Computerized Medical Imaging and Graphics 39 (2015) 27–34

Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

journal homepage: www.elsevier.com/locate/compmedimag

Applying latent semantic analysis to large-scale medical image databases

Spyridon Stathopoulos*, Theodore Kalamboukis

Information Processing Laboratory, Department of Informatics, Athens University of Economics and Business, 76 Patission Str., 104.34 Athens, Greece

ARTICLE INFO

Article history:
Received 22 January 2014
Received in revised form 16 May 2014
Accepted 19 May 2014

Keywords:
SVD
LSA
CBIR
Feature selection
Data fusion
Text retrieval
Image retrieval
Classification

ABSTRACT

Latent Semantic Analysis (LSA) although has been used successfully in text retrieval when applied to CBIR induces scalability issues with large image collections. The method so far has been used with small collections due to the high cost of storage and computational time for solving the SVD problem for a large and dense feature matrix. Here we present an effective and efficient approach of applying LSA skipping the SVD solution of the feature matrix and overcoming in this way the deficiencies of the method with large scale datasets. Early and late fusion techniques are tested and their performance is calculated. The study demonstrates that early fusion of several composite descriptors with visual words increase retrieval effectiveness. It also combines well in a late fusion for mixed (textual and visual) ad hoc and modality classification. The results reported are comparable to state of the art algorithms without including additional knowledge from the medical domain.

1. Introduction

During the recent years, there is a continuous increase in digital medical imaging and the development of methods and tools for acquiring, archiving and communicating all forms of medical information. Moreover, images from many diagnostic modalities, such as Radiology, Visible light photography, Microscopy, etc., are actively used to support clinical decisions, medical research and education. The management of those very large heterogeneous collections is a challenging task. As a consequence, this has led to creating systems for storage retrieval management and distribution of medical images known as PACS (Picture Archiving and Communications Systems) [1]. However, according to [2], the users of PAC systems seem to consider the image manipulation, retrieval and comparison an important missing feature.

Meanwhile, immense amounts of medical image data are continuously produced during the daily clinical practice and research, especially from CT, MRI and PET modalities. It is evident that CBIR techniques must be able to process data of this scale efficiently.

Although text retrieval methods seem to provide acceptable results, it is evident from [3], that performance in image retrieval is still very far from being effective for several reasons: computational cost, scalability and performance. The use of textual information although is an effective way for describing semantic content, when used in image retrieval there is an additional overhead of manually annotating every image in the database. Images in the medical domain, however, are always associated with a short text (the radiologist report). Thus we are seeking for methods that fuse efficiently several low level visual features together with textual information to improve retrieval results. However, depending on the current context, such short descriptions of the images could be noisy due to the language ambiguity, such as polysemy and synonymy.

Latent semantic analysis [4] is a mathematical technique for extracting hidden relations of contextual usages of words in documents. Over the years, LSA has been used successfully in text retrieval with many applications [5], and seems to be an effective method to remove redundant information and noise from data. Although the computational complexity is high for large collections it is not prohibitive, mainly due to the sparsity of the textual data.

However, this is not the case in image retrieval, where images are represented with vectors in a compact and dense form. To our knowledge, the method has so far been used only with small size collections due to the high complexity of the approach. This article presents an effective and efficient approach of applying LSA by skipping the SVD analysis of the feature matrix. We show that our approach provides a much faster, stable and scalable solution. Also we show experimentally that our method fuses well several low level visual features together with textual information.

* Corresponding author. Tel.: +30 2108203575.
E-mail addresses: spstathop@aueb.gr (S. Stathopoulos), tzk@aueb.gr (T. Kalamboukis).

http://dx.doi.org/10.1016/j.compmedimag.2014.05.009

The rest of the paper is organized as follows. Section 2 presents a brief overview of previous relevant work and Section 3 describes our approach of applying LSA for retrieval in large-scale image collections. Section 4 describes the methods for representing images and the retrieval models used in conjunction with LSA. The following Section 5 provides a brief description of the datasets used for our experiments.

Section 6 demonstrates the effects of applying LSA to retrieval and classification. For the evaluation a wide range of experiments were applied using local and global descriptors in combination with state of the art methods, like Bag of Visual Words models (BoVW) based on SIFT and SURF features and Bag of Colors models (BoC) similar with the works in [6,7].

Finally, in Sections 7 and 8, results are presented from fusion of several different low-level visual descriptors separately an in combination with textual information. For comparison both late and early fusion methods were tested.

From all our experiments it is evident that LSA is superior to direct matching the image vectors and very fast in both, off-line pre-processing of the data and on-line at query evaluation.

2. Related work

LSA approach covers a wide area of applications, from text retrieval, cross language retrieval and classification to image annotation and retrieval. One of the earlier use of LSA for image retrieval, was in [8]. More work on fusing visual and textual information using LSA for multimedia retrieval and automatic annotation extraction shown promising results are presented in [9–12] to mention a few. However, in all these works, the size of the image collections used for experimentation were very limited, the largest contained 20,000 images, a rather small size for a search method to show its potentials.

In [13] we have proposed a “bypass” solution to the SVD problem which overcomes all its deficiencies, concerning time and space requirements and makes the method attractive for content-based image retrieval. From our experience of using the method in the ImageCLEF medical tasks over the past years [14,15], we perform here a systematic evaluation on the potentials of this approach testing several combinations of visual and textual information, into a semantic representation that is suitable for fast retrieval and classification on large scale image collections.

3. Our approach for large-scale LSA

LSA is based on the SVD factorization of a feature matrix, say C, $(m \times n)$ into a product $C = USV^T$, where U and V are orthogonal matrices, with the columns of U forming the right eigenvectors of the matrix CC^T, V^T the left eigenvectors of C^TC and S a diagonal matrix with singular values σ_i in the diagonal, $(\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r)$ and $r = rank(C)$ [16]. According to Eckart–Young theorem [17] a low rank approximation, $C_k = U_k S_k V_k^T$, for $k \ll r$, is the best approximation of matrix C.

In LSA the columns of the matrix $SV^T (= U^TC)$, are interpreted as the projections of the initial vectors (columns of C) into a space of k latent dimensions. For retrieval purposes, a user's query, q, is projected into the same kth dimensional space by $\hat{q} = U_k^T q$. The query $\hat{q}$ is evaluated using as scoring function the cosine similarity and results are presented in descending order. The SVD solution has a complexity of $O(m^2 n)$. In text retrieval we can take advantage of sparsity that significantly reduces the complexity to $O(cmn)$, where c is the average number of non zero elements per document $(c \ll m)$. In the case of CBIR, matrix C is dense and that makes the problem not feasible to solve for large collections. For example in one of our experiments matrix C, has a size of $4176 \times 306{,}000 \approx 10$ GB (in double precision) which makes SVD impossible to solve with our computer resources.

From our image representation we observe that the number, m, of visual features is of moderate size which lead us to an SVD-bypass solution and solving the eigenvalue problem (EVD) of the real symmetric matrix, $(m \times m)$ CC^T instead of the SVD(C). Thus our SVD-alternative algorithm is summarized as follows:

1 Solve the $m \times m$ eigenproblem $CC^TU = US^2$ for the k largest eigenvectors and corresponding eigenvalues.
2 Calculate the projections of the original images into the kth dimensional space $\hat{d}_j = U_k^T d_j$.
3 For a query q, calculate the similarity $score(\hat{q}, \hat{d}_j)$, by the cosine function, where $\hat{q} = U_k^T q$.

This solution though attractive has its drawbacks which may affect seriously the results. Indeed matrix CC^T requires less space and can be accommodated in memory, i.e. for our example requires only 140 MB. Furthermore the complexity of the EVD(CC^T) problem is $O(m^3)$ which, for $m \ll n$ makes the algorithm much faster. However this alternative to SVD solution maybe become unstable with the introduction of rounding error in the calculation of CC^T which may affect seriously the solution of EVD(CC^T) due to bad conditioning. In this case matrix, U, may be a bad approximation compared to that calculated from SVD. From the numerical point of view, it is usually better to work with the original data matrix and avoid the formation of cross products matrix. This can be especially important when small perturbations in the data can change the rank of the data matrix. In such cases the normal matrix will be much more sensitive to perturbations in the data than the original data matrix. However, this by no means implies that normal matrices should be avoided at any cost.

Matrices derived from the visual features of images are of integer type. Thus CC^T can be calculated block-wise with integer arithmetic ($CC^T = C_1C_1^T + \cdots + C_pC_p^T$). We split C into p blocks, $C = (C_1, \ldots, C_p)$ such that each C_i can be accommodated in memory and their multiplication is performed in integer arithmetic before any normalization. The complexity of the EVD approach $O(m^2n + m^3 + kmn)$, where the first factor corresponds to the calculation of CC^T, the second factor to the solution of EVD and the third factor to the projection of the initial vectors to a k dimensional space.

In the following we shall argument on the stability of the EVD solution.

3.1. On the stability of EVD

As it is known from Linear Algebra [16] for a symmetric matrix, A, we have a stable backwards algorithm which computes the solution of a slightly perturbed problem $(A + E)$. In this case the calculated eigenvalues λ_i' approximate the exact ones λ_i within machine accuracy. Furthermore the corresponding eigenvectors u_i' approximate u_i with full accuracy, $O(\epsilon)$, if σ_i are well separated no matter how small σ_i are. Thus if we could ensure that the eigenvalues are well separated then the approach offers a stable and very fast solution. In [5] it was found that, in the case of text retrieval the statistical significance of LSI dimensions which, are related to the squared of the singular values, follow a Zipf-law, indicating that LSI dimensions represent latent concepts in the same way as words in texts. Following the work in [5] it was found that the same result holds in the case of image retrieval with the eigenvalues of the matrix CC^T. The eigenvalues obey the relation $\sigma_i^2 = \alpha \times i^\beta$, where α and β are constants depending on the collection. This means that there are a few large and well separated eigenvalues and a large number of small and close together ones. However, in LSA we are interested only on the largest eigenvalues and corresponding eigenvectors. In