Mining User Queries with Markov Chains: Application to Online Image Retrieval

Konstantinos A. Raftopoulos, *Member*, *IEEE*, Klimis S. Ntalianis, Dionyssios D. Sourlas, and Stefanos D. Kollias, *Member*, *IEEE*

Abstract—We propose a novel method for automatic annotation, indexing and annotation-based retrieval of images. The new method, that we call Markovian Semantic Indexing (MSI), is presented in the context of an online image retrieval system. Assuming such a system, the users' queries are used to construct an *Aggregate Markov Chain* (AMC) through which the relevance between the keywords seen by the system is defined. The users' queries are also used to automatically annotate the images. A stochastic distance between images, based on their annotation and the keyword relevance captured in the AMC, is then introduced. Geometric interpretations of the proposed distance are provided and its relation to a clustering in the keyword space is investigated. By means of a new measure of Markovian state similarity, the *mean first cross passage time* (CPT), optimality properties of the proposed distance are proved. Images are modeled as points in a vector space and their similarity is measured with MSI. The new method is shown to possess certain theoretical advantages and also to achieve better Precision versus Recall results when compared to Latent Semantic Indexing (LSI) and probabilistic Latent Semantic Indexing (pLSI) methods in Annotation-Based Image Retrieval (ABIR) tasks.

Index Terms—Markovian semantic indexing, image annotation, query mining, annotation-based image retrieval

1 Introduction

Even though humans tend to associate images with high-level concepts, the current computer vision techniques extract from images mostly low-level features and the link between low-level features and high-level semantics of image content is lost. Neither a single low-level feature nor a combination of multiple low-level features has explicit semantic meaning in general. In addition, the similarity measures between visual features do not necessarily match human perception [1] and, thus, retrieval results of low-level approaches are generally unsatisfactory and often unpredictable. This is what is called the semantic gap: the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation. However, the retrieval process fails also due to the sensory gap: the gap between the object in the world and the information in a (computational) description assigned to a recording of that object. While the former gap brings in the issue of users' interpretations of images and how it is inherently difficult to capture them in visual content, the latter gap makes recognition from image content challenging due to limitations in recording and description capabilities. Currently, only 10 percent of online image files have a professional description (annotation). As a result, image search engines are only able to deliver precision of around 42 percent and recall of around 12 percent [2], while 60 percent of search engine visitors use at least two different search engines since they are not satisfied by the retrieved content. The most common complaint is that search engines do not recognize content semantics. Additionally, about 77 percent of searchers change keywords more than once because they cannot detect content of interest [3], [4].

Annotation-Based Image Retrieval (ABIR) systems are an attempt to incorporate more efficient semantic content into both text-based queries and image captions (i.e.. Google Image Search, Yahoo! Image Search). The Latent Semantic Indexing (LSI)-based approaches that were initially applied with increased success in document indexing and retrieval, were incorporated into the ABIR systems to discover a more reliable concept association. However, the level of success in these attempts is questionable; a reason for this lies in the sparsity of the per-image keyword annotation data in comparison to the number of keywords that are usually assigned to documents.

We introduce the Markovian Semantic Indexing (MSI), a new method for automatic annotation and annotation-based image retrieval. The properties of MSI make it particularly suitable for ABIR tasks when the per image annotation data is limited. The characteristics of the method make it also particularly applicable in the context of online image retrieval systems.

The rest of the paper is organized as follows: In the next section, we first present related work and our contribution. The proposed approach (MSI) is presented in Section 3 together with a proximity measure (distance). In Section 4, the geometric interpretation and the optimal properties of the proposed distance are examined. In Section 5, MSI is

- *K.A. Raftopoulos, K.S. Ntalianis, and S.D. Kollias are with the Image, Video, and Multimedia Systems Laboratory (IVML), Computer Science Division, School of Electrical and Computer Engineering, National Technical University of Athens (NTUA), Electrical Engineering Building—1st Floor—Room 1.1.23 Iroon Polytexneiou 9, Zografou 15780, Athens, Greece.*
  E-mail: {raftop, kntal}@image.ntua.gr, stefanos@cs.ntua.gr.
- *D.D Sourlas is with the Elais-Unilever Hellas SA, Unilever-Knorr Factory, Legaki 22, AI Rentis 18233, Athens, Greece.*
  E-mail: dionyssis.sourlas@unilever.com.

Manuscript received 16 Dec. 2009; revised 25 Apr. 2011; accepted 25 Sept. 2011; published online 14 Oct. 2011.
Recommended for acceptance by A. Zhang.
For information on obtaining reprints of this article, please send e-mail to: tkde@computer.org, and reference IEEECS Log Number TKDE-2009-12-0852.
Digital Object Identifier no. 10.1109/TKDE.2011.219.

experimentally compared against two established methods from the literature, namely LSI and pLSI, under two scenarios: 1) of supervised MSI annotation and 2) of external annotations realized with unknown methods. A conceptual comparison between the methods and scalability concerns are also discussed in this section. The Conclusion follows in Section 6. Throughout this paper bold letters are vectors, capital letters are matrices, and standard letters are scalars. A bold letter with subscript is a column vector; its place in a matrix is implied through the subscript. If not explicitly mentioned otherwise, all vectors are considered column vectors except the equilibrium vectors that are traditionally considered as row vectors.

2 Related Work and Our Contribution

Growth in content-based retrieval has been unquestionably rapid. In the recent years, more than 200 content-based retrieval systems have been developed [5], the majority of which are based on low level features. In particular, they can be classified into two main categories: 1) those that perform semantics mining based on the analysis of textual information associated to images, such as annotations, assigned keywords, captions, alternative (alt) text in html pages or surrounding text, and 2) those that are based on the extraction of low-level visual features such as color, texture in order to perform alignment, classification, browsing, searching, summarization, etc. in image collections. Methods of the first category depend on laborious annotation, while the latter methods usually cannot effectively capture semantics. Additionally, some other techniques use both low-level features in the form of visual keywords [6] and text annotation to perform content-based operations but they usually demand the explicit involvement of users for linguistic annotation of pictures [7], [8].

Annotation-Based Image Retrieval systems incorporate more efficient semantic content into both text-based queries and image captions. A direct consequence is that methods initially developed for document retrieval may be suitable for ABIR systems, as well. Latent Semantic Indexing [9] was initially developed for document retrieval. Hofmann, based on the Aspect Model [10] introduced the probabilistic Latent Semantic Indexing (pLSI) [11] as an alternative to projection (LSI) or clustering methods for document retrieval. Latent Dirichlet Allocation (LDA) was proposed by Blei et al. [12] to address the limitations of pLSI regarding generalization and overfitting while Griffiths and Steyvers incorporated a Markov chain Monte Carlo technique to LDA [13]. Steyvers et al. introduced a new probabilistic model representing both authors and topics in document retrieval, and incorporating Gibbs sampling to overcome overfitting problems [14].

Attempts of applying LSI/pLSI-based techniques to discover a more reliable concept association in ABIR systems have been reported in [15], [16], [17], and [18].

In the computer vision literature joint models over image and text have been already explored. Barnard and Forsyth [19] proposed a statistical model with hybrid text/visual characteristics also based on the aspect model. Another model based on Bayesian incremental learning was proposed by Li et al. [20], while Fan et al, propose a scheme for multilevel automatic annotations that is based on local and global visual features [21].

The approach proposed herein, while stochastic in nature, raises the reasoning aspect of probabilities, as it defines explicit relevance relationships between keywords. Using network representations for capturing semantics is common in AI systems. In probability theory they are called Bayesian networks, causal nets, or influence diagrams [24]; in the Dempster-Shafer theory they are referred to as galleries [25], qualitative Markov networks [22], or constraint networks [26]. Many models of human reasoning are often portrayed in terms of associational graphs (e.g., semantic networks [27], constraint networks [26], inference networks [28], conceptual dependencies [29], and conceptual structures [30]) and have been shown to better reflect human plausibility [24].

2.1 Our Contribution

The methodology proposed in this work encompasses a novel (alternative) *probabilistic* approach for Annotation-Based Image Retrieval that, compared to LSI and pLSI, is better suited to sparsely annotated domains, like in image databases where, the per image sparse keyword annotation is also limited. It addresses in a more natural way the *zero frequency problem*, defined as the fact that the probability to find common keywords even in closely related images is typically small because the images are not annotated with *exactly* the same keywords. This problem is addressed here by means of an explicit relevance link between keywords that carries a probabilistic weight. We show that assigning logical connections between keywords by means of a Markovian model, permits better generalization over a sparsely annotated domain hence the proposed approach raises the reasoning aspect next to the numerical aspect of probabilities. The key idea behind the approach is to compensate for the sparse data by incorporating an annotation procedure of probabilistic *qualitative* reasoning that will propagate partial beliefs regarding connections between keywords. A mechanism that gains performance from mining the structure of the existing data rather that incorporating new data, as it happens with traditional models, is hence introduced.

Furthermore, even though automatic annotation and annotation-based image retrieval systems have been presented in the literature the proposed system is novel in the way it *unifies* these two tasks. Indeed both the automatic annotation and retrieval tasks are assumed in the implicit *user interaction* context, for dynamically mining semantics towards *qualitative probabilistic reasoning*. This has implications in the targeting aspect since the training is performed dynamically by the same users that are actually using the system. The unified Markovian setup behind the proposed system allows the retrieval technique to benefit from the underlying structure of the annotation data; at the same time the annotation data acquires concrete stochastic interpretation through the way it is treated by the retrieval process.

A comparison with LSI and pLSI in the application area of ABIR with Precision versus Recall diagrams on ground truth databases reveal that the proposed approach achieves better retrieval scores. A new measure of similarity between Markovian states, the *expected first Cross Passage Time* (CPT) is defined and its properties are examined. MSI is shown to be optimal with respect to CPT. Conceptual comparisons with LSI but also with pLSI reveal advantages for the proposed approach.

3 The Proposed Approach

In [31], the use of covariance matrices to represent image features demanded the representation to be ported on a Riemannian manifold before a distance could be defined. This paradigm elucidates the importance of treating the distance function and the feature representation *in a unified manner*. In our case, the images are represented by probability vectors assigned to them by the annotation procedure, as will be explained in the next paragraphs. A vector space therefore is available on the basis of which we will propose a distance that renders certain Markovian connectivity measures and has an optimal stochastic interpretation.

The proposed approach will be presented in the framework of an online image retrieval system (similar to Google image search) where users search for images by submitting queries that are made of keywords. The queries formed by the users of a search engine are semantically *refined*, the keywords representing concise semantics when compared to text in documents or other vocabulary related presentations. The aim is to improve user satisfaction by returning images that have a higher probability to be accepted (downloaded) by the user. The assumption is that the users search for images by issuing queries, each query being an ordered set of keywords. The system responds with a list of images. The user can download or ignore the returned images and issue a new query instead. During the training phase of the system the images are considered with no annotation. As the users issue queries and pick images the system annotates the images in an automatic manner and at the same time establishes relevance relations between the keywords as will be explained later on in the manuscript. The user never annotates the images explicitly, this happens by the system *transparently* from the user. At the testing phase the system uses the annotations available from the training phase but also the keyword relevance probability weights also evaluated during the training phase to return images that better reflect the users preferences and improve user satisfaction. This interactive procedure has implicit consequences that we exploit one by one in a step by step construction of the proposed system:

Step 1: The user implicitly relates the retrieved (downloaded) images to her/his query. By assuming Markovian chain transitions in the order of the keywords the aim of the proposed approach is to quantify logical connections between keywords. If some user relates image I_i to his query $\boldsymbol{q}_i$, where keyword k_2 follows keyword k_1 and this occurs m times, then the one step transition probability $p_i(k_1, k_2)$ is being updated as follows: if $p_i(k_1, k_2)$ is the current probability (before the update) based on M keywords then the new probability (based on $M + m$ keywords) is calculated by the *recurrent* formula

$$p_i(k_1, k_2) = \frac{M p_i(k_1, k_2) + m}{M + m}. \quad (1)$$

This procedure constructs a Markov chain where each keyword corresponds to a state. Each time a keyword appears in a query, its state counter is advanced; if another keyword follows in the same query, their interstate link counter is also advanced. The occurrences of the keywords but also the sequencing of these occurrences are both measured this way. The queries pertaining to an image are batch processed for this image, the counters are advanced, the probabilities are updated as above. Before the next set of queries is processed, the counters are cleared. The equilibrium state vector of the such constructed Markov Chain, for each image I_i, will be denoted by $\boldsymbol{\pi}_i$ and will represent the image from now on. This modeling approach is justified: 1) by the qualitative character of the conceptual approach that is adopted in this work, and 2) the targeting aspects of the MSI, whereby the objective is to capture user specific aspects (e.g., perception of images). Indeed, the fact that each sequence of keywords (query) comes from a specific user, filtered through her/his perception about the selected image, justifies the assignment of a logical connection of relevance between these two keywords, in addition to the individual connections between each keyword and the selected image. It is this logical connection that is favored by our modeling approach rather than computing the *distribution* of images over keywords in the traditional numerical fashion.

Step 2: Trying to compare directly the probability vectors $\boldsymbol{\pi}_i$ and $\boldsymbol{\pi}_j$ calculated in the previous step for two images, one faces the zero-frequency problem. By itself, the fact that a user puts certain keywords together in a query implicitly renders the keywords relative to each other regardless of the images that may or may not be picked by this user. We propose to use this and address the zero-frequency problem by *clustering* the keyword space into *similar* keywords. For this purpose, the Aggregate Markovian Chain (AMC) of all the queries asked by all users regardless of the selected images, is constructed in this step. The *kernel* of this process denoted by P_G, is calculated in a similar to the previous step manner by the recurrent formula of (1). P_G, even though a Markov kernel it will be used to cluster the keyword space rather than estimating an explicit probability distribution, hence the purpose of the AMC is to model keyword relevance.

Step 3: Optimization step. The AMC will be used to cluster the keyword space and define explicit relevance links between the keywords by means of this clustering. This clustering task is linked to the convergence characteristics of the AMC chain by evaluating the series $F_G(n) = \sum_{k=0}^{n} P_G^k$ where P_G is the AMC kernel. A suitable termination condition stops the series at the *desired* n where the slow convergence has taken over, but not before the rapid convergence has finished. The value of the determinant of $F_G(n)$ is used as a termination condition since the clusters in the rows of $F_G(n)$ will drop its rank and the determinant will become close to zero. $F_G(n)$ is the n-step *expected occupancies matrix* (Appendix). An optimization task is related to this procedure with respect to the total variance of the columns of $F_G(n)$, when projected on the direction of the eigenvectors of P_G. Further insight in this concept will be gained in the next section. From now on, we will use just F_G to note $\frac{1}{n+1} F_G(n)$, the n-step *expected fractional occupancies matrix* (Appendix), calculated at the desired n.

Step 4: The formal definition of the MSI distance can be provided at this step.

Definition 1. *Let $\boldsymbol{x}$ and $\boldsymbol{y}$ two images represented by their respective steady state probability row vectors $\boldsymbol{\pi}_x$ and $\boldsymbol{\pi}_y$. Let also $\Sigma({F_G}^T)$ be the covariance matrix of the zero-mean transpose expected fractional occupancies matrix of the Aggregate Markov Chain (AMC), calculated at the desired n. Then the MSI distance between images $\boldsymbol{x}$ and $\boldsymbol{y}$ is defined as*

$$\begin{aligned} d(\boldsymbol{x},\boldsymbol{y}) &= (\boldsymbol{\pi}_x - \boldsymbol{\pi}_y)\Sigma({F_G}^T)(\boldsymbol{\pi}_x - \boldsymbol{\pi}_y)^T \\ &= \boldsymbol{\delta}_{xy}\Sigma({F_G}^T)\boldsymbol{\delta}_{xy}^T, \end{aligned} \tag{2}$$

where the dimensionality of $\boldsymbol{\pi}_x$ and $\boldsymbol{\pi}_y$ has been extended to that of $\Sigma({F_G}^T)$ by filling in with zeros the respective coordinates.

The proposed distance is well defined since it is a generalized euclidean distance function, using a covariance matrix, which is always positive definite.

4 Geometric Interpretation and Optimality Properties of the MSI Distance

The proposed MSI distance $d(x,y)$ (Definition 1) can be viewed as measuring the total variance of the rows of ${F_G}^T$ when projected on the direction of the difference, $(\boldsymbol{\pi}_x - \boldsymbol{\pi}_y)$, of the two images. The ${F_G}^T$ is calculated once from all the data and thus the *direction* defined by the vector difference of the probability distributions of the two particular images is actually deciding their in between distance. To get further geometric and stochastic interpretations of the MSI distance one has to acquire insight into the mechanics of the Markovian convergence that produces the ${F_G}^T$ and its relation to certain *directions* in the keyword space. In the remaining of this section, we will investigate the geometrical meaning of the convergence process and its relation to the proposed distance with respect to state clusters and state connectivity measures. Standard terminology, concepts, and notation from the stochastic processes literature will be used. The reader is referred to the Appendix for a brief introduction.

4.1 The Convergence Process on the Eigenvectors of the Markov Kernel as a Basis for State Partition

In this section, we realize that the convergence from an initial condition to the respective equilibrium state happens on the direction of the eigenvectors of the Markov kernel and the relative rate of convergence in each direction is dictated by the respective eigenvalue. The reader is referred to the Appendix, for a brief introduction in the notation of $\boldsymbol{\tau}_i(n), \boldsymbol{t}_i(n)$, $\boldsymbol{t}_g(i)$, and their values $\tau_{ij}(n), t_{ij}(n)$, $t_g(ij)$, respectively as well as other relevant concepts used hereinafter.

The l_1 normalized eigenvector that corresponds to an eigenvalue of 1 is the equilibrium state for the respective initial condition since, as will be clear later in this section, the chain converges with rate 1 on this direction. We first present some useful derivations by expressing the transient components of the process in terms of the projections of the initial conditions on the eigenvectors. Let i, j, be nodes of some chain and $\boldsymbol{e}_i = \boldsymbol{\phi}_i(0), \boldsymbol{e}_j = \boldsymbol{\phi}_j(0)$ the respective initial conditions. Let also $\lambda_i, i = 1, 2, \ldots, N$ be the eigenvalues of P, $|\lambda_i| \leq 1, \forall i$ and $\boldsymbol{v}_i$ the respective normalized (with respect to the l_1 norm) *left hand* eigenvectors. Let $|\lambda_i| \geq |\lambda_j|$ when $i < j$, a nonincreasing ordering of the eigenvalues.

Let V be the matrix having rows $\boldsymbol{v}_i$, V^T the transpose of V and Λ the diagonal matrix of the eigenvalues of P that corresponds to V.

We have: $V^T P = \Lambda V^T \Leftrightarrow (V^T)^{-1} V^T P = (V^T)^{-1}\Lambda V^T \Leftrightarrow P = (V^T)^{-1}\Lambda V^T \Leftrightarrow P(V^T)^{-1} = (V^T)^{-1}\Lambda$, thus if V is the matrix with columns the *left-hand* eigenvectors of P, then $(V^T)^{-1}$ is the matrix with columns the *right-hand* eigenvectors of P and we denote it with $V' = \{v'_{ij}\}$.

From the obvious $V'V^T = I$, we conclude that the ith row of V' holds the coordinates of the initial condition $\boldsymbol{e}_i$ when this is projected on the column eigenvectors of V.

Let now i and j be two initial conditions. The respective initial vectors can be expressed as linear combinations of the $l_1 - normalized$ eigenvectors

$$\begin{aligned} \boldsymbol{e}_i &= \boldsymbol{\phi}_i(0) = v'_{1i}\boldsymbol{v}_1 + v'_{2i}\boldsymbol{v}_2 + \cdots + v'_{Ni}\boldsymbol{v}_N \\ \boldsymbol{e}_j &= \boldsymbol{\phi}_j(0) = v'_{1j}\boldsymbol{v}_1 + v'_{2j}\boldsymbol{v}_2 + \cdots + v'_{Nj}\boldsymbol{v}_N. \end{aligned} \tag{3}$$

Employing an ordering for the eigenvectors compatible with the ordering of the eigenvalues ($\boldsymbol{v}_1$ is the eigenvector that corresponds to the eigenvalue λ_1), and assuming that the states i and j belong to the same chain, the system's response n-steps into the process, when starting from i and j, can be expressed as

$$\begin{aligned} \boldsymbol{\phi}_i(n) &= \boldsymbol{v}_1 + \lambda_2^n v'_{2i}\boldsymbol{v}_2 + \cdots + \lambda_N^n v'_{Ni}\boldsymbol{v}_N \\ &= \boldsymbol{v}_1 + \boldsymbol{t}_i(n) \\ \boldsymbol{\phi}_j(n) &= \boldsymbol{v}_1 + \lambda_2^n v'_{2j}\boldsymbol{v}_2 + \cdots + \lambda_N^n v'_{Nj}\boldsymbol{v}_N \\ &= \boldsymbol{v}_1 + \boldsymbol{t}_j(n). \end{aligned} \tag{4}$$

Here, we look into one chain only, thus the indices are modified in a way that $\boldsymbol{v}_1$ is the only equilibrium eigenvector, $\lambda_i < 1, \forall i > 1$ and N is the cardinality of the chain. The significance in this expression is in that it shows the structure of the transient components through their coordinates on the eigenbase. The vector of the expected number of occupancies, n steps into the process that started from node i, is then

$$\begin{aligned} \bar{\boldsymbol{\eta}}_i(n) &= \sum_{k=0}^{n} \boldsymbol{\phi}_i(k) \\ &= (n+1)\boldsymbol{v}_1 + v'_{2i}\boldsymbol{v}_2 \sum_{k=0}^{n} \lambda_2^k + \cdots + v'_{Ni}\boldsymbol{v}_N \sum_{k=0}^{n} \lambda_N^k \\ &= (n+1)\boldsymbol{v}_1 + \sum_{p=2}^{N} \frac{1-\lambda_p^n}{1-\lambda_p} v'_{pi}\boldsymbol{v}_p \\ &= (n+1)\boldsymbol{v}_1 + \sum_{k=0}^{n} \boldsymbol{t}_i(k) = (n+1)\boldsymbol{v}_1 + \boldsymbol{\tau}_i(n), \end{aligned} \tag{5}$$

and the vector of the expected fractional occupancies; if started from i, is

$$\begin{aligned} \boldsymbol{f}_i(n) &= \frac{\bar{\boldsymbol{\eta}}_i(n)}{n+1} \\ &= \boldsymbol{v}_1 + \frac{1}{n+1}\sum_{p=2}^{N} \frac{1-\lambda_p^n}{1-\lambda_p} v'_{pi}\boldsymbol{v}_p \\ &= \boldsymbol{v}_1 + \frac{\boldsymbol{\tau}_i(n)}{n+1}. \end{aligned} \tag{6}$$

This expression shows the structure of the transient sum components through their coordinates on the eigenbase.

The difference of the expected occupancies of each node at step n due to starting the system from i rather than j is then

$$\bar{\boldsymbol{\eta}}_i(n) - \bar{\boldsymbol{\eta}}_j(n) = \sum_{k=0}^{n} \boldsymbol{\phi}_i(k) - \sum_{k=0}^{n} \boldsymbol{\phi}_j(k) = \sum_{k=0}^{n} (\boldsymbol{\phi}_i(k) - \boldsymbol{\phi}_j(k)),$$

which, taking into account (4) and (5) becomes

$$\begin{aligned} \bar{\boldsymbol{\eta}}_i(n) - \bar{\boldsymbol{\eta}}_j(n) &= \sum_{p=2}^{N} \sum_{k=0}^{n} \lambda_p^k (v'_{pi} - v'_{pj}) \boldsymbol{v}_p \\ &= \sum_{p=2}^{N} \frac{1-\lambda_p^n}{1-\lambda_p} (v'_{pi} - v'_{pj}) \boldsymbol{v}_p \\ &= \boldsymbol{\tau}_i(n) - \boldsymbol{\tau}_j(n). \end{aligned} \tag{7}$$

Hence, the difference of the transient sum components expressed on the eigenbase. Notice that since i and j are in the same chain, the term $(n+1)\boldsymbol{v}_1$, in (5), vanishes in the difference of (7) and thus the result converges. This is not the case in general, since, if i and j are in different chains, they have different equilibrium vectors, say, $\boldsymbol{v}_1$ and $\boldsymbol{u}_1$ and therefore the respective term becomes $(n+1)(\boldsymbol{v}_1 - \boldsymbol{u}_1)$ tending to infinity. We can overcome this problem, if we convert the multidesmic processes to monodesmic, by introducing small, close to zero, probabilities connecting the different chains. In the sequel, we will always consider monodesmic processes and thus the term *Markovian chain* will be used hereinafter.

The difference of the expected fractional occupancies always converges and in the case of a monodesmic process it is

$$\begin{aligned} \boldsymbol{f}_i(n) - \boldsymbol{f}_j(n) &= \frac{\bar{\boldsymbol{\eta}}_i(n) - \bar{\boldsymbol{\eta}}_j(n)}{n+1} \\ &= \frac{1}{n+1} \sum_{p=2}^{N} \frac{1-\lambda_p^n}{1-\lambda_p} (v'_{pi} - v'_{pj}) \boldsymbol{v}_p \\ &= \frac{\boldsymbol{\tau}_i(n) - \boldsymbol{\tau}_j(n)}{n+1}. \end{aligned} \tag{8}$$

As n goes to infinity, (7) becomes

$$\bar{\boldsymbol{\eta}}_i(\infty) - \bar{\boldsymbol{\eta}}_j(\infty) = \sum_{p=2}^{N} \frac{v'_{pi} - v'_{pj}}{1-\lambda_p} \boldsymbol{v}_p = \boldsymbol{t}_g(i) - \boldsymbol{t}_g(j), \tag{9}$$

and this is the transient sum difference expressed on the eigenbase. This expression will be useful in proving optimality concepts of the proposed distance in the next sections, since it links the mechanics of the convergence to the actual process statistics.

Let now i, j be two states of the same chain with j recurrent. The mean first passage time $\bar{\theta}_{ij}$ from state i to state j is the expected number of transitions the process needs to reach j for the first time, starting from i

$$\bar{\theta}_{ij} = \frac{t_g(jj) - t_g(ij)}{\pi_j}, \tag{10}$$

where $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_N)$ the equilibrium state of the chain. Equations (9) and (10) show that the mean first passage time from j to i is related to the elements of V', the coefficients of the initial conditions $\boldsymbol{e}_i$ and $\boldsymbol{e}_j$ in their expression as a linear combination of the eigenvectors of P. In particular from (9) and (10) we get

$$\pi_j \bar{\theta}_{ij} = \frac{v'_{2j} - v'_{2i}}{1-\lambda_2} v_{j2} + \cdots + \frac{v'_{Nj} - v'_{Ni}}{1-\lambda_N} v_{jN}, \tag{11}$$

which means that the first passage time from j to i depends on the difference of the coordinates of the respective initial conditions on the eigenvectors of P amplified by the respective eigenvalue. The projections of the initial conditions on the eigenbase are the **columns** of the matrix V', namely the *right hand* eigenvectors of P. Therefore, while the left hand eigenvectors of V are the directions of convergence form the initial conditions to equilibrium, the right-hand eigenvectors of P describe this convergence in terms of direction and speed for each of the initial conditions in different time scales, according to the respective eigenvalues. The first passage time provides a measure of the connectivity nearness between two states and for our purpose, of establishing direct relevance links between keywords, it seems very important. Nevertheless, as is shown in Fig. 1, in the case of the nonrecurrent states 1 and 2, even though these states lead to identical futures, the first passage time between them cannot be defined, since the process never visits them both. To handle this and other similar situations of non recurrent states, we need a connectivity measure between *any* pair of states regardless of the way the process visits these states. We propose to achieve this, by relating the states based on the statistics of the whole process after visiting/starting from these states, and not by directly connecting the states themselves.

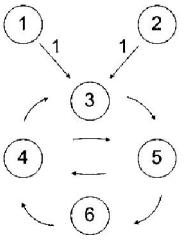


Fig. 1. A Markov chain showing two nonrecurrent states with identical future.

4.2 The Mean First Cross Passage Time

Now that we have arranged certain concepts regarding the directions of convergence and their stochastic meaning we will attempt to connect these concepts to the mechanics of the proposed MSI distance. First, we need a connectivity measure between the Markovian states that will serve as the explicit probabilistic *relevance link* between the keywords.

In Fig. 1, we can see the states 1 and 2 being isolated from each other but identical as far as the process is concerned after starting from these states. We cannot define any metric that involves directly these two states, like the first passage time, since the two states are isolated from each other and a passage between them does not exist. We will have then to compare them through their connectivity to other states and for this purpose, we now introduce a new measure of state similarity that we call *mean first cross passage time* (CPT). The

mean first cross passage time between states i and j will measure the difference of the passages of these two states to all the other states. Due to the memoryless property, once the process reach the same state, lets say k, from i and from j then the different initial condition (i or j) has no effect for the process any more, therefore, the mean first cross passage time between two states will be the expected time it takes for the process to cancel out the fact of one of these two states being the initial condition rather than the other. Intuitively, the CPT will be large between states that there is no state to which *both of them* connect with short passages. The CPT is calculated next.

First, we choose an arbitrary state k. If the first passage crossing from i and from j happened at k then the process from i visited k for the first time after an average of $\bar{\theta}_{ik}$ transitions and the process from j also visited k for the first time after an average of $\bar{\theta}_{jk}$ transitions. The difference of the mean first passage times of i and j to k is the expected time k will have to wait for the second passage after it was visited by the first passage and this is the CPT given that the first crossing of passages happened at state k. To say this formally, if i, j arbitrary and k recurrent and the mean first passage time from i to k and from j to k is $\bar{\theta}_{ik}$ and $\bar{\theta}_{jk}$, respectively, then the *mean first cross passage time* between i and j given that the first crossing of passages will happen at state k is $|\bar{\theta}_{ik} - \bar{\theta}_{jk}|$. To overcome the problem of nonrecurrent states and to also remove the condition that the first crossing of passages will happen at state k, we can assume zero π_k for nonrecurrent states k and add up the between i and j first cross passage times at k, weighted by the probability of the process being at state k. So finally

Definition 2. *For any arbitrary pair of Markovian states i and j, their expected first Cross Passage Time is denoted by $\bar{\psi}_{ij}$ and is given by*

$$\bar{\psi}_{ij} = \sum_{k \in N} |\bar{\theta}_{ik} - \bar{\theta}_{jk}| \pi_k, \tag{12}$$

which from (11) becomes:

$$\begin{aligned} \bar{\psi}_{ij} &= \sum_{k \in N} |\pi_k \bar{\theta}_{ik} - \pi_k \bar{\theta}_{jk}| \\ &= \sum_{k \in N} |\frac{v'_{2j} - v'_{2i}}{1 - \lambda_2} v_{k2} + \cdots + \frac{v'_{Nj} - v'_{Ni}}{1 - \lambda_N} v_{kN}| \end{aligned} \tag{13}$$

The existence of a *Perron root* λ_2 (modulus close to one) means that no matter where the process starts from, lets say from i, at a large time n its state vector will be $\boldsymbol{v}_1 + \lambda_2^n v'_{2i} \boldsymbol{v}_2$ because according to (4) the other terms will vanish, so the chain will move toward equilibrium with the rate that λ_2 shrinks the coefficient v'_{2i} which is slow if λ_2 is close to one. If the process was started from j instead of i then at a large n its state vector would be $\boldsymbol{v}_1 + \lambda_2^n v'_{2j} \boldsymbol{v}_2$ and the difference of the state vectors at this point due to starting the process from j rather than i would be in $|v'_{2j} - v'_{2i}|$. In terms of clustering, this indicates that a 1D clustering of the projections v'_{2i} of the initial conditions $\boldsymbol{e}_i$ on the eigenvector $\boldsymbol{v}_2$ corresponds to a clustering of the states with respect to the final convergence to equilibrium when started from these states and thus with respect to their CPT. The last expression of CPT in (13), gives an interpretation to the significance of the difference $|v'_{2j} - v'_{2i}|$ just described, since in (13) it is shown that it carries most of the weight in the calculation of the CPT. Hence, for a pair of images represented by just one keyword each, their in-between MSI distance is directly related to the CPT between these keywords. Indeed, if $I_1 \equiv \boldsymbol{\pi}_1 = \boldsymbol{e}_i$ and $I_2 \equiv \boldsymbol{\pi}_2 = \boldsymbol{e}_j$ where $\boldsymbol{e}_i, \boldsymbol{e}_j$ *row* vectors with zero entries in all but the ith and jth places, respectively, then with the help of (9) and (10) it follows that

$$\begin{aligned} d(I_1, I_2) &= (\boldsymbol{e}_i - \boldsymbol{e}_j) \Sigma(F_G)^T (\boldsymbol{e}_i - \boldsymbol{e}_j)^T \\ &= \frac{1}{(n+1)^2} \sum_{k \in N} (\bar{\eta}_{ik} - \bar{\eta}_{jk})^2 \\ &= \frac{1}{(n+1)^2} \sum_{k \in N} ((\bar{\eta}_{kk} - \bar{\eta}_{ik}) \quad - (\bar{\eta}_{kk} - \bar{\eta}_{jk}))^2 \\ &= \frac{1}{(n+1)^2} \sum_{k \in N} ((t_g(kk) - t_g(ik)) - (t_g(kk) - t_g(jk)))^2 \\ &= \frac{1}{(n+1)^2} \sum_{k \in N} (\pi_k \bar{\theta}_{ik} - \pi_k \bar{\theta}_{jk})^2. \end{aligned}$$

From (13), we verify that the above MSI distance between the two keywords (and images represented by only these keywords) is directly related to the CPT and we realize that indeed an explicit relevance link has been established between the keywords. It is this link that measures the differences between the keywords in the two images and brings back the reasoning aspect of the proposed approach. Furthermore, the CPT between two states (keywords) depends on the *clustering* of the states that was achieved through the convergence process of the AMC. Therefore, it is the clustering of the state space that is evolved in the definition of the relevance links. In case that more keywords are used to represent the images, by generalizing the above we can infer that their MSI distance is weighting the per keyword-pair relevance link by the difference of their probability to represent the two images. We will now show that this happens in an optimal manner.

4.3 Optimality of the Proposed Distance

The CPT is not only a measure of the relevance between keywords that is modeled in the MSI distance but also gives stochastic interpretation by quantifying the importance of the coordinates of $\boldsymbol{v}'_2$ with respect to connectivity between states. As was shown in (13) and discussed at the end of Section 4.2, the difference between any pair of the coordinates of this vector is the best largest approximation to the CPT between the corresponding states. Consequently, the sum of all the pairwise absolute differences of the coordinates of this vector is the largest best approximation of the total CPT between all pairs of states of the network and thus $\boldsymbol{v}'_2$ is the direction where the CPT projects with maximum variance. We will now show that is also the direction that maximizes the proposed distance.

Proposition 1. *Let $\Sigma(F_G{}^T)$ the covariance matrix of the transpose zero-mean expected fractional occupancies matrix of a Markovian chain with kernel P and d the proposed distance defined in Section 3 as: $d(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{\pi}_x - \boldsymbol{\pi}_y) \Sigma(F_G{}^T) (\boldsymbol{\pi}_x - \boldsymbol{\pi}_y)^T$ then $d(\boldsymbol{x}, \boldsymbol{y})$ is maximized on the direction of maximum CPT.*

Proof. From simple inspection we see that $d(\boldsymbol{x}, \boldsymbol{y})$ is the total variance of the rows of $F_G{}^T$ when projected on the

direction of $(\boldsymbol{\pi}_x - \boldsymbol{\pi}_y) \equiv \boldsymbol{\delta}_{xy}$, therefore when $d(\boldsymbol{x}, \boldsymbol{y})$ achieves global maximum this total variance is maximized. In our case, this direction, on which the total variance is maximized, can be identified, since the rows of ${F_G}^T$ are formed by a Markovian generative process.

From (6), we have that the ith *column* of ${F_G}^T$ is given by: $\boldsymbol{v}_1 + \frac{1}{n+1}\sum_{k=2}^{N} \frac{1-\lambda_k^n}{1-\lambda_k} v'_{ki}\boldsymbol{v}_k$, thus the same column of the zero mean ${F_G}^T$ is given by: $\frac{1}{n+1}\sum_{k=2}^{N} \frac{1-\lambda_k^n}{1-\lambda_k} v'_{ki}\boldsymbol{v}_k$. With simple calculations, we can then verify that the ith *row* of this same matrix is given by $\frac{1}{n+1}\sum_{k=2}^{N} \frac{1-\lambda_k^n}{1-\lambda_k} v_{ki}\boldsymbol{v'}_k$. As n increases the factor

$$\frac{1-\lambda_2^n}{1-\lambda_2}$$

is reduced in a smaller rate than all the other factors $\frac{1-\lambda_k^n}{1-\lambda_k}$. Recall that the desired n_*, in the optimization procedure of the convergence process of F_G, terminates the series when the convergence on the fast shrinking eigenvectors have been completed. This means that at this point the direction of $\boldsymbol{v}'_2$ is the one of total maximum variance since the rows of ${F_G}^T$ are spread by the maximum factor of $\frac{1-\lambda_2^{n_*}}{1-\lambda_2}$ on this direction (the other factors have converged to $\frac{1}{1-\lambda_k}$). But from above, this is the direction of total maximum CPT thus it is proven that the proposed distance is maximized on the direction of maximum expected first cross passage time. □

This last proposition means that the same difference between a pair of coordinates in $\boldsymbol{\pi}_x$ and $\boldsymbol{\pi}_y$, respectively, is penalized more when the corresponding states have higher CPT. The maximum penalty happens when $(\boldsymbol{\pi}_x - \boldsymbol{\pi}_y)$ is on the direction of $\boldsymbol{v}'_2$, where the respective states have maximum CPT.

5 Experimental Results

We compare the proposed method to the LSI and pLSI approaches in two scenarios.

The first experiment is a comparison to LSI, since the limited number of images used in this experiment do not permit reliable comparison to pLSI. The full features of the proposed distance (MSI) are demonstrated in this experiment since the generative process of the aggregate Markov chain during the automatic annotation of images was available to us as is explained later on. Sixty four images that form two intuitive classes were used for this experiment, 32 images related to the term *Greek* and considered to belong to the first class, and 32 images related to the term *Hawaiian* are considered to belong to the second class. First, the distance of the 64 images from the query *Greek Islands* is calculated and ranked for both methods and the results are examined. Then, a complete distance table is built for all the in-between distances of these 64 images using both methods. A precision versus recall diagram is presented for comparison.

In the second experiment, the full features of our method cannot be demonstrated since the scenario involves a publicly available ground-truth database, during the annotation of which we had no control [23]. The Aggregate Markov Chain, necessary for our method cannot be reliably constructed. Nevertheless this experiment serves as a comparison to pLSI in the ability to extract latent features in the case of already annotated databases, the annotation having been performed with unknown methods. For such cases, we propose a modification to the standard MSI approach that involves an explicit step for dimensionality reduction (since the implicit dimensionality reduction through the clustering of the keyword space cannot by applied). This comparison to the pLSI, again with precision versus recall diagrams, is performed using the ground-truth database of [23].

5.1 Comparison to LSI When the Proposed Markovian Annotation is Available

We asked a group of University students to let us record their click-through data (query q, returned ranking r, and clicks c) when interacting with Google Image Search using the English language. The students were asked at different periods to simulate different target groups with specific interests (tourists, sports fans, artists, etc.) since the adaptability of the proposed system to the user's preferences had to be tested. A simple proxy was used for keeping a log file with the query-ID, URL, and click-log. From the recorded sessions over a six months interval, for each selected image, its individual Markov chain (based on the query terms as explained above) was built; only the selected images among the images that were returned by Google Image Search were considered. The equilibrium state vector of the respective Markovian chain was then used to automatically annotate each image. In a similar manner, from the total number of queries from all users, regardless of the chosen images, a Markov chain, representing the Aggregate Markov Chain, was constructed, serving as a dynamically adaptive *taxonomy* for the keywords already seen in the user queries. A subpart of this network, that formed a cluster around the keywords *Greek* and *Hawaian* is shown graphically in Fig. 2. The transition probabilities lower than 0.1 were omitted.

In Table 2, we can see a ranking of 64 images according to their semantic distance from the query *Greek Islands* for the two methods under comparison. The first column holds the initial enumeration for each image, columns 4-24 hold the equilibrium state vector of the Markov chain representing each image, the second and third columns of this table hold the ranking with respect to the distance from the query *Greek Islands* for the MSI and LSI methods, respectively. The query *Greek Islands*, is represented by the vector $[Greek(0.5), Islands(0.5)]$. As can be seen in the ranking of the images returned from the proposed system (MSI) in Table 2, the semantic relevance of the keywords implied by the Markovian network of the aggregate Markov chain, permits intelligent retrieval, involving dependencies beyond simple keyword co-occurrences. While, e.g., images (1), (2), and (6) are the three closest to the query, as is expected, since they share the same keywords, images (3), (4), and (5) come in the 29th, 25th and 17th places, respectively, even if they share the same exactly keywords with images (1), (2), and (6). Before (3), (4) and (5), we meet in places from 4th to 16th a lot of images that include additional keywords. This happens because, as can be seen in the Aggregate Markov Chain of Fig. 2, starting the process from initial conditions (or transient states) equal to the vector representing each of these images (e.g., (15), (14), (7), (8), (11), (9), etc), the network is closer, in the sense of expected number of

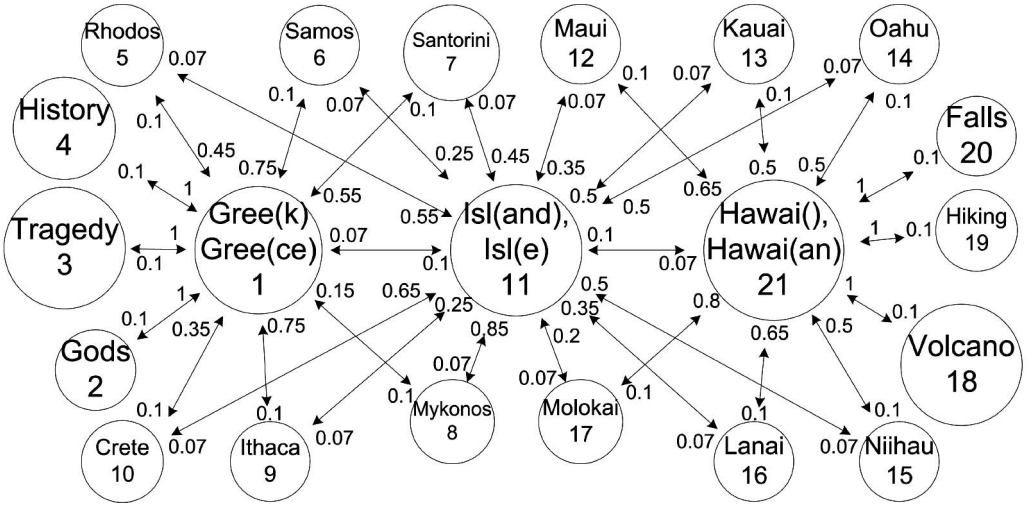


Fig. 2. The Aggregate Markovian Process of the experiment in Section 5.1 is shown in this figure. Only the part of the network forming a cluster around the concepts of *Greek*, *Hawaiian*, and *Islands* is graphically shown. The transition probabilities leading outside the displayed portion of the network have been set to zero and their weight distributed evenly to the respective connections within the displayed portion of the network. Transition probabilities less than 0.05 have been omitted.

transitions, to reach at a state represented by the query, than it is when starting from the initial conditions that correspond to the images (3),(4), or(5).

A similar situation is observed further down in the ranking, where, e.g., image (14) precedes a lot of images that include both query keywords, even if it includes only one of the two query keywords. This is also explained by the aggregate process of Fig. 2, where it has been captured that the majority of users are more likely to ask either of the query keywords *island* or *Greek*, after they ask *Santorini*, than they are after asking any of the other island names.

We also observe in this ranking, that all images involving *Greek* precede all images involving *Hawaii* as it should happen and is clearly depicted in the Markov chain of AMC. This is so, since all Greek islands have a higher probability to lead to both the query terms than Hawaiian islands are. We also notice that the images involving any of the three terms (*God*, *Tragedy*, *History*) linked only to *Greek* but not to *island*, come after the images involving Greek islands, but before the images involving Hawaiian islands, as it should be desirable, since even if they are linked only to one keyword, all the keywords involving Hawaiian islands lead, with a high probability, away from at least one of the query keywords.

Regarding the LSI rankings we observe that while initially it returns more intuitive results, like for example images (1)-(6) that are the closest to the query since they share *exactly* the same keywords, further down in the ranking we see nonintuitive results like, e.g., in the 10th place we meet image #40, a Hawaiian island, being closer to the query *Greek Island* than a lot of images involving Greek islands. The consistency we observed with the MSI, that all Greek islands preceded all the Hawaiian islands, is lost in the LSI ranking.

To further examine the behavior of the two methods we constructed a distance table between all pairs of the 64 images and considering images (1)-(32) forming the class *Greek* and images 33-64 forming the class *Hawaiian*, a Precision versus recall diagram was constructed for the two methods.

For the proposed (MSI) approach the steps in constructing the distance table is as follows:

1. Starting from the Markov kernel P of Table 1 add a small quantity ϵ to all the one-diagonal elements (elements lying on the superdiagonal) and subtract it from any random nonzero element in the same line. This way we convert the process to monodesmic (chain), according to Section 4.1 while keeping the matrix stochastic and without altering the process statistics.
2. Calculate the expected fractional occupancy matrix $F_G = \frac{1}{(n+1)} \sum_{k=1}^{n} P^k$ at the desired n, where n is estimated according to the discussion in Section 3. We present results for $\mathrm{n} = 1, \mathrm{n} = 5$, $\mathrm{n} = 10$, $\mathrm{n} = 12$, $\mathrm{n} = 14$, $\mathrm{n} = 15$. For calculating the powers of P there

TABLE 1
The Markov Kernel that Correpsonds to the Chain of Fig. 2

	K1 GRE	K2 GOD	K3 TRA	K4 HIS	K5 RHO	K6 SAM	K7 SAN	K8 MYK	K9 ITH	K10 CRE	K11 ISL	K12 MAU	K13 KAU	K14 OAH	K15 NIH	K16 LAN	K17 MOL	K18 VOL	K19 HIK	K20 FAL	K21 HAW
GRE	-	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	-	-	-	-	-	-	-	-	-	-
GOD	1.0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
TRA	1.0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
HIS	1.0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
RHO	0.45	-	-	-	-	-	-	-	-	-	0.55	-	-	-	-	-	-	-	-	-	-
SAM	0.75	-	-	-	-	-	-	-	-	-	0.25	-	-	-	-	-	-	-	-	-	-
SAN	0.55	-	-	-	-	-	-	-	-	-	0.45	-	-	-	-	-	-	-	-	-	-
MYK	0.15	-	-	-	-	-	-	-	-	-	0.85	-	-	-	-	-	-	-	-	-	-
ITH	0.75	-	-	-	-	-	-	-	-	-	0.25	-	-	-	-	-	-	-	-	-	-
CRE	0.35	-	-	-	-	-	-	-	-	-	0.65	-	-	-	-	-	-	-	-	-	-
ISL	0.07	-	-	-	0.07	0.07	0.07	0.07	0.07	0.07	-	0.07	0.07	0.07	0.07	0.07	0.07	-	-	-	0.07
MAU	-	-	-	-	-	-	-	-	-	-	0.35	-	-	-	-	-	-	-	-	-	0.65
KAU	-	-	-	-	-	-	-	-	-	-	0.5	-	-	-	-	-	-	-	-	-	0.5
OAH	-	-	-	-	-	-	-	-	-	-	0.5	-	-	-	-	-	-	-	-	-	0.5
NIH	-	-	-	-	-	-	-	-	-	-	0.5	-	-	-	-	-	-	-	-	-	0.5
LAN	-	-	-	-	-	-	-	-	-	-	0.35	-	-	-	-	-	-	-	-	-	0.65
MOL	-	-	-	-	-	-	-	-	-	-	0.2	-	-	-	-	-	-	-	-	-	0.8
VOL	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1.0
HIK	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1.0
FAL	-	-	-	-	-	-	-	-	-	-	0.1	-	0.1	-	0.1	-	-	-	-	-	1.0
HAW	-	-	-	-	-	-	-	-	-	-	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	-

is no need for matrix multiplication since, according to (6), an eigenvalue decomposition of P is enough to calculate F_G at any n, only the powers of the eigenvalues need to be calculated.

3. Calculate the zero mean ${F_G}^T$ by subtracting the mean row from each row of ${F_G}^T$ and then calculate the covariance matrix of ${F_G}^T$ and denote it $\Sigma({F_G}^T)$.
4. For each pair of rows $\boldsymbol{r}_i, \boldsymbol{r}_j$ of Table 2 calculate their distance as $(\boldsymbol{r}_i - \boldsymbol{r}_j)\Sigma({F_G}^T)(\boldsymbol{r}_i - \boldsymbol{r}_j)^T$.

In the case of the query $\boldsymbol{q} = [Greek(0.5),\ Islands(0.5)]$, considered as a row vector, its distance from all the rows $\boldsymbol{r}_i$ of Table 2 was calculated by $(\boldsymbol{q} - \boldsymbol{r}_i)\Sigma({F_G}^T)(\boldsymbol{q} - \boldsymbol{r}_i)^T$, where the dimensionalities are arranged equal by filling in zeros in the appropriate places.

For the LSI approach the steps in constructing the distance table is as follows:

1. Starting from the keyword-image matrix A, which is the transpose of Table 2, we perform a singular value decomposition $A = USV^T$ of this matrix for the desired dimensionality k. Since the dimensions in this experiment are not many we used the full dimensionality $k = 21$ and just a reduction to $k = 10$.
2. Since the columns of SV^T represent the images in their LSI representation, we follow the standard LSI approach and calculate the distance between the columns of SV^T as the *cosine* of their angle.

In the case of the query $\boldsymbol{q} = [Greek(0.5),\ Islands(0.5)]$, considered as a column vector, first it is projected on the image space by $\hat{\boldsymbol{q}} = \boldsymbol{q}^T US^{-1}$ and then its distance to all the images is measured by the cosine of its angle with each column of SV^T.

In Fig. 4, we can see the Precision versus Recall graphs for the MSI for $\mathrm{n} = 1, 5, 10, 12, 14, 15$ and the LSI for $\mathrm{k} = 21$ (no dimensionality reduction) and $\mathrm{k} = 10$. Its is evident that LSI cannot capture the latent features that are captured by means of the AMC used in MSI, since the best LSI result, achieved as was expected when no dimensionally reduction is applied, is very poor compared to the MSI results. For MSI, we see how the choice of n is critical for the accuracy of the method since as n increases the accuracy increases rapidly until a maximum is achieved at $n = 10$ where the best MSI result ranks almost all the images in the correct class. For $n > 10$ the accuracy drops slowly and is comparable to the best LSI result only at $n > 15$.

In Fig. 3, we can see the distance table between the first 15 Greek (enumerated as G1-G15) and the first 15 Hawaiian (enumerated with the corresponding to Table 2 indices as H33-H47) images of Table 2, where we observe the perfect score of the MSI, for $n = 10$, where all the first 9 rankings for all the 30 images are in the correct class as opposed to the LSI rankings where many images are matched in the wrong of the two classes.

These results justify our attempt of using a probabilistic approach for semantic inference in sparse domains. The Precision versus Recall measure achieves almost perfect scores for certain *mixing* values of the proposed model but the annotation has been performed implicitly by the click through data, therefore we see the proposed system to *dynamically* adapt to the specific group of users that are actually using the system. The need of reverting to external taxonomy systems for assigning relevance metrics between keywords, and, therefore, face the problem of evaluating the compatibility between those systems and the semantics behind the *actual* users that are using the system is alleviated in MSI.

The mechanism behind LSI correlates keywords that appear at the same image and images when they are annotated with the same keyword, starting therefore from a keyword/image frequency matrix a singular value decomposition provides a more compact representation of keywords and images in a space of fewer dimensions.

In our approach, we need a keyword/keyword 1 step transition probability square matrix (AMC) to infer relationships between keywords only and then we just use the

TABLE 2
64 Images Arranged with Respect to Their Distance from the Query *"Greek Islands"*

IMG #	RANK MSI	RANK LSI	KEYWORDS			
1	1	1	**GRE**(0.5)	-	**ISL**(0.5)	-
2	2	2	**GRE**(0.4)	-	**ISL**(0.6)	-
3	29	6	**GRE**(0.3)	-	**ISL**(0.7)	-
4	25	3	**GRE**(0.8)	-	**ISL**(0.2)	-
5	17	4	**GRE**(0.2)	-	**ISL**(0.8)	-
6	3	5	**GRE**(0.7)	-	**ISL**(0.3)	-
7	6	29	**GRE**(0.2)	-	**ISL**(0.3)	**CRE**(0.5)
8	7	25	**GRE**(0.1)	-	**ISL**(0.3)	**CRE**(0.6)
9	9	24	**GRE**(0.3)	-	**ISL**(0.2)	**CRE**(0.5)
10	15	11	**GRE**(0.2)	-	**ISL**(0.1)	**CRE**(0.7)
11	8	31	**GRE**(0.4)	-	**ISL**(0.2)	**ITH**(0.4)
12	20	58	**GRE**(0.2)	-	**ISL**(0.3)	**ITH**(0.5)
13	10	62	**GRE**(0.1)	-	**ISL**(0.1)	**MYK**(0.8)
14	5	63	-	-	**ISL**(0.1)	**MYK**(0.9)
15	4	38	**GRE**(0.3)	-	**ISL**(0.1)	**MYK**(0.6)
16	26	61	**GRE**(0.1)	-	**ISL**(0.2)	**MYK**(0.7)
17	16	12	**GRE**(0.4)	-	**ISL**(0.1)	**MYK**(0.5)
18	12	9	**GRE**(0.4)	-	**ISL**(0.1)	**SAN**(0.5)
19	13	7	**GRE**(0.5)	-	-	**SAN**(0.5)
20	18	18	-	-	**ISL**(0.4)	**SAN**(0.6)
21	11	17	**GRE**(0.1)	-	**ISL**(0.1)	**SAN**(0.8)
22	24	36	**GRE**(0.3)	-	-	**SAN**(0.7)
23	27	32	**GRE**(0.3)	-	**ISL**(0.1)	**SAM**(0.6)
24	28	30	**GRE**(0.4)	-	**ISL**(0.2)	**SAM**(0.4)
25	14	19	**GRE**(0.3)	-	**ISL**(0.4)	**SAM**(0.3)
26	19	40	**GRE**(0.2)	-	**ISL**(0.1)	**RHO**(0.7)
27	22	59	-	-	**ISL**(0.1)	**RHO**(0.9)
28	21	15	**GRE**(0.2)	-	-	**RHO**(0.8)
29	23	23	**GRE**(0.4)	-	**ISL**(0.3)	**RHO**(0.3)
30	31	8	**GRE**(0.5)	-	-	**HIS**(0.5)
31	30	20	**GRE**(0.7)	-	-	**TRA**(0.3)
32	32	55	**GRE**(0.5)	-	-	**GOD**(0.5)
33	58	64	-	**HAW**(0.4)	-	**FAL**(0.6)
34	62	41	-	**HAW**(0.3)	-	**HIK**(0.7)
35	38	50	-	**HAW**(0.5)	-	**VOL**(0.5)
36	61	49	-	**HAW**(0.1)	**ISL**(0.5)	**MOL**(0.4)
37	49	37	-	**HAW**(0.2)	**ISL**(0.3)	**MOL**(0.5)
38	46	16	-	**HAW**(0.1)	**ISL**(0.6)	**LAN**(0.3)
39	63	26	-	**HAW**(0.5)	-	**LAN**(0.5)
40	40	10	-	-	**ISL**(0.5)	**LAN**(0.5)
41	59	56	-	**HAW**(0.3)	**ISL**(0.3)	**LAN**(0.4)
42	47	22	-	**HAW**(0.4)	**ISL**(0.2)	**NIH**(0.4)
43	36	60	-	**HAW**(0.4)	**ISL**(0.1)	**NIH**(0.5)
44	55	42	-	**HAW**(0.3)	-	**NIH**(0.7)
45	64	47	-	**HAW**(0.6)	**ISL**(0.2)	**OAH**(0.2)
46	48	54	-	-	**ISL**(0.1)	**OAH**(0.9)
47	50	45	-	**HAW**(0.3)	**ISL**(0.2)	**OAH**(0.5)
48	41	13	-	**HAW**(0.3)	-	**OAH**(0.7)
49	56	28	-	**HAW**(0.2)	**ISL**(0.3)	**OAH**(0.5)
50	45	43	-	**HAW**(0.4)	**ISL**(0.3)	**KAU**(0.3)
51	42	52	-	**HAW**(0.7)	**ISL**(0.1)	**KAU**(0.2)
52	53	57	-	**HAW**(0.2)	**ISL**(0.1)	**KAU**(0.7)
53	60	51	-	-	**ISL**(0.1)	**KAU**(0.9)
54	52	21	-	**HAW**(0.3)	**ISL**(0.2)	**MAU**(0.5)
55	43	46	-	-	**ISL**(0.4)	**MAU**(0.6)
56	54	14	-	-	**ISL**(0.3)	**MAU**(0.7)
57	37	53	-	**HAW**(0.7)	**ISL**(0.1)	**MAU**(0.2)
58	44	27	-	-	**ISL**(0.7)	**MAU**(0.3)
59	51	48	-	**HAW**(0.5)	**ISL**(0.5)	-
60	57	34	-	**HAW**(0.7)	**ISL**(0.3)	-
61	39	44	-	**HAW**(0.4)	**ISL**(0.6)	-
62	35	33	-	**HAW**(0.2)	**ISL**(0.8)	-
63	33	35	-	**HAW**(0.2)	**ISL**(0.8)	-
64	34	39	-	**HAW**(0.6)	**ISL**(0.4)	-

Each row represents an image. The first column holds the image tag number, the second and third columns hold the ranking number with respect to the MSI and LSI distances from the query "Greek Islands," respectively for this image. The rest of the entries correspond to the steady state probabilities of the respective Markov chain representing each image. Only the zero entries that help for better illustration appear as dashes, the other zero entries heve been ommited. The ranking for the MSI distance is implied by the AMC of (Fig. 2). Thirty of these images are shown in the distance table of (Fig. 3) for rows 1-15 and 33-47 of this table.

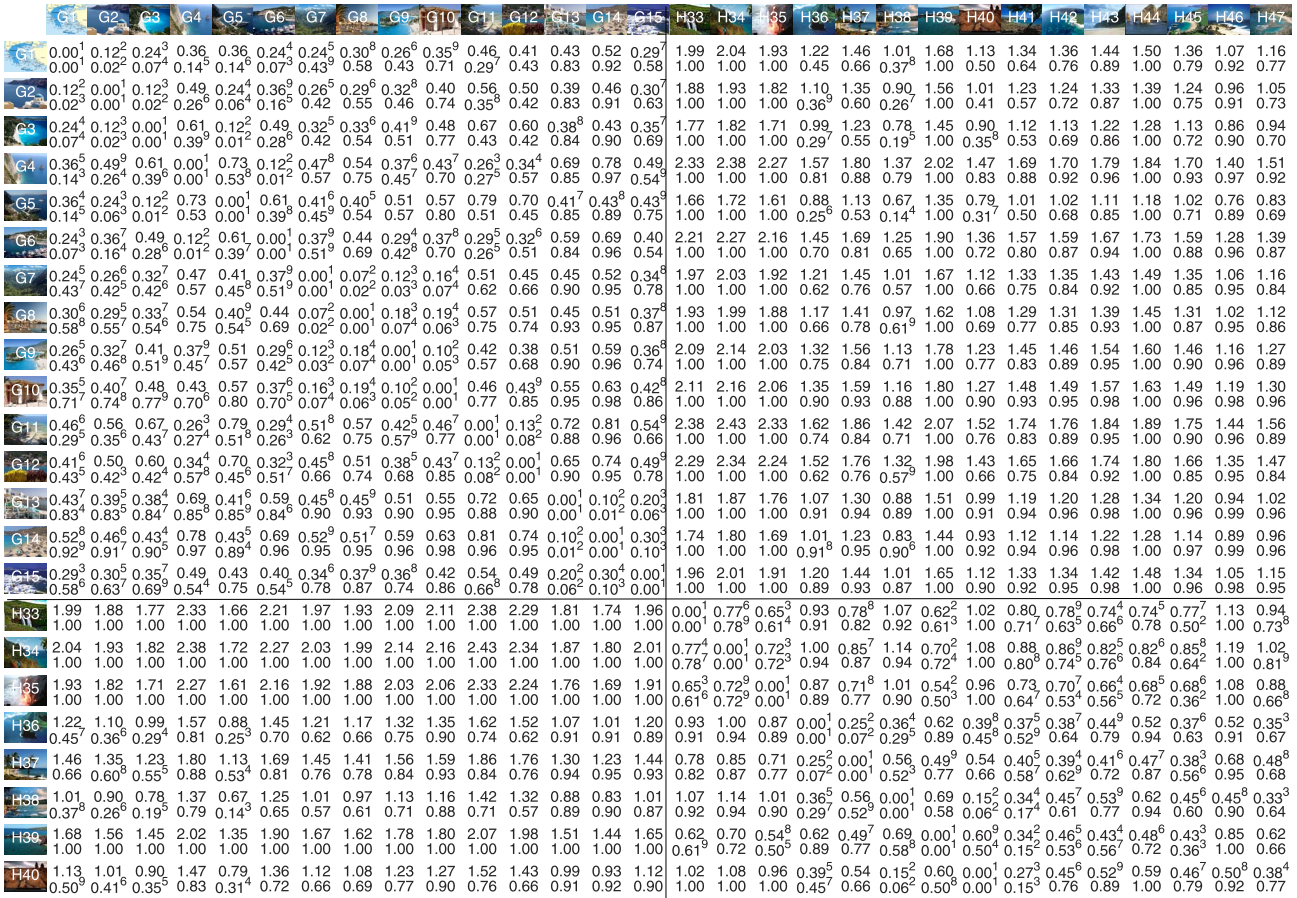


Fig. 3. The distance table between 30 images of two classes. Images $G1$-$G15$ are the first 15 images of Table 2 and belong to the class *Greek* whereas images $H33$-$H47$ are represented by the corresponding 15 rows of the same table and belong to the class *Hawai*. The proposed distance (MSI) (upper number) and the latent semantic indexing (lower number) are compared. The first nine best matches are marked with the respective superscripts. MSI achieves better results since all the first nine matches are correct (the images match to images in the same class), where many of the first nine matches are not correct for the LSI. The precision versus recall diagram for this experiment for all the 64 images of Table 2 is shown in Fig. 4.

keyword/image frequency matrix to represent the image vectors, on the difference of which the covariance matrix will be projected.

The mechanism behind MSI is revealed in the behavior of the system with respect to the power n. We see that as n increases the results are improved until the best Precision Recall result is achieved at $\mathrm{n} = 10$. This behavior shows that we managed to improve the retrieval without adding observations in the system but by propagating relevance connections defined on the existing observations. At the time where LSI or pLSI approaches would need more observations to improve results, the proposed system uses the convergence of the Markov chain (by raising the AMC to the appropriate power) to gain on the existing observations by mining deeper qualitative inferences between keywords.

5.2 Comparison to pLSI Using External Annotation

When we have no control over the annotation method used for the images neither we have access to the user queries that were used to annotate the images (if such a method has even been used), many of the advantages of the proposed method (MSI) vanish since the AMC cannot be constructed. Nevertheless, we still want to incorporate such cases in the proposed approach and compare to pLSI by modifying our method to permit AMC construction with dimensionality reduction whenever only a keyword-image matrix is available.

Our approach assumes that the construction of the AMC has happened progressively from the user queries and in such a case a reduction of dimensionality can be realized by aggregation/disaggregation techniques of the Markovian states, when clusters of states appear in the form of a block

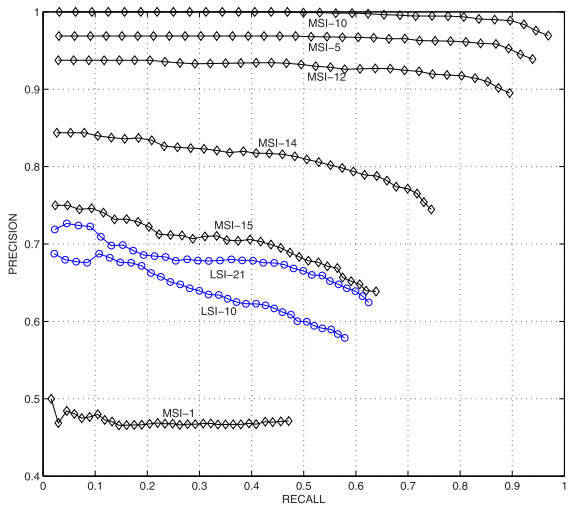


Fig. 4. Precision versus Recall comparisons between Latent Semantic Indexing and the proposed system for various parameters of the two algorithms. The experiment is performed on the 64 images of Table 2. The first 32 images belong to the class *Greek* and the next 32 to the class *Hawaiian*. The diamond graphs represent the results of the proposed distance (implied by the general process of Table 1) for the cases of $n = 1, n = 5, n = 10, n = 12, n = 14, n = 15$. The circle graphs represent the results when the LSI distance is used with the figure/keyword frequency matrix of Table 2 for the parameters of $k = 10$ and $k = 21$. See text for more.

diagonal AMC. Details on fast such methods can be found in [34].

When we have no access to the query logs, the AMC will have to be constructed from the image annotations themselves, treating each images' set of keywords as a query related to this image. By doing this, an AMC will be constructed with dimensions equal to the total number of keywords seen by the system. For a fair comparison to LSI, we will have to apply a reduction of dimensionality to this AMC and then project the images on this reduced space before we measure their distance. This can be achieved by choosing the k *principal components*, where k the desired dimensions after the reduction. Because AMC is a square matrix, we can proceed with the eigen decomposition and choose the k principal components that best approximate the AMC in these dimensions, ending up with an AMC_k of dimension k. For the modified MSI approach therefore the steps in constructing the distance table is as follows:

1. Parse the text annotation of the images included in the ground-truth of [23], assign an index for each unique keyword and build the one-step transition between keywords probability matrix P_G, which is the AMC, considering the annotation of each image as a query related to this image. Similarly to step 1 of Section 6.1, in order to convert the process to monodesmic, add a small quantity ϵ to all the one-diagonal elements (elements lying on the super-diagonal) and subtract it from any random nonzero element in the same line.
2. Perform the eigen decomposition $P_G = VDV^{-1}$ and calculate F_G at the desired n from (6) and the criteria of Section 3 .
3. Calculate the zero-mean ${F_G}^T$ by subtracting the mean and perform the eigen decomposition of the covariance matrix of ${F_G}^T$ as $\Sigma({F_G}^T) = V_1 D_1 {V_1}^{-1}$.
4. Calculate $B = D_{1k} {V_1}_k^{-1}$, where D_{1k} is a square $k \times k$ submatrix of D_1, holding the k largest eigenvalues of $\Sigma({F_G}^T)$ and ${V_1}_k^{-1}$ the submatrix of the k rows of ${V_1}^{-1}$ that correspond to these eigenvalues.
5. Calculate the reduced k-dimensional $\Sigma_k({F_G}^T)$ from $\Sigma_k({F_G}^T) = cov(B^T)$, cov meaning the covariance matrix. At this point we have managed to reduce the dimensionality of $\Sigma({F_G}^T)$ by projecting on the k principal components. Now we need to project the image vectors in the same space.
6. if A is the matrix having rows the image vectors project the image vectors in the k-dimensional space by $A_k = A V_{1k} {D_{1k}}^{-1}$, where V_{1k} the submatrix of the k columns of V_1 that correspond to the k largest eigenvalues of $\Sigma({F_G}^T)$.
7. For every pair or rows $\boldsymbol{r}_i, \boldsymbol{r}_j$ of A_k calculate their distance by $(\boldsymbol{r}_i - \boldsymbol{r}_j)\Sigma_k({F_G}^T)(\boldsymbol{r}_i - \boldsymbol{r}_j)$.

If we do not need to raise to a power, step 2 above can be omitted. This is the case where the annotation is of unknown origin and there is no Markovian connection to the keyword data, in which case MSI performs similar to LSI. The main difference between the modified proposed approach and LSI, when no power is applied, is in the AMC matrix. Indeed the MSI method uses a square keyword/keyword matrix instead of the nonsquare keyword/image matrix of LSI to reduce the dimensionality of the keyword space. The proposed method (MSI), with respect to LSI, has the advantage of the keyword/keyword matrix being a square matrix thus applying eigen decomposition (as opposed to singular value decomposition that LSI invokes on nonsquare matrices), algorithms that are better optimized can be utilized.

To evaluate the modified proposed distance (MSI) in the case of external annotations of unknown origin we compare to the probabilistic Latent Semantic Indexing method, a widely used and generally accepted method for indexing based on latent variables. We use the ground-truth database of [23], with 1,109 images, in 20 classes of about 50 images in each class, annotated with a total of 437 keywords, each annotation being a text string of up to 25 keywords (the classes *barcelona2* and *cambridge* were not used since there are only images but without annotations). We use this data set to do the comparison by means of Precision versus Recall diagrams. Our implementation of pLSI was according to Peter Gehler, available on line at [35].

In Fig. 5, the best matching results for each of the classes are shown. The best matching curve for both methods was achieved at $k = 200$ where we observe that MSI (solid line) performs better for most of the classes. In Fig. 6, we can see the overall Precision versus Recall for all the classes combined where we can verify that both methods achieve the best result for $k = 200$ (k is the number of dimensions after the reduction) but MSI gives better results.

For $k > 200$ both methods produce curves in the space below $k = 300$ and for $k < 200$ the results for both methods are degrading at the same rate.

Though in this experiment there is no control over the annotation procedure and therefore the complete benefits of the proposed system cannot be revealed, we still see that at 200 dimensions, MSI performs better than pLSI does at any dimensionality.

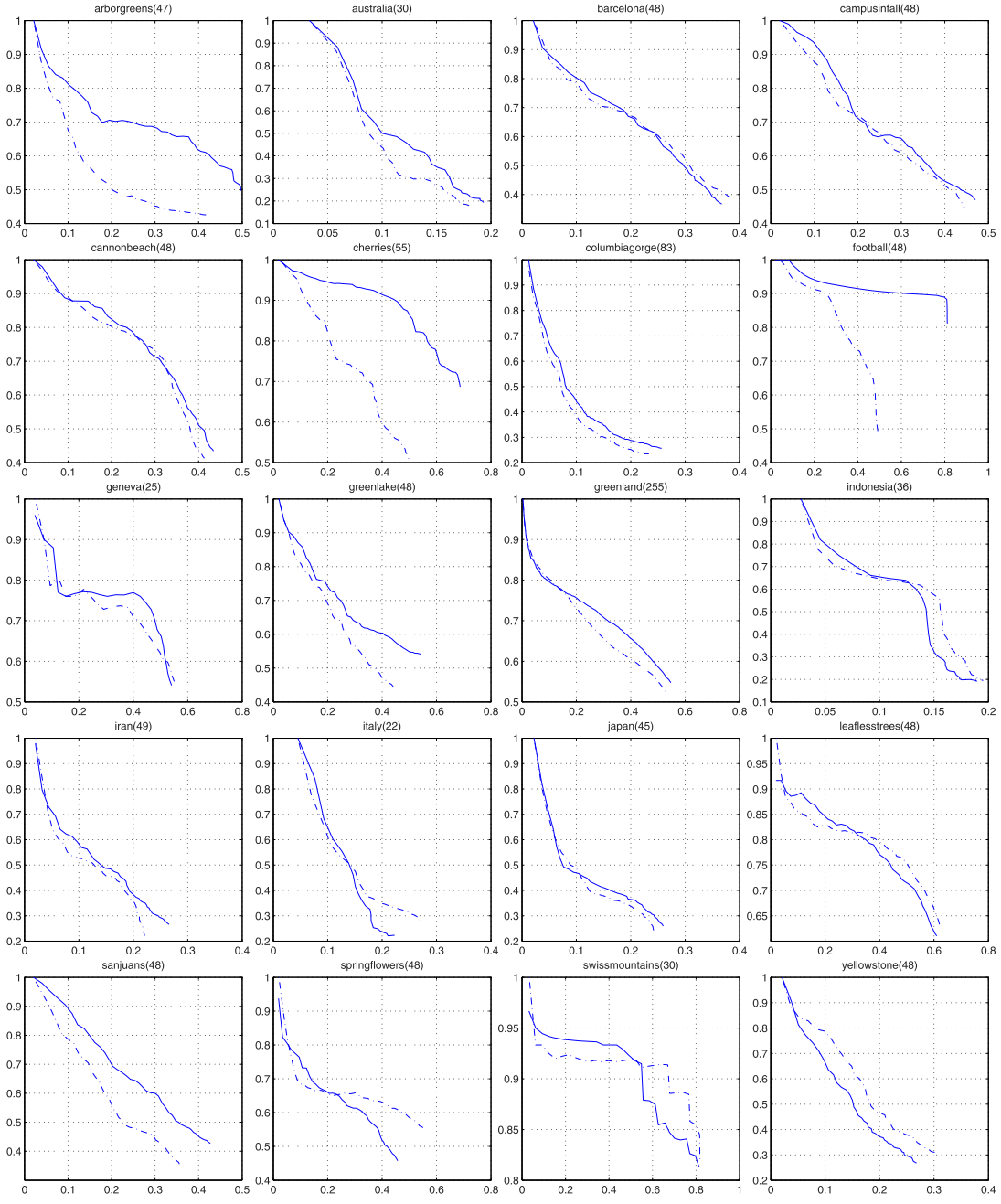


Fig. 5. Precision versus Recall comparisons between MSI (solid lines) and Probabilistic Latent Semantic Indexing (dashed lines) for each of the classes of the ground-truth database of [23]. The name of each class appears as the title for each plot, the number in the parenthesis is the cardinality, the images contained in the class. Only the best result, achieved for both methods at $k = 200$, is shown in the graphs. The proposed method (MSI) performs better in most of the classes even though there was no control over the annotation method. In Fig. 6, the overall Precision versus Recall diagram combined for all the classes is shown for various parameters of the two methods.

5.3 Conceptual Comparisons and Scalability Issues

LSI achieves a reduction of dimensionality based on a L2-optimal approximation of matrices by means of Singular Value Decomposition (SVD). While SVD is well understood, its use in this context is somewhat ad hoc since there is not direct statistical interpretation in the use of the *Frobenious* norm for optimization. pLSI on the other hand which is based on the *aspect model*, defines a proper statistical model but the optimization is achieved through the Expectation Maximization algorithm which has its own problems of overfitting, sensitivity to initial conditions, and generalization on new unseen data [12]. On the other hand, the proposed MSI/AMC approach incorporates an optimization step that relies on the convergence characteristics of the Markovian chain captured in the AMC kernel. These can be identified through an eigen decomposition of the AMC, thus the optimization step in the proposed model not only is optimal but is also stochastic with respect to the connectivity of the Markovian states and thus directly interpretable in the keyword-relevance links.

Both LSI and pLSI treat the *automatic indexing* and the *query-based retrieval* tasks by means of a standard cosine matching. While the *cosine distance* is widely used and generally accepted it has no direct interpretation in relation

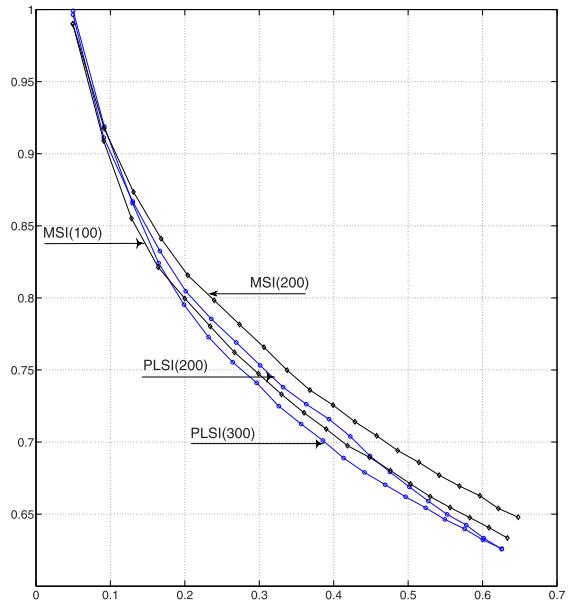


Fig. 6. Precision versus Recall comparisons between Probabilistic Latent Semantic Indexing and the proposed system (MSI) for various parameters of the two algorithms. The experiment is performed on the 1,109 images of the ground-truth database [23]. The diamond graphs represent the results of the proposed distance for the cases of $k = 100$ and $k = 200$. The circle graphs represent the results when the pLSI distance is used with $k = 200$ and $k = 300$. In Fig. 5, the Precision versus Recall is shown for each individual class separately.

to the underlying models in both of these methods. Furthermore, numerical problems have to be addressed when the norm of both vectors is close to zero. The method proposed in this work (MSI/AMC) on the other hand, incorporates automatic indexing and query matching tasks within the whole framework, the distance being defined in an optimal manner, directly interpretable in relation to the clustering of the Markovian states.

Dimensionality reduction in LSI, is achieved by thresholding all but the k largest singular values. In pLSI, k is represented by the *bottleneck* random variable in the aspect model formulation but the value of k is chosen a-priori, usually different values of k are examined and combined. In the proposed MSI model, a reduction of dimensionality can also be considered in the clusters of the keywords, achieved by the convergence procedure. In that case, the analogous to k is the number of clusters and thus it is calculated automatically from the convergence threshold.

A comparison in terms of computational complexity also suggests advantages for MSI. The powers of the AMC kernel in the first implementation as well as the PCA treatment in the second implementation that are involved in the respective MSI optimization steps are both computed accurately and efficiently, since, according to (6), only the eigenvalues of the AMC kernel need to be raised to the respective power. The EM algorithm on the other hand, is an iterative method which is only guaranteed to find a local maximum of the likelihood function.

The inference engine of the MSI approach, lies in the clustering of the state space, since this clustering arranges the states into groups of *relevance*. In terms of scalability, one could examine the degree of this clustering with respect to the size of the system. The degree of state clustering in Markov Chains has been thoroughly studied, usually referred to as the *coupling degree* in the context of Near Completely Decomposable Markov Chains and their faster convergence to equilibrium [32], [34]. The coupling degree of the chain, quantifies the degree of state clustering one can assume in terms of state connectivity. While in these cases a higher coupling degree leads to slower convergence, which is not desired if one is interested for the equilibrium state of the chain, in terms of the MSI distance a higher coupling degree means better differentiation between states in terms of connectivity and therefore better performance of the proposed system. In most real-life applications, the Markov chains we come across are sparse and they more or less possess some structure, that is, the ratio of the number of nonzero elements to the total number of elements in the underlying chain is small; moreover, the magnitude and location of these nonzero elements is not random, hence higher coupling degrees can be expected as the size of the system increases.

6 Conclusion

We proposed the Markovian Semantic Indexing, a new method for mining user queries by defining keyword relevance as a connectivity measure between Markovian states modeled after the user queries. The proposed system is dynamically trained by the queries of the same users that will be served by the system. Consequently, the targeting is more accurate, compared to other systems that use external means of nondynamic or nonadaptive nature to define keyword relevance. A stochastic distance, in the form of a generalized euclidean distance, was constructed by means of an *Aggregate Markovian Chain* and proved to be optimal with respect to certain Markovian connectivity measures that were defined for this purpose. A comparison to Latent Semantic Indexing and probabilistic Latent semantic Indexing revealed certain theoretical advantages of the proposed method (MSI). Experiments have shown that MSI achieves better retrieval results in sparsely annotated image data sets. A comparison to LSI on 64 images gathered from the Google Image Search and annotated in a transparent way by the proposed system, revealed certain advantages for the MSI method, mainly in retrieving images with deeper dependencies than simple keyword cooccurrence. Another comparison to pLSI was performed using the ground-truth annotated database of [23] of 1,109 images after modifying the proposed method to incorporate AMC construction and dimensionality reduction in external annotations. The results of Precision versus Recall for this experiment revealed that MSI at 200 dimensions achieve better scores than pLSI does at any dimensionality.

Appendix

The Markovian Process and Its Transient Behavior

Let $P = \{p_{ij}\}$ the one step transition probability matrix of a Markovian process of N states. A *chain* is a kind of collective trapping state, a set of states the process never leaves once enters that set. A transient state cannot be a member of a chain but can be associated with one if it is possible for the process to enter that chain from that transient state. All

states in a chain are therefor recurrent. The n-step transition probability matrix we denote $\Phi(n) = \{\phi_{ij}(n)\}$ and it is

$$\Phi(n) = P^n = \underbrace{P * P * \cdots * P}_{\text{n times}}.$$

The ith *row* of $\Phi(n)$ corresponds to the probability *vector* of the system being in each of the states after n steps (transient response n steps into the process), if started from state i, or equivalently from the initial condition $\boldsymbol{e}_i$, where $\boldsymbol{e}_i$ the $i\text{th}, N$-dimensional orthocanonical vector which has 1 at the ith place and zeros elsewhere (an example for three dimensions is: $\boldsymbol{e}_1 = (1,0,0),\ \boldsymbol{e}_2 = (0,1,0),\ \boldsymbol{e}_3 = (0,0,1)$). As n goes to infinity the matrix $\Phi(n)$ approaches the limit matrix $P^\infty = \Phi$. All states belonging to the same chain will have identical rows in the matrix Φ, each chain therefor can itself be thought of as a monodesmic Markov process[1] having its own equilibrium probability vector, the identical rows of its states. The transient states associated with more than one chain will have a row in the matrix Φ that is the sum of the rows corresponding to each chain with which they are associated, weighted by the probabilities of entering each of these chains. In the general case conclusively, a Markovian process can be thought of a collection of its chains together with any possible transient states associated with them. A Markovian process of more than one chain will not converge to a single equilibrium probability vector. Depending on the initial condition the equilibrium vector will be the one of the corresponding chain the initial condition is associated with. In this case the nodes associated with any particular chain could be identified by the corresponding group of identical rows of the limiting matrix Φ. The convergence process in a monodesmic process has a single equilibrium point. This is the point where all transient state series from all initial conditions meet eventually. The rate of this convergence is proportional to the product of the eigenvalues of the transition probability matrix, since the rate of convergence in each dimension equals the corresponding eigenvalue. An eigenvalue of 1 corresponds to no convergence along the corresponding direction, the number of eigenvalues that are equal to 1 equals the number of chains in the process. The matrix $T(n) = \Phi(n) - \Phi$, the transient matrix, contains the transient components of the Markov process. The sum of these matrices up to time n is the transient n-sum matrix $T_\Sigma(n) = \sum_{k=0}^{n} T(k)$. The z-transform of $T(n)$ evaluated at 1 is the transient sum matrix $T_g(1)$ that contains the infinite sum of the transient components of the Markovian process and can be calculated from

$$T_g(1) = \sum_{n=0}^{\infty}(P^n - \Phi) = [I - P + \Phi]^{-1} - \Phi. \tag{14}$$

It is clear that $T_g(1)$ is the limit of $T_\Sigma(n)$. The rows of $T_\Sigma(n), T(n)$ and $T_g(1)$ will be denoted $\boldsymbol{\tau}_i(n), \boldsymbol{t}_i(n)$, and $\boldsymbol{t}_g(i)$ and their values with $\tau_{ij}(n), t_{ij}(n)$ and $t_g(ij)$, respectively. The expected number of occupancies of the nodes at time n, if starting from i, is

$$\bar{\boldsymbol{\eta}}_i(n) = \sum_{k=0}^{n} \boldsymbol{\phi}_i(k). \tag{15}$$

The expected fractional occupancies from i is

$$\boldsymbol{f}_i(n) = \frac{\bar{\boldsymbol{\eta}}_i(n)}{n+1} = x_{i1}\boldsymbol{v}_1 + \frac{\boldsymbol{\tau}_i(n)}{n+1}. \tag{16}$$

The vector of fractional occupancies $\boldsymbol{f}_i(n)$ converges to the equilibrium vector $x_{i1}\boldsymbol{v}_1$.

1. A Markovian process of a single chain, having a limiting matrix with identical rows, duodesmic if the process consists of two chains, tridesmic of three chains, etc.

ACKNOWLEDGMENTS

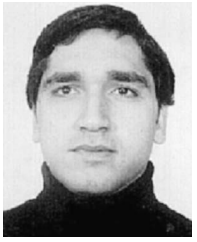
The authors would like to thank the reviewers for their constructive comments that helped significantly in improving this paper.

REFERENCES

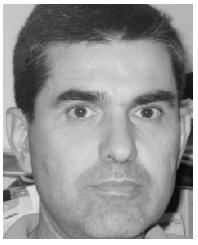
[1] S. Santini and R. Jain, "Similarity Measures," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, no. 9, pp. 871-883, Sept. 1999.
[2] K. Stevenson and C. Leung, "Comparative Evaluation of Web Image Search Engines for Multimedia Applications," *Proc. IEEE Int'l Conf. Multimedia and Expo*, July 2005.
[3] comScore's Report Article, "Comscore's Qsearch 2.0 Service," *comScore's Report Article*, www.comscore.com, 2007.
[4] B.J. Jansen, A. Spink, and T. Saracevic, "Real Life, Real Users, and Real Needs: A Study and Analysis of User Queries on the Web," *Information Processing and Management*, vol. 36, no. 2, pp. 207-227, 2000.
[5] R. Datta, D. Joshi, J. Li, and J.Z. Wang, "Image Retrieval: Ideas, Influences, and Trends of the New Age," *ACM Computing Surveys*, vol. 40, no. 2, pp. 1-60, 2008.
[6] A. Bhattacharya, V. Ljosa, J.-Y. Pan, M.R. Verardo, H. Yang, C. Faloutsos, and A.K. Singh, "Vivo: Visual Vocabulary Construction for Mining Biomedical Images," *Proc. IEEE Fifth Int'l Conf. Data Mining*, Nov. 2005.
[7] J. Li and J. Wang, "Real-Time Computerized Annotation of Pictures," *Proc. ACM 14th Ann. Int'l Conf. Multimedia*, 2006.
[8] D. Joshi, J.Z. Wang, and J. Li, "The Story Picturing Engine - A System for Automatic Text Illustration," *ACM Trans. Multimedia Computing, Comm. and Applications*, vol. 2, no. 1, pp. 68-89, 2006.
[9] M.W. Berry, S.T. Dumais, and G.W. O'Brien, "Using Linear Algebra for Intelligent Information Retrieval," *SIAM Rev.*, vol. 37, no. 4, pp. 573-595, 1995.
[10] T. Hofmann, "Probabilistic Latent Semantic Indexing," *Proc. 22nd Int'l Conf. Research and Development in Information Retrieval (SIGIR '99)*, 1999.
[11] T. Hofmann, "Unsupervised Learning by Probabilistic Latent Semantic Analysis," *Machine Learning*, vol. 42, no. 1/2, pp. 177-196, 2001.
[12] D.M. Blei and A.Y. Ng, and M.I. Jordan, "Latent Dirichlet Allocation," *J. Machine Learning Research*, vol. 3, pp. 993-1022, 2003.
[13] T.L. Griffiths and M. Steyvers, "Finding Scientific Topics," *Proc. Nat'l Academy of Sciences USA*, vol. 101, no. suppl. 1, pp. 5228-5235, 2004.
[14] M. Steyvers, P. Smyth, M. Rosen-Zvi, and T. Griffiths, "Probabilistic Author-Topic Models for Information Discovery," *Proc. 10th ACM SIGKDD Conf. Knowledge Discovery and Data Mining*, 2004.
[15] Z. Guo, S. Zhu, Y. Chi, Z. Zhang, and Y. Gong, "A Latent Topic Model for Linked Documents," *Proc. 32nd Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR)*, 2009.
[16] T.-T. Pham, N.E. Maillot, J.-H. Lim, and J.-P. Chevallet, "Latent Semantic Fusion Model for Image Retrieval and Annotation," *Proc. 16th ACM Conf. Information and Knowledge Management (CIKM)*, 2007.
[17] R. Datta, D. Joshi, J. Li, and J.Z. Wang, "Image Retrieval: Ideas, Influences, and Trends of the New Age," *ACM Computing Surveys*, vol. 40, no. 2, article 5, pp. 1-60, 2008.
[18] F. Monay and D. Gatica-Perez, "On Image Auto-Annotation with Latent Space Models," *Proc. ACM Int'l Conf. Multimedia (MM)*, 2003.

[19] K. Barnard and D. Forsyth, "Learning the Semantics of Words and Pictures," *Proc. Int'l Conf. Computer Vision,* vol. 2, pp. 408-415, 2001.
[20] L.-J. Li and G. Wang, and L. Fei-Fei, "OPTIMOL: Automatic Online Picture Collection via Incremental Model Learning," *Int'l J. Computer Vision,* vol. 88, no. 2, pp. 147-168, 2010.
[21] J. Fan and Y. Gao, and H. Luo, "Integrating Concept Ontology and Multitask Learning to Achieve More Effective Classier Training for Multilevel Image Annotation," *IEEE Trans. Image Processing,* vol. 17, no. 3, pp. 407-426, Mar. 2008.
[22] G. Shafer, P.P. Shenoy, and K. Mellouli, "Propagating belief Functions in Qualitative Markov Trees," *Int'l J. Approximate Reasoning 1,* vol. 4, pp. 394-400, 1987.
[23] L.G. Shapiro, "GroundTruth Database," http://www.cs.washington.edu/research/imagedatabase/groundtruth/, Univ. of Washington, 2012.
[24] J. Pearl, *Probabilistic Reasoning in Intelligent Systems.* Morgan Kaufmann, 1988.
[25] L.D. Lowrance, T.D. Garvey, and T.M. Strat, "A Framework for Evidential Reasoning Systems," *Proc. Fifth Nat'l Conf. Artificial Intelligence (AAAI '86),* pp. 896-901, 1986.
[26] U. Montanari, "Networks of Constraints, Fundamental Properties and Applications to Picture Processing," *Information Science,* vol. 7, pp. 95-132, 1974.
[27] W. Woods, *Representation and Understanding,* D. Bobrow and A. Collins, eds. Academic Press, 1975.
[28] R.O. Duda, P.E. Hart, and N.J. Nilsson, "Subjective Bayesian Methods for Rule-Based Inference Systems," *Proc. Nat'l Computer Conf. and Exposition (AFIPS),* vol. 45, pp. 1075-1082, 1976.
[29] R. Schank, "Conceptual Dependency: A Theory of Natural Language Understanding," *Cognitive Psychology,* vol. 4, pp. 552-631, 1972.
[30] J.F. Sowa, *Conceptual Structures: Information Processing in Mind and Machine.* Addison-Wesley, 1984.
[31] O. Tuzel, F. Porikli, and P. Meer, "Pedestrian Detection via Classification on Riemannian Manifolds," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 30, no. 10, pp. 1713-1727, Oct. 2008.
[32] R. Howard, *Dynamic Probabilistic Systems.* John Wiley and Sons, Inc., 1971.
[33] G. Zhen, Z. Shenghuo, C. Yun, Z. Zhongfei, and G. Yihong, "A Latent Topic Model for Linked Documents," *Proc. 32nd Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR '09),* 2009.
[34] W.J. Stewart, *Numerical Solution of Markov Chains.* Princeton Univ. Press, 1994.
[35] http://people.kyb.tuebingen.mpg.de/pgehler/code/index.html, 2012.

Konstantinos A. Raftopoulos received the degree in mathematics from the University of Athens, Greece, in 1994, the master of science degree in computer science and the engineer degree, in 1996 and 1998, respectively, both from the University of California at Los Angeles (UCLA), California, and the doctorate degree from the National Technical University of Athens (NTUA), Athens, Greece. He has worked as a researcher at the UCLA Data Mining Lab, the Laboratory of Neuro Imaging, Brain Research Institute, UCLA Medical School, the Telecom Lab at NTUA, and the Image, Video and Multimedia Systems Laboratory at NTUA. Dr Raftopoulos is also an entrepreneur with start up companies in IT and industrial technology. He has been involved in several national, NIH, and EU research projects. His research is on semantic web, knowledge mining and representation, shape representation and analysis, machine learning, and biologically inspired computational models. He is a member of the IEEE.



Klimis S. Ntalianis received the diploma degree and the PhD degree in electrical and computer engineering, both from the National Technical University of Athens (NTUA), Athens, Greece, in 1998 and 2002, respectively, with support from the National Scholarship Foundation. He has also received a scholarship from the Institute of Communications and Computers Systems at NTUA as one of the best PhD students. He is the author of more than 45 scientific articles. His research interests include 3-D image processing, video summarization, content-based indexing, and retrieval as well as video source modeling and transmission. Dr Ntalianis is an assistant professor at the Technical Educational Institute of Athens and had been involved in many national and EU projects.



Dionyssios D. Sourlas received the five year diploma from the National Technical University of Athens in 1988 and a PhD degree from the University of California, Los Angeles (UCLA), both in chemical engineering. He has been an assistant professor in chemical engineering at the University of Missouri-Rolla, and since 2002, he has been a technical manager with Unilever Hellas. His published research and research interests include the areas of control theory (decentralized and robust control and modeling), optimization methods (global and infinite dimensional optimization), process operations (polymers), and biological systems, safety and toxicology. He is the recipient of 1998 AIChE-CAST Division Ted Peterson Best Student Paper Award for "Research on Decentralized Control."

Stefanos D. Kollias received the diploma in electrical engineering from the National Technical University of Athens (NTUA) in 1979, the MSc degree in communication engineering from the University of Manchester in England in 1980, and the PhD degree in signal processing from the Computer Science Division of NTUA in 1984. In 1982, he was given a COMSOC Scholarship by the IEEE Communication Society. Since 1986, he has served as a lecturer and an assistant and associate professor in the ECE Department at NTUA. From 1987 to 1988, he was a visiting research scientist in the Department of Electrical Engineering and the Center for Telecommunications Research at Columbia University in New York. Since 1997, he has been a professor at NTUA and Director of the Intelligent Multimedia Systems Lab. His research interests include multimedia analysis, retrieval, semantic web and knowledge technologies, human computer interaction, multimodal emotion recognition, artificial intelligence, neural networks, interoperability, and semantic metadata analysis of cultural content. He has published more than 200 papers, 100 of which were in international journals. He is editor of the book *Multimedia Content and the Semantic Web*, Wiley, 2005. He has been associate editor of the *IEEE Transactions on Neural Networks*, *IEEE Transactions on Multimedia*, and the *Neural Networks Journal*. He is a member of the executive committee of the European Neural Network Society (2007-today). He has been general chair of the VLBV '01 (Int'l Workshop on Multimedia Analysis), the ICANN '06 (Int'l Conference on Neural Networks) and of CBMI '09 (Content-based Multimedia Indexing). He has been a national representative in the EC Group on Cultural Heritage (2004-2007) and in the Member State Expert Group on CH (2007-today). He is a member of the IEEE.