

# 基于内容的词典式图像库生成 Hash 编码的算法

杜文超, 苏胜君, 施伟斌

(上海理工大学 光电信息与计算机工程学院, 上海 200093)

**摘要:** 主要研究了基于内容的词典式图像库产生图像 Hash 编码的算法。建立一个词典结构的图像库, 提取图像库中所有的块特征向量, 并进行聚类分析, 对聚类后的每一部分进行二进制编码, 从而生成二进制索引(词典)库, 将待查图像引入图像库进行匹配, 生成一串二进制代码的路径, 即 Hash 编码。实验证明, 基于词典结构生成的 Hash 编码能够很好地反映视觉上的变化, 但是, 对视觉上相似的图像细微的变化反应并不剧烈, 能够较好地表达图像的整体情况。

**关键词:** CBIR; 特征向量; PCA; 图像 Hash; 词典式构造

中图分类号: TN911.79

文献标识码: A

## An algorithm of image hashing based on image dictionary of CBIR

DU Wen Chao, SU Sheng Jun, SHI Wei Bin

(School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 20093, China)

**Abstract:** This article is to examine an algorithm of image hashing in content-based image retrieval (CBIR). First of all, establish an image dictionary, and extract all pieces of characteristic vector from the image database and clustering them. Then generate a binary indexed (dictionary) database. Investigate processed image into images, generating a string matching of binary code that is Hash. Experimental results show that it can report well the change of sense of sight, but not insensitive to small local changes, it is suitable for distinguishing different images.

**Key words:** CBIR; feature vectors; PCA; image hashing; dictionary structure

图像 Hash 是用图像中提取的短序列来标识该图像, 可广泛用于内容认证和检索等领域。与数字水印不同, 提取 Hash 并不对图像数据进行修改, 因此能完好地保持原图像的品质。密码学 Hash 函数如 SHA21、MD5 对数据输入非常敏感, 任何细微的改变都会完全改变输出的 Hash 值, 故不适用于图像。实际应用中往往要对图像进行正常的数字处理如增强、几何变换等, 其内容并未发生实质性改变, 因此希望图像 Hash 保持不变。图像 Hash 函数  $H(\cdot)$  应该满足 2 个条件: (1) 图像 I1 和图像 I2 相似时,  $Hash(I1)$  和  $Hash(I2)$  以很大的概率相同或十分接近; (2) 图像 I1 和图像 I2 不相似时,  $Hash(I1)$  和  $Hash(I2)$  应以很大的概率不同。如用于检索, 则要求内容相似的图像有相近的 Hash, 内容无关的不同图像 Hash 值差异较大。

近年来, 图像 Hash 引起了广泛关注。Venkatesan 等人<sup>[1]</sup> 在图像小波域中用互不重叠的矩形进行伪随机分割, 取低频子带各矩形区域系数的平均值和高频子带矩形区

《微型机与应用》2010 年第 4 期

中系数的方差为特征值, 量化后输入 Reed2Muller 解码器产生中间 Hash, 再由线性编码得到最终 Hash 值。虽然小波系数统计特性较稳健, 却不能很好地反映图像内容, 特别是恶意产生的内容, 因此抵抗攻击的能力有限。另一种思路是寻求 DCT/DWT 系数间的近似不变性, 如针对 JPEG 压缩提取 2 个 DCT 系数之间的不变关系<sup>[2]</sup>, 该方法对 JPEG 较稳健, 但对其他攻击却缺乏稳健性。此外还可用 DWT 分解中父子节点间的统计依赖关系得到稳健的结构式数字签名, 克服某些 Hash 或脆弱水印的不足<sup>[3]</sup>。

本文提出了一种基于图像库的词典式结构产生的图像 Hash 算法, 在检索过程中引入了图像的 Hash, 对图像场景的局部变化不敏感, 能较好地地区分不同图像。

### 1 特征向量提取及分析

#### 1.1 特征向量的提取

对于基于内容的图像检索的研究, 其目的不是从图像数据库中查询具有完全相同视觉内容的图像, 那样是

欢迎网上投稿 www.pcachina.com 47

# 图形、图像与多媒体

Image Processing and Multimedia Technology

没有任何意义的。本文所讨论的图像 Hash 必须能允许视觉上略微的变化<sup>[4]</sup>,包括几何变形、对比度增强和有损压缩编码等。因此,要选取基于低层次特征的特征向量。除了选取色彩和纹理,不同的特征向量能充分描述视觉上的图像的各种不同的属性。因此,所需要的特征值应该包括统计矩的灰度值、彩色组成和纹理。把图像分成不同的小模块,计算这些模块的值,并且构造它们在降维后的特征向量,本文基于这一点引入 11 个模块的特征量,分别为灰度平均值  $\mu$ 、方差  $\sigma^2$ 、偏斜度  $s$ 、峰度  $k$ 、 $C_b$  的平均值  $m_a$ 、 $C_b$  的方差  $V_a$ 、 $C_r$  的平均值  $m_b$ 、 $C_r$  的方差  $V_b$ 、粗糙度  $S$ 、对比度  $C$  和定向度  $\theta$ 。

在实际应用中,不需要所有的图像模块都用来提取图像的特征量,载有很少信息量的模块应予摒弃。在应用中所选择信息的衡量标准是由矩阵的模块灰度层的奇异值决定的,本文选用  $SV1$  与  $SV2$  的比值  $R$ ,因为它反映了模块的粗糙度,带有少量有用信息的平坦模块具有数值较大的  $R$ ;相反,视觉上起伏较大的模块中  $R$  的数值较小。特殊情况是,信息为常量值的模块中  $R$  趋向无穷大,而模块仅包含噪声时,  $R=1$ 。因此,  $R$  可以作为第 12 个特征量,并且第  $i$  个模块的特征向量为:

$$u_i = [\mu \ \sigma^2 \ s \ k \ m_a \ V_a \ m_b \ V_b \ S \ C \ \theta \ R] \quad (1)$$

阈值  $T$  可以当作是决定某个模块是否可以用来提取特征向量的条件。当  $R < T$  时,模块有意义,反之则被忽略。

由于这 12 个特征向量并非从统计意义上是相互独立的,因此要求降维效果能够形成一个更紧密的模块描述符。采用在大量图像上主分量分析的方法(PCA)来帮助实现这个目标。

## 1.2 主分量分析

主分量分析(PCA)主要用于数据降维。对每一幅子块图像提取出一个 12 维的特征向量,对于整个图像子块库,可以组成一个很多维的特征向量,用这些特征向量来区分不同的子块图像。对于这些特征向量而言,向量里的某些特征值本身比较相似,用它来区分不同的子块特征向量,作用会非常小,而对这些数据进行计算和存储将会花费大量的时间和存储空间。所以笔者的目的是找那些变化大的元素。

提取一个 12 维的特征向量

$$u_i = [\mu \ \sigma^2 \ s \ k \ m_a \ V_a \ m_b \ V_b \ S \ C \ \theta \ R] \quad (2)$$

令

$X = (X_1, X_2, X_3, \dots, X_n)$ , 其中  $X_i = u_i^T$  为  $u_i$  的转置矩阵。

首先做标准化变换,即:

$$\begin{cases} Y_1 = a_{11} X_1 + a_{12} X_2 + a_{13} X_3 + \dots + a_{112} X_{12} \\ Y_2 = a_{21} X_1 + a_{22} X_2 + a_{23} X_3 + \dots + a_{212} X_{12} \\ \dots \\ Y_m = a_{m1} X_1 + a_{m2} X_2 + a_{m3} X_3 + \dots + a_{m12} X_{12} \end{cases} \quad (3)$$

记为  $Y = AX^T$ 。

对于  $Y$  还有以下几点要说明:

(1)  $Y_i$  和  $Y_j$  ( $i \neq j, i, j = 1, 2, \dots, m$ ) 是不相关的或正交的;

(2) 在  $X_1, X_2, \dots, X_m$  的所有线性组合中,  $Y_1$  的方差是最大的,即:  $A_1 X^T$  的方差是最大的;  $Y_2$  与  $Y_1$  不相关,而且其方差是第二大的;依此类推,  $Y_m$  与  $Y_1, Y_2, \dots, Y_{m-1}$  都不相关,并且其方差是第  $m$  大的<sup>[5]</sup>。即  $Y_i (i=1, 2, \dots, m)$  就是所要求的前  $M$  个主分量。

PCA 使得特征向量中各个元素相互独立。如图 1 所示,最大的 6 个主成分含有至少 98% 的能量。这 6 个主成分足以代表图像模块的信息。因此,保留初始 6 位主成分作为最终的特征向量。

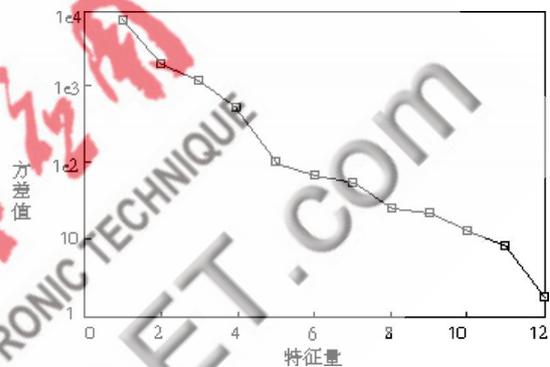


图 1 12 个特征分量所代表的图像信息值

## 1.3 特征向量的聚类分析

特征向量的提取是依赖图像的,提取后的每一个特征向量反映着一个图像子块的信息,视觉相似的图像其特征向量应接近;反之,视觉差异大的图像,特征向量之间距离应当大,对图像进行分类,可以看作是对这些特征向量进行的分类。选取了  $k$  均值聚类算法进行特征向量的聚类,聚类后,将  $n$  维的特征向量划分为  $k$  个聚类,同时满足同一聚类中的向量相似度较高,而不同聚类中的向量相似度较小。反映到图像块上,视觉上相似度大的图像集合在一起,而相似度小的在不同集合中。在聚类运算中,聚类相似度是利用各聚类中对象的均值所获得一个“中心对象”(引力中心)来进行计算的,保留每一个子集的中心特征向量。

基于本文采用的聚类算法,建立了一个四类的特征向量字典库,图像库中图像被分割成  $16 \text{ bit} \times 16 \text{ bit}$  大小的子块,建立子块图像库,用  $k$  均值算法对这些子块图像进行聚类分析。选取第一级聚类的 4 个子库中的典型视觉相似的子块,如图 2 所示。

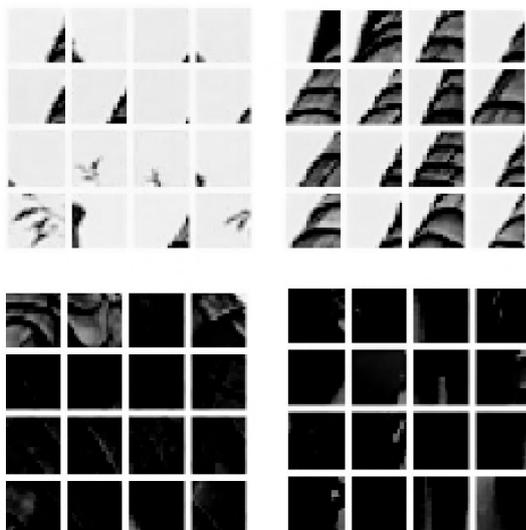


图2 第一级取4个子块库中典型视觉相似子块

2 词典图像库的 Hash 编码

2.1 图像词典式结构模型

在汉语词典中，通常通过汉字的某一特征如拼音、部首等来查询所要得到的汉字。同理，可以把一幅图像子块看成词典中的一个汉字，而把特征向量作为进行图像匹配的“拼音”或“部首”，从而得到待查子块图像所对应的“页码”，也就是找到匹配子块图像的路径，将其用二进制编码的形式表示出来，即图像的 Hash 码。借鉴查词典的方法，将其应用到图像的检索上，会达到和查字典一样的效果，大大提高了图像检索的准确度及效率。

图像词典是一个分级分类结构，对图像库中的每幅图片按照一定的大小分割成长、宽相等的子块，可以从每个子块中提取相应的特征向量，构成一个特征向量集。这些特征向量全依赖于图像的内容属性，因此，要对图像进行分类，只需要对这些特征向量进行聚类。为了尽量使内容上最相似的图像子块聚集在一起，要求构造更多的子库，如果在一级上构造太多，聚类的中间向量差距较小容易出现错误，而且词典显得庞大，繁琐。为了解决这些问题，本文提出了一种分级结构，如图3所示。

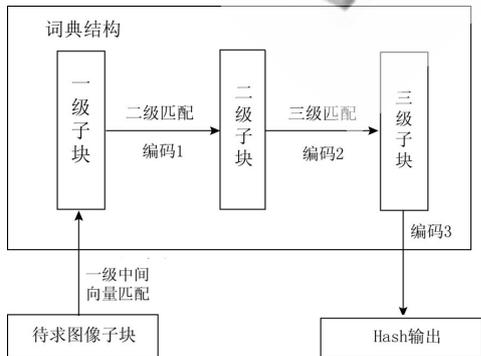


图3 图像词典式结构模型

2.2 生成 Hash 编码算法描述

本文研究的基于词典式结构的图像 Hsah 算法，主要分成两部分完成。

第一部分是对图像词典库的构造，为了计算方便，易于实现，建立一个  $M$  级词典库，每一级的聚类都为  $K$  类，即： $K=K_1=K_2=\dots=K_m$ 。建立这样一个图像词典库，依次要完成以下内容：

(1)对库中图像进行去噪及规格化处理。去噪主要是为了尽量减少噪声对图像内容特征的影响，规则化处理可以使图像具有相同的像素点个数；

(2)进行图像分割。把图像库中图像分割成大小相等的  $f \times f$  像素的子块，假设可以分割成  $N$  块，依次标记为  $1, 2, \dots, N$ 。对于图像分割到最后，行或者列不足  $f$  个像素点的情况，则可将不足部分舍弃。一般情况下，取  $f \times f = 8 \text{ bit} \times 8 \text{ bit}$  或者  $16 \text{ bit} \times 16 \text{ bit}$  的子块；

(3)对库中子块图像进行特征向量提取，由于图像本身数据量较大，并且经过实验验证，所提取的数据经过 PCA 主分量分析后得到其中 6 个向量就足以反映图像子块的内容，因此，得到一组 6 维的特征向量  $Feature(n)$ ,  $n=1, 2, \dots, N$ ；

(4)建造子块图像库的分级结构。由于提取的特征向量  $Feature$  是基于子块图像内容的数据，反映了子块的内容，要将子块库分成下一级的子库，只需对相应的特征向量进行聚类，每一类的中心特征向量记作  $FeaC(k)$ ,  $k=1, 2, \dots, K$ 。每一级分级结构如图4所示。对子库用二进制代码命名，依次记作： $Bin(k)$ ,  $k=1, 2, \dots, K$ 。

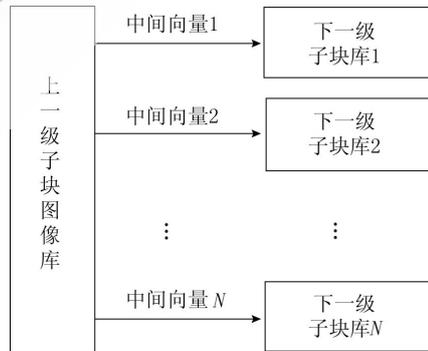


图4 词典式结构 Hash 编码生成模型

第二部分是 Hash 码的生成。

(1)依照构造图像词典中对图像的处理过程，对待求图像进行去噪、规格化处理、图像分割、特征向量提取及 PCA 主分量分析，得到待求图像子块的 6 维特征向量，记作： $Fea(i)$ ,  $i=1, 2, \dots, N$ 。

(2)图像子块匹配过程。待求图像所有子块依次对图像词典库中子块进行匹配，匹配过程分级进行，每一级采用相同的方法。对于不同的图像，匹配算法的效果有所不同，本文采用的匹配算法是基于欧几里德距离的查找，每一子块向量  $Fea(i)$  匹配到最相似的下一级子块库

## 图形、图像与多媒体

Image Processing and Multimedia Technology

的中间向量  $FeaC(k), k=1, 2, \dots, K$ , 对于第  $i$  级结构, 可以得到相应的子块库的二进制编码  $Bin_i(k), k=1, 2, \dots, K$ , 称为所在级子块库的 Hash 子码, 记作  $H(Bin_i)$ , 如图 5 所示。

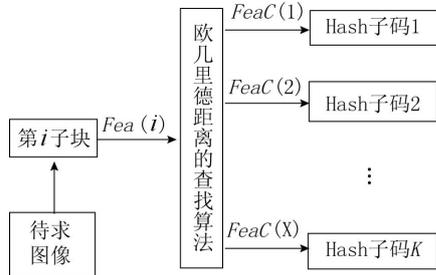


图 5 基于欧几里德距离的查找匹配的 Hash 编码生成而对于在第  $M$  级下所包含的子块, 不再进行聚类分析, 取  $H(Bin(n)), n=1, 2, \dots, N$ 。

所有子块匹配完成后, 把每一子块的 Hash 子码拼接起来, 即:

$Hash = \text{strcat}(H(Bin_1), H(Bin_2), \dots, H(Bin_M), H(Bin(n)))$   
 $\text{strcat}$  函数是进行字符串的拼接操作, 得到要求的图像的 Hash 编码。

### 2.3 Hash 编码的 Hamming 距离分析

Hamming 距离的大小代表着 2 个序列中存在差异对应位的数量, 在这里体现了图像 Hash 结果的差异。图像越相似, Hash 间的 Hamming 距离越小; 反之, 图像视觉上相差越大, Hamming 码越大。图 6 所示是一组视觉上相似的连续图像, 以图中左上角图像为样本, 随着视觉

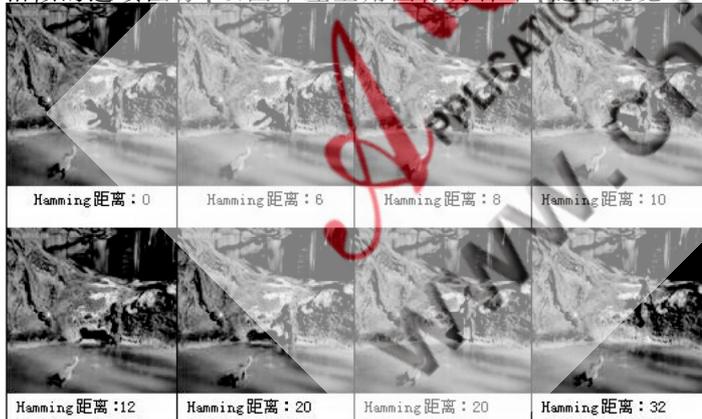


图 6 场景相似图像之间的 Hamming 距离

(上接第 46 页)

- [5] CHRYSAFIS C, ORTEGA A. Line based, reduced memory, wavelet image compression[J]. IEEE Trans on Image Processing, 2000, 9(3): 378-389.
- [6] SWELDENS W. The lifting scheme: a custom-design construction of biorthogonal wavelets [J]. Appl Comput Harmonic Anal, 1996, 3 (15): 186-200.

(收稿日期: 2009-12-04)

的上的变化, 图像 Hamming 距离会发生相应变化。但是, 由于词典库生成的 Hash 编码较长, Hash 编码差异所占 Hash 编码中的很少一部分, 因此图像 Hash 能较好地表达图像的整体情况, 场景中的局部变化不会导致 Hash 的剧烈改变。

庞大的图库、多样的分类、迅速的匹配将是今后词典式图像索引的发展方向与侧重点。对于大型的图像库, 要求具有运算效率高、精确度高、安全性高等特点。基于词典式结构产生的图像 Hash, 能精确反映图像内容特征, 视觉上越相似的图像, 其 Hash 编码越相似, 反之, 视觉上差异较大时, 则 Hash 编码差别越大。而当图像内容发生改变时, 图像的 Hash 也随之改变。

### 参考文献

- [1] VENKATESAN R, KOON S-M, JAKUBOWSKI M H, et al. Robust image hashing [C]. IEEE International Conference on Image Processing. Vancouver, BC, Canada, 2000, 3: 664-666.
- [2] LIN S, OZSU M T, ORIA V, et al. An extendible hash for multi-precision similarity querying of image databases [C]. Proceedings of 27th Very Large Data Bases Conference (VLDB) Rome, Italy, 2001: 221-230.
- [3] LU C S, LIAO H YM. Structural digital signature for image authentication [J]. IEEE Transactions on Multimedia, 2003, 5(2): 161-173.
- [4] TAMURA H. Textural features corresponding to visual perception. IEEE Transactions on Systems, Man and Cybernetics, 1978, 8(6): 460-472.
- [5] HANLEY J A, MCNEIL B J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology, 1982, 143:29-36.
- [6] JOHNSON S C. Hierarchical clustering schemes, Psychometrika, 1967, 2:241-254.

(收稿日期: 2009-10-14)

### 作者简介

杜文超, 男, 1985 年生, 硕士研究生, 主要研究方向: 图像处理与测试。

苏胜男, 女, 1970 年生, 博士生, 讲师, 主要研究方向: 图像处理与通信工程。

施伟斌, 男, 1961 年生, 副教授, 主要研究方向: 光学软件测试与通信。

### 作者简介

陈曙涛, 男, 1984 年生, 硕士研究生, 主要研究方向: 图像处理及 VLSI 设计;

罗桂娥, 女, 1962 年生, 教授, 主要研究方向: 信息融合技术、数字图像处理、智能仪器开发。