

一种适合 Web 图像检索的图像降维算法研究^{*}

鲁珂 赵继东 曾家智

(电子科技大学计算机科学与工程学院 成都 610054)

摘要 本文对现有 Web 图像检索技术现状进行了归纳,分析阐明了图像降维算法在基于内容的 Web 图像检索技术中的地位和作用。在介绍了几种经典图像降维方法后,重点介绍了国外近来提出的基于拉普拉斯特征值映射(LE)的图像降维算法。针对 Web 环境下图像检索必须进行剧烈降维的特点,本文进而对基于 KL 变换的主成分分析(PCA)算法和基于 LE 的图像降维算法进行了实验分析和比较。实验结果表明:对于需要进行维数剧烈缩减的 Web 图像检索来说,基于 LE 的图像降维算法可以获得最佳的效果。

关键词 Web 图像检索,降维,拉普拉斯特征值映射,主成分分析

Research on an Arithmetic of Image Dimensionality Reduction for Web Environment

LU Ke ZHAO Ji-Dong ZENG Jia-Zhi

(School of Computer Science and Engineering, UESTC, Chengdu 610054)

Abstract In this paper, the authors survey recent research of Web image retrieval, analyze and present the important effect of image dimensionality reduction on content-based Web image retrieval. After introducing several classical methods of image dimensionality reduction, the paper gives a detailed introduction on the arithmetic of image dimensionality reduction that based on Laplacian Eigenmap. Due to the characteristic that image retrieval needs drastically dimensionality reduction in Web environment, we performed several experiments to evaluate the effectiveness between the arithmetic that based on LE and the PCA arithmetic that based on KL transform. Experimental results show that the arithmetic hat based on LE give excellent performance for image retrieval in Web environment.

Keywords Web image retrieval, Dimensionality reduction, Laplacian eigenmap, PCA

1 引言

随着互联网的高速发展,网络上的各类资源日益丰富,尤其是直观形象的多媒体信息备受人们的青睐,因而急剧增加。人们对多媒体信息的检索需求也就随之而来。传统的网络信息检索主要集中于文字的检索,在多媒体方面的研究并不是很多。互联网上的多媒体以图像为主,因此网络图像的检索就成为了目前研究的热点。目前,已经有不少的搜索引擎提供网络图像的检索服务,如 Google、Ditto、PicSearch、Ixquick、Mamma、百度等。

互联网上图像的检索经历了两个阶段:第一阶段是以关键字为基础的检索。第二阶段是以图像自身的内容为基础的检索。

基于文本的检索技术发展已经成熟。如 Page-Rank 方法、概率方法、位置方法、摘要方法、分类或聚类方法、词性标注法等,不仅技术发展较为成熟,同时分析和实现的难度略小。但这种方法需要较多的人工参与,而且不同的人对于同一张图像的理解也不相同,这就导致对图像的标注没有一个统一的标准,因而检索的结果不能很好地符合用户的需求。

基于内容的检索不同于基于关键字的检索,它不需要过多的人工参与,而利用图像自身的特征(如颜色、纹理、形状等)来进行检索,具有较强的客观性。基于内容的图像检索建立在多媒体信息的内容语义上,能够更为客观地反映媒体本

质的特征。但是,因为图像的规模一般要大于纯粹的文本信息,因此,基于内容的图像检索在检索的速度和效率上要求更高。目前已有不少应用于实践环境的基于内容图像检索系统,如由 IBM 公司开发的最早商业化 QBIC 系统,以及由哥伦比亚大学研发的 WebSeek 系统、麻省理工学院研发的 Photobook 系统等。

基于内容的 Web 图像检索技术,首先需要从 Web 中剥离图像,组成图像集,对图像集中的各个对象进行基于内容的特征分析、相似度匹配。现有互联网的规模巨大,而且,图像特征向量的维数的数量级一般高达 10^2 ,而索引结构的性能是随着维数的增大而迅速降低的,在维数较高(>10)时,其效率还不如顺序扫描。在 Web 环境中,快速高效是系统自身资源和用户对图像检索系统的共同要求,如何在保证不丢失图像主要特征的前提下尽可能地缩减图像维数,就成为了基于内容的 Web 图像检索技术的关键。

2 当前图像检索技术中的常用降维算法

在基于内容的图像检索技术中,常用的两种缩减方法是 Karhunen-Loeve 变换(KLT)和按列聚类。

KL 变换和它的变种在面部识别、特征图像、信息分析和主成分分析(Principal Component Analysis)等领域内的应用已经得到了深入的研究。Ng 和 Sedighian^[1]曾采用特征图像的方法来实现维数缩减,Faloutsos 和 Lin^[2]提出了用于维数

^{*} 本文工作受到电子科技大学学科建设重点基金支持。鲁珂 讲师,博士生,主要研究领域为网络图像检索,网络体系结构;赵继东 讲师,博士生,主要研究领域为计算机网络;曾家智 教授,博士生导师,主要研究方向为计算机网络与通信。

缩减的 KLT 快速逼近算法。研究表明,大多数的实数集合(视觉特征向量)可以大量地缩减维数,并且对检索效果不会产生明显的影响。Chandrasekaran^[3]等开发了低秩奇异值分解(Singular Value Decomposition)更新算法,它能够被高效而稳定地应用于 KL 变换。由于图像检索系统是一个动态系统,不断会有新的图像加入到系统中,索引结构也需要相应地进行动态更新。奇异值分解算法就提供了这样一种动态更新索引结构的工具。

除了 KL 变换,聚类是另一种实现维数缩减的有力工具。聚类技术被广泛地应用于模式识别、语音分析、信息检索等领域。通常的聚类方法是将相似的对象(如模式、信号和文档等)聚合在一起,以实现识别或分组等功能,即所谓的按行聚类。同样,聚类也可以按列进行,从而缩减特征空间的维数^[4]。实验表明这种方法非常简单有效。

3 基于拉普拉斯特征值映射(LE)的图像降维算法

以上介绍的维数缩减技术相对成熟,而且已经得到了大量应用。但在维数剧烈缩减的情况下,上述方法并不能很好地保留图像的非线性结构,导致图像特征丢失,识别准确度大幅降低。拉普拉斯特征值映射(LE)^[5]是一种最近提出的用于流形学习的算法。它可以很好地保留图像空间的非线性的几何结构,在维数剧烈缩减时,这些保留下来的非线性结构有助于保持图像特征^[6]。因此,对于高维数缩减的图像检索来说,基于 LE 的图像降维算法可以获得最佳的效果。

下面,我们介绍如何利用 LE 来将图像映射到一个低维空间^[7]。我们考虑将一个权图映射为一条线段以使相连的点能尽可能地接近。在这里,顶点(如 $x_1, x_2, x_3, \dots, x_n$)表示图像处于一个高维欧几里得空间中。如果两点之间足够接近,那么它们之间就连上一条边。设定 $y = \{y_1, y_2, y_3, \dots, y_n\}^T$ 作为这里的映射。找到一个最佳映射的合理办法是使下面的目标函数值为最小。

$$\sum_{ij} (y_i - y_j)^2 W_{ij}$$

其中, W 是权矩阵,它的定义如下:

- i. 当 x_i 是 x_j 的 K-近邻或者 x_j 是 x_i 的 K-近邻时, $W_{ij} = 1$;
- ii. 否则, $W_{ij} = 0$;

这个最小值的问题可以简化为求:

$$\arg \min y^T L y \quad (\text{即求使得 } y^T L y \text{ 为最小的 } y)$$

$$\text{满足条件: } y^T D y = 1$$

这里, $L = D - W$ 是拉普拉斯矩阵。 D 是对角权矩阵,它的元素就是 W 的行(或列,因为 W 是对称矩阵)元素的数值和。拉普拉斯矩阵是对称矩阵且具有半正定的性质,它可以被看作是由图 G 的顶点定义的函数的一个算子。

当相邻的两点 x_i 和 x_j 被映射后反而远离时,含有我们选择的权矩阵 W_{ij} 的目标函数会产生很大的补偿。因此,最小化目标函数就是试图确保如果 x_i 和 x_j 紧邻,那么 y_i 和 y_j 同样紧邻。

该算法分步骤如下:

第一步:(创建邻接图)如果节点 i 是 j 的 K-近邻或者节点 j 是 i 的 K-近邻,则节点 i 与 j 之间用一条边连接。

第二步:(确定权重)只有节点 i 与 j 之间被一条边连接时,它们之间的权 $W_{ij} = 1$ 。

第三步:(特征值映射)如果图 G 按上述的构造方法被完全连通(如果没有完全连通,则将已连通的各局部用第三步的办法处理),按下面这个常见的方程计算特征值和特征向量。

$$L y = \lambda D y \quad (1)$$

假定 $y_0, y_1, y_2, \dots, y_{k-1}$ 是方程(1)的解,特征值按顺序排列为: $0 = \lambda_0 < \lambda_1 < \dots < \lambda_{k-1}$ 。我们将 $\lambda_0 = 0$ 及其对应的特征向量 y_0 剔除,依次用后面的 m 个特征向量来映射内嵌的 m 维欧几里得空间。

$$x_i = (y_1(i), y_2(i), \dots, y_m(i))$$

这里的 $y_j(i)$ 表示特征向量 y_j 的第 i 个元素。

4 基于 LE 的降维算法与 PCA 降维算法的比较试验

主成分分析(PCA)是基于 KL 变换进行降维的一种经典技术,它通过保留原始特征(n 维)的协方差矩阵的 m ($m < n$) 个占显著贡献的特征向量(这些特征向量是由原始特征线性变换得到的)来构建新的 m 维空间,将原始的 n 维数据映射到新的 m 维空间中来降维。为了比较 LE 算法与 PCA 算法应用于大型图像数据库时的效果,我们进行了实验比较。

在实验中使用的图像数据库由从 Corel 图库选出的包括 79 个语义类的 10000 个图像,构成一个大型的多种类的图集。检索出来的某个图像如果和求检图像属于同一个类别,我们就认为该图像是一个正确的结果。在我们的系统中使用了三种类型的颜色特征和三种类型的纹理特征。特征空间的维数是 435。我们设计了一种自动的反馈方案来模拟真实用户操作下的检索过程。在每一次循环中,系统从前 100 个匹配结果中选出头 3 个不正确的结果作为不相关的样例,同时选出最多 3 个正确结果作为相关的样例(上一次循环选出的相关样例在这一次不参与选择)。这些自动产生的反馈被加入到待检索样例集中,从而使检索结果最佳。为了评价算法的性能,我们定义了检索准确度:

$$\text{检索准确度} = (\text{头 } N \text{ 个检索结果中相关图像的个数}) / N$$

实验结果见图 1 和图 2。当特征空间的维数被 PCA 剧烈地缩减时,其性能下降极快。在第 0 次循环中,没有用户的相关反馈被引入。因此,检索结果只是基于低维空间中的特征表示。我们可以看到,LE 的表现比 PCA 好得多。在 PCA 产生的 2 维子空间中,3 次循环后头 10 个返回结果的准确度只达到大约 10%,而在 LE 产生的 2 维子空间中,在 1 次循环后头 10 个返回结果的准确度就超过了约 40%。这些结果显示 LE 特别适合于需要剧烈缩减维数的场合。当降维空间的维数是 100 时,在 PCA 子空间和 LE 子空间第 0 次循环的准确度分别是 36% 和 39%。

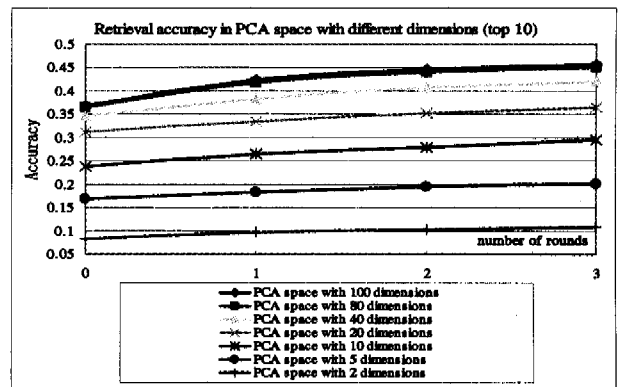


图 1 在 PCA 空间的检索效果

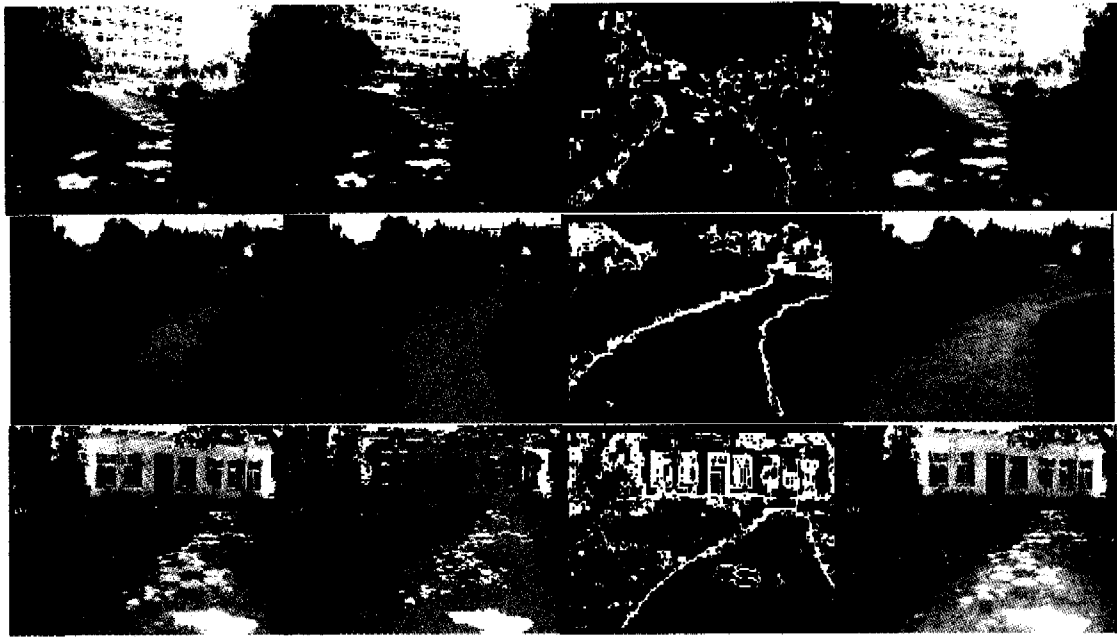


图9 实验结果(从左到右依次为原始图像,本质图像,过滤后的边缘图像和结果图像)

(上接第 256 页)

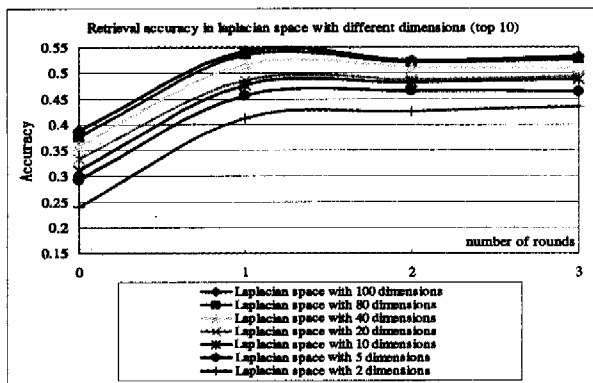


图2 在 LE 空间中的检索效果

总结 基于内容的 Web 图像检索技术是当前的研究热点。在 Web 环境中,快速高效是检索系统自身和用户的共同要求,因此,在保证不丢失图像主要特征的前提下尽可能地缩减图像维数是基于内容的 Web 图像检索技术的关键。拉普拉斯特征值映射(LE)是一种最近提出的用于流形学习的算法。它可以很好地保留图像空间的非线性的几何结构,在维数剧烈缩减时,这些保留下来的非线性结构有助于保持图像特征。从以上的实验分析我们可以发现,对于需要进行维数剧烈缩减的 Web 图像检索来说,基于 LE 的图像降维算法可

以获得最佳的效果。因此,我们认为,它是一种新型的最适用于 Web 图像检索的降维算法。

参考文献

- 1 Ng R, Sedighian A. Evaluating multi-dimensional indexing structures for images transformed by principal component analysis. In: Proc. SPIE Storage and Retrieval for Image and Video Database, 1996
- 2 Faloutsos C, Lin K-Ip (David). Fastmap: A fast algorithm for indexing, data-mining and visualization of traditional and multimedia datasets. In: Proc. Of SIGMOD, 1995, 163~179
- 3 Chandrasekaran S, Manjunath B S, Wang Y F, et al. An eigenspace update algorithm for image analysis. Comput. Vis., Graphics, and Image Proc. 1997
- 4 Salton G, McGill M J. Introduction to Modern Information Retrieval. McGraw-Hill Book Company, New York, 1982
- 5 Belkin M, Niyogi P. Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering. Advances in Neural Information Processing Systems 14, Vancouver, Canada, 2001
- 6 He Xiaofei, King O, Ma Wei-Ying, et al. Learning a Semantic Space from User's Relevance Feedback for Image Retrieval. IEEE Trans. On Circuit and Systems for Video Technology, 2003, 13 (1)