性。形式化描述如下: 令  $X = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$  表示  $d$  维特征向量集合,  $h(\mathbf{x}_i)$  代表一个哈希函数, 局部保持哈希的目标是生成二进制向量  $\mathbf{y}_i = [h_1(\mathbf{x}_i), \dots, h_k(\mathbf{x}_i)]$ , 使得  $\mathbf{y}_i$

在海明空间下保持特征之间的数据共性。

将数据  $\mathbf{x}_i$  映射为  $k$  位的二进制码  $\mathbf{y}_i$ , 共需要  $k$  个哈希函数进行  $k$  次海明嵌入。将数据映射为二进制码的哈希函数可归结为矩阵投影和二值化两个部分, 定义如下:

$$\mathbf{y}_i = \text{sgn}(\mathbf{P}\mathbf{x}_i^T + \mathbf{T}) \quad (2)$$

其中  $\mathbf{P}$  是  $k * d$  的矩阵,  $\mathbf{T}$  是  $k * 1$  向量。式(2)中投影矩阵  $\mathbf{P}$  和阈值向量  $\mathbf{T}$  是问题求解的关键。

### 2.2 模型求解

本节首先介绍投影矩阵的求解过程, 即通过定义基于数据共性的相似度矩阵, 求解有监督的局部保持投影, 得到保持特征数据局部结构的投影矩阵, 然后介绍保持数据共性的阈值生成算法。

#### 2.2.1 投影矩阵求解算法

局部保持哈希将数据共性作为监督信息, 通过有监督的学习来训练投影矩阵  $\mathbf{P}$ 。首先根据数据共性定义相似度矩阵。令相似度矩阵为  $\mathbf{W}$ , 特征向量  $\mathbf{x}_i$  和  $\mathbf{x}_j$  之间的相似度为  $w_{ij}$ , 本文定义了两种计算  $w_{ij}$  的方式, 如下式所示:

$$w_{ij} = \begin{cases} \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/\beta) & \text{如果 } \mathbf{x}_i \text{ 和 } \mathbf{x}_j \text{ 具有数据共性} \\ 0, & \text{其他} \end{cases} \quad (3)$$

$$w_{ij} = \begin{cases} 1, & \text{如果 } \mathbf{x}_i \text{ 和 } \mathbf{x}_j \text{ 具有数据共性} \\ 0, & \text{其他} \end{cases} \quad (4)$$

式(4)将矩阵  $\mathbf{W}$  的元素设置为 0 或 1, 即具有数据共性的特征之间相似度为 1, 否则相似度为 0。

实验部分将比较两种相似度计算方式对特征表示结果的影响。相似度矩阵不需要人工标注, 而是在对训练集进行特征挖掘时基于数据共性信息自动生成。

令数据  $\mathbf{x}$  在矩阵  $\mathbf{P}$  上的投影为  $\mathbf{y}'$ , 即  $\mathbf{y}' = \mathbf{P}\mathbf{x}^T$ 。训练的目标是使得相似度越大的数据之间投影值也越相似, 相似度较小的数据之间投影值距离较远。因此, 本文将目标函数定义为

$$\operatorname{argmin} \sum_{i,j} \|\mathbf{y}'_i - \mathbf{y}'_j\| w_{ij} \quad (5)$$

正如前面所讨论的, 视觉上相似的特征在向量空间下不一定相似, 因此欧氏距离不能有效地表达数据共性。为了揭示数据的非线性结构, 需采用流形的学习方法。传统的流形学习方法, 比如 ISO-MAP<sup>[11]</sup> 和 LE<sup>[12]</sup>, 只定义在训练集上, 没有显式的映射, 因此学习得到的模型不能推广到新的测试集。局部保持投影(LPP)方法<sup>[13]</sup>对 LE 进行了线性化, 定义了投影矩阵, 利用线性逼近的思想将数据从高

维空间投影到低维流形空间。LPP 不仅能保持数据的局部结构,还能将训练结果推广到测试集上。由于 LPP 是无监督的,局部保持哈希根据式(3)和(4)定义的相似度矩阵,通过有监督 LPP 来生成投影向量  $P$ ,可得到以下推导:

$$\sum_{i,j} (y'_i - y'_j)^2 w_{ij} = 2\text{tr}(\mathbf{P}^T \mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{P})$$

其中  $\text{tr}$  为矩阵的迹。 $\mathbf{D}$  是对角矩阵,  $D_{ij} = \begin{cases} \sum_{k=1}^n w_{ik}, & i = j \\ 0, & i \neq j \end{cases}$ 。 $\mathbf{L} = \mathbf{D} - \mathbf{W}$  是拉普拉斯矩阵。矩阵

$\mathbf{D}$  中,  $D_{ii}$  对应着第  $i$  个数据点, 值越大反映出第  $i$  个数据点越重要, 投影后的  $y'_i$  就越重要。因此, 为了消除在投影时的缩放问题, 引入以下约束:

$$\mathbf{Y}' \mathbf{D} \mathbf{Y}'^T = 1 \Rightarrow \mathbf{P}^T \mathbf{X} \mathbf{D} \mathbf{X}^T \mathbf{P} = 1$$

最小化问题转换为

$$\begin{aligned} & \text{argmin}(\text{tr}(\mathbf{P}^T \mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{P})) \\ & \text{s.t. } \mathbf{P}^T \mathbf{X} \mathbf{D} \mathbf{X}^T \mathbf{P} = 1 \end{aligned}$$

最后, 通过求解公式

$$\mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{P} = \lambda \mathbf{X} \mathbf{D} \mathbf{X}^T \mathbf{P} \quad (6)$$

中的广义特征值问题, 可得到投影矩阵  $\mathbf{P}$ 。

综上, 局部保持哈希基于数据共性定义了相似度矩阵以及最小化目标函数, 然后根据改进后的 LPP 算法求解最小化问题, 从而得到投影矩阵  $\mathbf{P}$ 。

## 2.2.2 阈值求解算法

阈值用来对投影后的数据进行二值化, 从而实现二进制编码。本文中阈值生成算法与求解投影矩阵是两个独立的问题, 分别进行求解。通常情况下, 从查询效率的角度考虑, 阈值的选取尽量对数据空间均衡划分。本节则同时考虑数据空间划分对查询效率和查询精度的影响, 分别定义了划分熵值(partition entropy, PE) 和错误划分率(false ratio, FR)。通过减小错误划分率, 在二进制编码过程中尽量保持数据之间的共性。

一方面, 阈值的选择应使得划分熵值最大, 从而提高查询效率。划分熵值越大, 说明对数据空间的划分越均衡, 即子空间中索引的数据越均匀, 从而使最终生成的二进制码也具有更好的查询性能。阈值  $t_i (1 \leq i \leq d)$  的划分熵值形式化定义为

$$\begin{aligned} \text{PE}(t_i) &= -p_1 \log p_1 - p_2 \log p_2 \\ (p_1 &= \Pr(y' \leq t_i), p_2 = \Pr(y' > t_i)) \end{aligned} \quad (7)$$

其中  $y'$  表示数据  $\mathbf{x}$  在  $\mathbf{P}_i$  上的投影值。

另一方面, 选择的阈值应减小错误划分率, 尽量保持数据共性, 从而提高查询精度。错误划分率指

的是将具有数据共性的特征划分到不同子空间下的概率。通过减小错误划分率, 可以在划分后的空间下尽量保持数据共性, 从而实现保持数据共性的二进制编码。令  $y' = \mathbf{P}_i \mathbf{x}^T$  表示数据  $\mathbf{x}$  在投影向量  $\mathbf{P}_i (1 \leq i \leq d)$  上的投影值, 对  $y'$  进行二值化后生成二进制码的第  $i$  位。具体来说, 错误划分率定义为阈值  $t_i$  介于具有数据共性的投影数据之间的概率, 即:

$$\text{FR}(t_i) = \Pr(\min\{Y'\} < t_i < \max\{Y'\}) \quad (8)$$

其中  $Y'$  表示一组具有数据共性的特征数据在  $\mathbf{P}_i$  上的投影值。

令参数  $\alpha$  和  $\beta$  分别代表划分熵值和错误划分率的权重, 则阈值  $t_i$  可通过求解公式

$$\arg \min(\alpha \text{FR}(t_i) - \beta \text{PE}(t_i)) \quad (9)$$

中的最优化问题得到。

如果  $\alpha = 0$ , 则子空间中的数据被均等划分, 划分后的数据空间完全均衡。如果  $\beta = 0$  则错误划分率的最小值为 0, 此时  $t_i$  大于数据空间下的最大值或小于最小值, 对数据空间的划分是没有意义的。因此,  $\beta$  取值应大于 0。本文把划分熵值和错误划分率设置为相同的权重, 即  $\alpha = \beta = 1$ 。

实验部分对式(9)进行近似求解, 即在训练集上尝试不同的投影值作为阈值, 并计算相应的错误划分率和划分熵值, 然后选择使式(9)取最大值时的阈值。对投影矩阵  $\mathbf{P}$  中的每个  $\mathbf{P}_i (1 \leq i \leq d)$ , 依次按照公式计算阈值  $t_i$ , 从而得到阈值向量  $\mathbf{T}$ 。

在训练数据集上计算得到  $\mathbf{P}$  和  $\mathbf{T}$  后, 在新的测试集上根据矩阵  $\mathbf{P}$  对数据集进行投影, 然后根据阈值向量  $\mathbf{T}$  对投影数据进行二值化, 从而实现保持数据共性的二进制编码。

## 3 实验结果及分析

本节首先介绍实验采用的评测数据集, 然后介绍本文方法的离线训练过程, 最后将本文方法与现有方法作对比。

### 3.1 两个评测数据集

(1) 从 INRIA Holiday 图像库中提取的具有数据共性的 SIFT 特征区块

在 Holiday 库中选取不同场景下的图像, 并对这些图像执行基于数据共性的鲁棒特征挖掘算法。通过该算法生成了约 15000 组具有数据共性的特征, 共 91400 个。该特征集合作为评测数据集, 如

图2所示。查询数据采用以下方法生成:在评测数据集中随机选取500组具有数据共性的特征,并在每组中选取一个作为查询,共得到500个查询数据。评测基准(baseline)则是与查询数据有数据共性的所有特征。

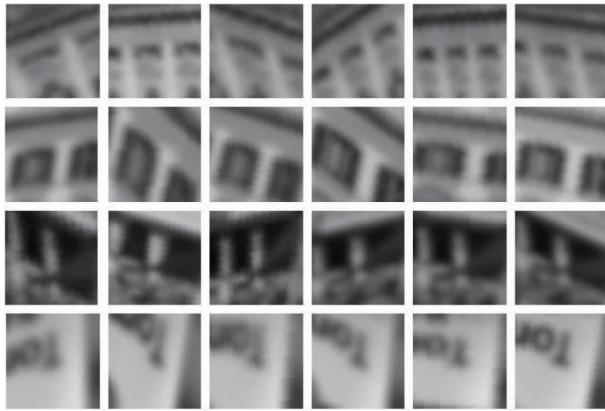


图2 具有数据共性的特征区块数据集示例

#### (2)加入了干扰图像的Ukbench数据集<sup>[14]</sup>

Ukbench对1550个物体从不同的视角在光照等条件下分别拍摄4次,共10200幅图像。实验加入了10000幅干扰图像,并用SIFT方法提取了7.8M的特征数据。查询数据则从每组图像中随机选取一幅,共500幅。评测基准是查询图像对应的组中的四幅图像。图3为数据集示例。

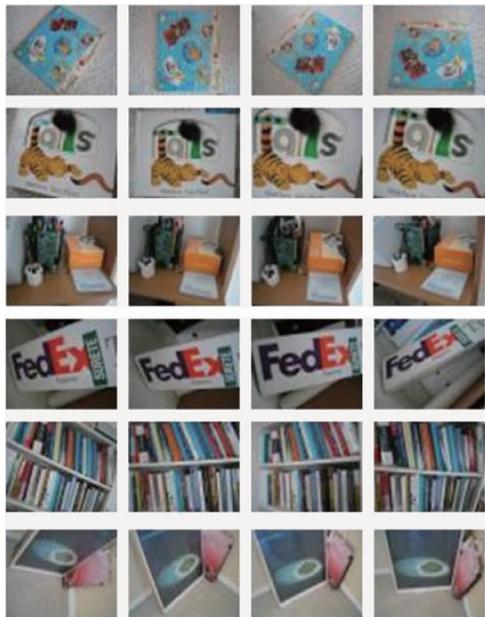


图3 Ukbench图像库示例

### 3.2 离线训练

训练数据是从具有数据共性的SIFT特征区块数据集中随机选取的10000组具有共性的特征。训练过程包含两步,即计算投影矩阵和阈值。相似度矩阵在特征挖掘过程中根据特征之间的数据共性构造。实验采用有监督的LPP来计算投影矩阵。计算得到投影矩阵后,分别在投影数据的每一维上计算阈值。图4以投影数据的第一维为例,依次选取每个数据的投影值作为阈值,并计算错误划分率和划分熵值。通过计算满足式(9)的阈值,兼顾了数据共性和均衡划分。图4中实心圆点标注了在第一维上错误划分率-划分熵值(FR-PE)取最优值时的阈值。

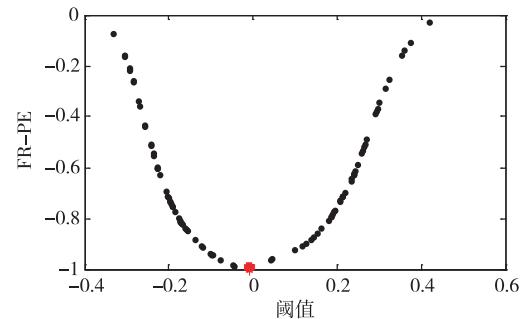


图4 最优阈值选择

如果重点考虑错误划分率,则需要在式(9)中增大 $\alpha$ 的权重。由于查询时对二进制码执行范围查询可以弥补错误划分率,因此本文将 $\alpha$ 和 $\beta$ 设置为相同的权重,以在查询效率和精度之间取得较好折中。

### 3.3 算法性能比较

首先在SIFT特征区块数据集上对本文方法进行验证。实验采用返回数据比例和召回率作为检索的评测标准。召回率较高而返回数据比例较低时,意味着检索到的正确结果多而噪声少。在上一节,通过离线训练得到了二进制编码的参数:投影矩阵和阈值向量。基于该参数,对从特征区块中提取的SIFT数据进行基于数据共性的二进制编码,即执行局部保持哈希(LPH),然后验证二进制码的检索性能。

在本次实验中,对局部保持哈希(LPH)和局部敏感哈希(locality sensitive Hashing, LSH)<sup>[15]</sup>、谱哈希(spectral Hashing, SH)<sup>[9]</sup>以及线性判别分析哈希(LDAHash)<sup>[8]</sup>的二进制编码方法作了比较。基于LSH的二进制编码方法采用随机投影,并假设数据

空间是均匀分布的。谱哈希(SH)是无监督的学习方法,其最优化函数的目标是在映射空间下保持数据的欧氏距离。LDAHash 是有监督的学习方法,利用了特征的视觉信息。但是 LDAHash 考虑的是数据的全局结构,不能有效揭示数据的流形结构。这三种方法分别代表了随机的方法,无监督的学习方法和有监督的针对全局数据结构的学习方法,能够全面地验证本文方法的有效性。实验将式(3)对应的特征表示方法记为 LPH,式(4)中相似度矩阵元素设置为 0 或 1 的特征表示方法标记为 LPH-1。

实验将二进制码位数设置为 64,在检索过程中将海明距离从 0 设置到 32,得到了本文方法和现有方法的性能比较结果,即返回数据比例和召回率的曲线图,如图 5 所示。

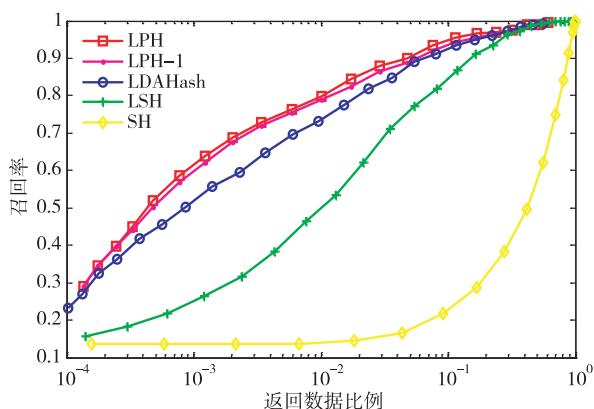


图 5 特征区块数据集上的性能比较

从图 5 可以看出,有监督的 LPH 和 LDAHash 与无监督的 SH 和 LSH 相比,性能有显著提升。LPH 产生的二进制码与 LDAHash 相比,具有更好的性能。这是因为 LPH 保持了数据的局部结构,其哈希函数充分利用了数据共性。通常情况下 SH 比 LSH 性能要好,但在本次实验中 SH 表现最差。这是因为在本数据集上,评测的基准是具有数据共性的特征,而不是欧氏空间下的最近邻。SH 的目标是保持欧氏空间下的近邻关系,因此导致了较低的性能。这也说明了在本数据集上,欧氏距离并不能很好地反映特征的相似度即数据共性。

下面的实验在具有复杂图像变化的 Ukbench 数据集上执行目标检索。该图像库在不同的光照、视角等条件下进行拍摄,对目标检索具有一定的挑战性,如图 3 所示。实验中二进制码位数设置为 32 位,范围查询的海明半径设置为 2。图 6 给出了结果集大小与召回率的关系。

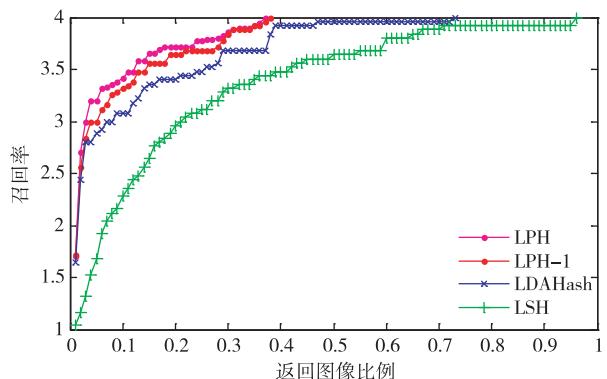


图 6 Ukbench 数据集上的性能比较

数据集共包含了 20200 幅图像和 7.8M 特征数据,LPH-1 的召回率比 LDAHash 平均高出 6.03%,比 LSH 平均高出 45.61%。在该数据集上,LPH 的表现优于 LSH 和 LDAHash。因此,LPH 在具有复杂图像变化的数据集上比上述两种方法更加鲁棒。此外,从图 5 和图 6 可以看出,LPH 比 LPH-1 性能略高,表明了采用式(3)定义的相似度矩阵略优于采用式(4)。

对图像库提取的 SIFT 特征数据占用了 7.89G 的存储空间,而 LPH 仅需要 78.2M。二进制码的内存消耗大约是原始特征数据的百分之一,这对大规模图像检索是至关重要的。在本次实验执行目标检索时,只关注二进制码之间的性能比较,因此对图像查询结果排序只采用了简单的投票机制。在实际系统中,可以采用几何一致性验证等后续处理步骤,来进一步提高检索精度。

## 4 结 论

本文提出了基于数据共性的鲁棒特征挖掘与表示方法,在视觉空间和向量空间下深入挖掘视觉特征之间的数据共性,并基于数据共性对特征进行二进制码表示。实验表明,局部保持哈希与现有二进制码方法相比能够更好地应对复杂图像变化。这是由于现有方法仅仅对高维数据进行压缩,并没有考虑特征提取过程。本文则挖掘特征提取过程中潜在的数据共性,并将数据共性与二进制码表示相融合,从而得到了更鲁棒的特征表示。

## 参考文献

- [1] Mikolajczyk K, Tuytelaars T, Schmid C, et al. A comparison of affine region detectors. *International Journal of Computer Vision*, 2005, 65(1-2):43-72

- [ 2 ] Forssen P E, Lowe D G. Shape Descriptors for maximally stable extremal regions. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV), Rio de Janeiro, Brazil, 2007. 1-8
- [ 3 ] Wu Z, Ke Q, Sun J, et al. A multi-sample, multi-tree approach to bag-of-words image representation for image retrieval. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV), Kyoto, Japan, 2009. 1992-1999
- [ 4 ] Morel J M, Yu G. ASIFT: A new framework for fully affine invariant image comparison. *SIAM Journal on Imaging Sciences*, 2009, 2(2): 438-69
- [ 5 ] Gao K, Zhang Y, Luo P, et al. Visual stem mapping and geometric tense coding for augmented visual vocabulary. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Providence, Rhode Island, 2012. 3234-3241
- [ 6 ] 张伟,高科,张勇东等. 基于鲁棒特征挖掘的图像副本检测系统. 计算机辅助设计与图形学学报, 2013, 25(6): 1-9
- [ 7 ] Salakhutdinov R, Hinton G. Semantic hashing. *International Journal of Approximate Reasoning*, 2009, 50(7): 969-78
- [ 8 ] Strecha C, Bronstein A, Bronstein M, et al. LDAHash: improved matching with smaller descriptors. *IEEE Trans on PAMI*, 2012, 34(99): 1-9
- [ 9 ] Weiss Y, Torralba A, Fergus R. Spectral hashing. In: Proceedings of Neural Information Processing Systems (NIPS), Canada, 2008. 1-8
- [ 10 ] Torralba A, Fergus R, Weiss Y. Small codes and large image databases for recognition. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Anchorage, USA, 2008. 1-8
- [ 11 ] Tenenbaum J B, Silva V and Langford J C. A global geometric framework for nonlinear dimensionality reduction. *Science*, 2000, 290(5500): 2319-2323
- [ 12 ] Belkin M, Niyogi P. Laplacian Eigenmaps for dimensionality reduction and data representation. *Neural computation*, 2003, 15(6): 1373-1396
- [ 13 ] He X, Niyogi P. Locality preserving projections. In: Proceedings of Neural Information Processing Systems (NIPS), Vancouver, Canada, 2003.
- [ 14 ] Nister D, Stewenius H. Scalable recognition with a vocabulary tree. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, New York, USA, 2006(2): 2161-2168
- [ 15 ] Datar M, Immorlica N, Indyk P, et al. Locality-sensitive hashing scheme based on p-stable distributions. In: Proceedings of the 20th Annual Symposium on Computational Geometry (SCG), New York, USA, 2004. 253-262

## Robust feature representation based on data commonalities

Yue Jinpeng \* \*\* , Shi Zhongzhi \* , Zhang Wei \*\* \*\*\* , Hu Hong \* , Zhang Jianhua \* \*\*

(\* Key Laboratory of Intelligent Information Processing, Institute of Computing Technology,  
Chinese Academy of Sciences, Beijing 100190)

(\*\* University of Chinese Academy of Sciences, Beijing 100049)

(\*\*\* Advanced Computing Research Laboratory, Institute of Computing Technology,  
Chinese Academy of Sciences, Beijing 100190)

### Abstract

To improve the performance of image retrieval, the effects of image variations on the robustness of visual features were studied, and then a novel robust feature representation method was proposed. The method mines robust features' data commonalities, the general characters among the robust features extracted under different image variations, and then represents features in binary codes based on the data commonalities. In the feature mining stage, the commonalities are obtained based on the visual information and the similarities of features in a vector space. In the feature representation stage, the commonalities of features are off-line learned and the data with commonalities are represented as similar binary codes by using the locality preserving Hash (LPH). This method can perform well under complex image variations, for it fuses data commonalities implicit in the feature extracting process with feature representation techniques. The experimental results show that the precision of the method can be improved by more than 6% compared with the existing methods in image retrieval applications.

**Key words:** image content analysis, visual feature extracting, robust feature representation, data commonalities