International Journal on Artificial Intelligence Tools

Building an image annotation and tourism recommender system

KONSTANTINOS PLIAKOS
Department of Informatics, Aristotle University of Thessaloniki
Thessaloniki, 54124, Greece.
kpliakos@aiia.csd.auth.gr

CONSTANTINE KOTROPOULOS
Department of Informatics, Aristotle University of Thessaloniki
Thessaloniki, 54124, Greece.
costas@aiia.csd.auth.gr

Received (Day Month Year)
Revised (Day Month Year)
Accepted (Day Month Year)

The interest in image annotation and recommendation has been increased due to the ever rising amount of data uploaded to the web. Despite the many efforts undertaken so far, accuracy or efficiency still remain open problems. Here, a complete image annotation and tourism recommender system is proposed. It is based on the probabilistic latent semantic analysis (PLSA) and hypergraph ranking, exploiting the visual attributes of the images and the semantic information found in image tags and geo-tags. In particular, semantic image annotation resorts to the PLSA, exploiting the textual information in image tags. It is further complemented by visual annotation based on visual image content classification. Tourist destinations, strongly related to a query image, are recommended using hypergraph ranking enhanced by enforcing group sparsity constraints. Experiments were conducted on a large image dataset of Greek sites collected from `Flickr`. The experimental results demonstrate the merits of the proposed model. Semantic image annotation by means of the PLSA has achieved an average precision of 92% at 10% recall. The accuracy of content-based image classification is $82,6\%$. An average precision of 92% is measured at 1% recall for tourism recommendation.

Keywords: Probabilistic latent semantic analysis (PLSA); Clustering; Image Classification; Image Annotation; Recommender systems; Hypergraph.

1. Introduction

Nowadays, the overwhelming number of images uploaded to the web in conjuction with the constantly rising popularity of social media sharing platforms has led to an indisputable need for efficient annotation and recommender systems. Browsing through this vast volume of data resorts primarily to search engines, which harness the text information provided by the image tags or the image title. Image annotation is a crucial procedure, as it affects considerably both the retrieval accuracy of search engines and the organization of the images uploaded to the web. Several photo

sharing websites like `Flickr`[a], `Instagram`[b] or `PicasaWebAlbum`[c] enable the users to annotate images, facilitating image search and content description, bridging this way the gap between the semantic and the visual content of an image. However, quite often, the user provided tags are far from being accurate, some are redundant, or even contain false information. Consequently, an accurate and efficient annotation model is of crucial importance.

Tourism is a vital economic sector for Greece as well as other countries. Nowadays, the sector of e-tourism is thriving, launching a plethora of initiatives or entrepreneurial plans. The way people choose their tourist destination has evolved. Traditional ways, such as brochures or a simple web search have been substituted by tourist recommender systems. Despite the effort made so far, there are persisting problems in achieving satisfactory efficiency and accuracy to be addressed, amplifying the need of efficient tourism recommender systems. Here, a tourism recommender system is proposed and experimental results are disclosed, demonstrating its great potential.

In the past, many efforts have been made toward handling large scale tourism-related image datasets. In Ref. 1, the problem of the vast amount of images was handled by building an Internet landmark recognition engine, resorting to efficient object recognition and unsupervised clustering techniques. In Ref. 2, a cluster-based landmark and event detection scheme was presented that was based on clustering performed on both visual and tag similarity graphs. In Ref. 3, the problem of landmark classification in large-scale image datasets was addressed. Models for these landmarks were learnt by a multiclass support vector machine, using vector-quantized interest point descriptors as features. The classification rate was significantly improved by using temporal constraints and textual information distilled from image tags.

Image annotation has been studied extensively in the past using probabilistic models or graph-based methods. In Ref. 4, an image and video annotation model was proposed based on the joint probability distribution of tags and image feature vectors. The tag probabilities were computed, using a multiple Bernoulli model and the probabilities of image features were obtained, using non-parametric kernel density estimates. In Refs. 5 and 6, image auto-annotation models were implemented by applying the probabilistic latent semantic analysis (PLSA)[7]. In Ref. 8, image annotation was refined by integrating the PLSA with a Random Walk model. The PLSA was employed to estimate the posterior probabilities of each tag for an image. Next, a label similarity graph was constructed by a weighted linear combination of labels and visual similarities. A random walk process over the label graph was performed, mining the correlation of the tags and further refining this way the image annotation process. A joint probabilistic model was proposed for simultaneous image

[a] https://www.flickr.com/
[b] http://instagram.com/
[c] http://picasa.google.com/

classification and annotation in Ref. 9. It was based on a multi-class extension of the supervised Latent Dirichlet Analysis (sLDA) introduced in Ref. 10. Graph-based methods were proposed for tagging in Refs. 11 and 12, capturing the information from multi-type interrelated objects. In Ref. 13, a K-NN sparse graph-based semi-supervised approach was proposed for image annotation. A label training refinement strategy was proposed within this K-NN graph-based learning framework, handling the noise in the training labels by using a dual regularization for both the quantity and sparsity of the noise.

Hypergraph models have been exploited in many works dealing with recommendation problems. In Ref. 14, a unified hypergraph model was proposed for music recommendation. Music recommendation was treated as a hypergraph ranking problem, using both social media information and music acoustic-based content. In Ref. 15, a hypergraph model was employed for personalized image recommendation, harnessing the high-order relations among the users, the tags, and the images. The recommendation results were improved by enforcing group sparsity constraints. In Ref. 16, image tagging and geo-location prediction were treated simultaneously, modelling the users, the user social groups, the images, the image tags, and the geo-tags within a unified hypergraph framework. The hypergraph learning process was accelerated, using Singular Value Decomposition (SVD) analysis[17]. In Ref. 18, news recommendation was formulated as a ranking problem on fined-grained hypergraphs and a transductive inference approach was proposed to solve the so-called *cold-start* problem, (i.e., the problem of recommending news on a new user without much reading history). Ranking on hypergraphs has been also employed in Refs. 19, 20, 21. A related work in tourism recommendation is that of L. Cao *et al.*[22], where recommendation was based on clustering of geotagged images by location and visual matching.

The main contribution of this paper is in the development of a complete image annotation and tourism recommender system, extending and complementing the work presented in Ref. 23. In particular, the problems of semantic image annotation, content-based image classification, and tourism recommendation are addressed in a unified framework. To begin with, geo-tagged images crawled from `Flickr` are clustered by means of the GPS coordinates (latitude, longitude), forming several geographical clusters, called *geo-clusters* hereafter. The geo-clusters are then sorted according to their density (i.e., the number of images they contain). This way, places of interest (POIs) are defined. The underlying rationale is that popular tourist destinations attract more visitors, who upload more geo-tagged photos to the web. The text information (e.g., titles, tags) associated to the images that belong to each geo-cluster is concatenated, forming a geo-cluster derived document. A term-document matrix is then created, and the PLSA[7,24,25], properly initialized, is applied to it. The PLSA is used to represent the documents as probability distributions of topics treated as unobserved class variables. By applying the PLSA to a term-document matrix the relations between the terms and the documents are captured by observing the probability distribution between the documents and the generated topics as

well as between the topics and terms. Here, the PLSA is the heart of the method proposed for semantic image annotation and pertains the hypergraph ranking employed for tourism recommendation. The experimental results, obtained by replacing the PLSA with the LDA[26], did not show any improvement. Thus, the PLSA is preferred due to its simplicity. The semantic annotation is performed by assigning the most strongly related terms with the geo-cluster derived document to the associated geo-cluster as well as all images belonging to it.

Semantic image annotation is complemented by exploiting the visual attributes of images. The visual content annotation is based on visual content classification, treating the class label as a global image description. Here, the Support Vector Machine (SVM) classifier[27] is fed by GIST descriptors[28] in order to classify each image to a predefined number of classes, outperforming the simple method used in Ref. 23. Among the most common kernel functions used in the SVM are the linear, the polynomial, the Gaussian radial basis function (RBF), and the sigmoid. It was found that the RBF kernel outperforms the other ones.

Tourism recommendation is addressed using a hypergraph. A hypergraph is defined as a set of vertices made by concatenating different kinds of objects (e.g., documents, topics, terms) and hyperedges linking these vertices. In contrast to simple graphs, multi-link relations between the vertices are captured by hypergraphs[29]. Here, the hypergraph vertices are the annotation terms, the geo-cluster derived documents, and the latent topics derived by the PLSA. This way, the relations computed by the PLSA between the geo-cluster derived documents and the topics as well as the vocabulary terms are modeled. Tourism recommendation is treated as a hypergraph ranking problem and the top ranked geo-clusters are recommended as tourist destinations. Here, the recommendation resorts to the hypergraph ranking enhanced by enforcing group sparsity constraints, such as the group Least Absolute Shrinkage and Selection Operator (LASSO)[30,31,32]. This way, both the sparsity and the group structure of the data are exploited. The effect of each object group in the recommendation process is controlled separately by assigning them different weights. Accordingly, a more advanced method is developed than that in Ref. 23. In order to accelerate the solution of the optimization problem, an efficient method is proposed, using the Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) proposed in Ref 33. This algorithm is an extension of the Iterative Shrinkage-Thresholding Algorithm (ISTA), preserving its computational simplicity, but achieving a significantly faster rate of convergence.

The block diagram of the proposed model is depicted in Fig. 1. The user gives a test image as input to the system and the image is annotated semantically, geographically, and visually, as described in Sections 3.1 and 3.3. Proceeding to tourism recommendation, the query vector is set, as in Section 4.2. The hypergraph model is formed by the geo-cluster derived documents, topics, and terms, exploiting the results of the annotation process. Hypergraph ranking is applied to geo-cluster derived documents, topics, and terms and the top ranked geo-cluster documents are recommended as tourism POIs.

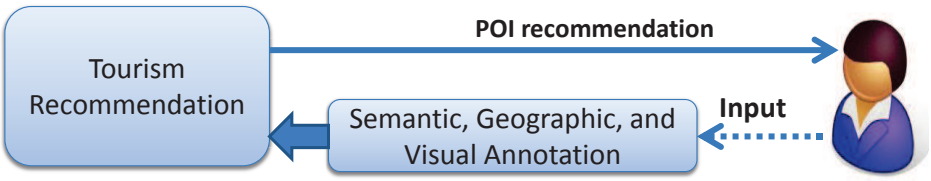

Fig. 1. Annotation and recommender system.

Promising experimental results are disclosed. In particular, an average precision of 92% at 10% recall is reported for semantic image annotation. The accuracy of content-based image classification of 310 test images over 14 classes is 82,6%. For tourism recommendation, an average precision of 92% is measured at 1% recall, indicating the effectiveness of the proposed recommendation method.

The structure of this paper is organized as follows. The dataset and the tourism POI identification method are described in Section 2. In Section 3, the image annotation is detailed. Hypergraph construction is explained in Section 4.1. The hypergraph ranking model is analyzed in Section 4.2. The optimization solution enforcing group sparsity is detailed in Sections 4.3 and 4.4. The outline of the proposed system is presented in Section 5. In Section 6, experimental results are presented, demonstrating the effectiveness of the proposed method. Conclusions are drawn and topics for future research are suggested in Section 7.

2. Dataset and tourism POI Identification

Popular tourist destinations attract more visitors, who upload more geo-tagged images on social media sharing platforms. To properly organize such geo-tagged images into geographical clusters, an hierarchical clustering algorithm was applied, that is based on geographical distances computed by the "Haversine formula" [d]. Thus, from 50000 geo-tagged randomly selected images related to Greece, that were collected from `Flickr`, 4660 geo-clusters were formed. From these geo-clusters, only the 500 most dense were considered as tourism POIs, containing 31814 images. The number of 500 was chosen taking into account the trade-off between the computational complexity and the information preservation. Next, a document was created for each geo-cluster by concatenating the text information (e.g., title, tags) available in all the images assigned to the geo-cluster.

Text information related to 150000 images, not including the 50000 images mentioned above, was also crawled from `Flickr` in order to properly capture the context of the tourism application. All characters were converted to lower case. Unreadable symbols and redundant information were removed. A vocabulary of unique words was generated along with their frequencies. Terms appearing with frequency less than a threshold (e.g., 100) were eliminated, yielding a vocabulary of 1901 terms.

[d] http://www.movable-type.co.uk/scripts/latlong.html

3. Image Annotation

3.1. *Image Annotation Using Semantic Topics*

The PLSA performs a probabilistic mixture decomposition, which associates an unobserved class variable to co-occurrences of terms and documents. The PLSA models each term in a document as a sample from a mixture model. The mixture components are multinomial random variables that can be interpreted as topic representations. The data generation process can be described as[7,24]: 1) select a document d with probability $P(d)$, 2) pick a latent topic z with probability $P(z|d)$ and, 3) generate a term t with probability $P(t|z)$.

Let $t \in T = \{t_1, t_2, \cdots, t_k\}$ be a vocabulary term and $d \in D = \{d_1, d_2, \cdots, d_m\}$ denote a document. The joint probability model is defined by the mixture:

$$P(t,d) = P(d)\ P(t|d) \quad \text{where} \tag{1}$$

$$P(t|d) = \sum_{z \in Z} P(t|z)\ P(z|d) \tag{2}$$

and $z \in Z = \{z_1, z_2, \cdots, z_n\}$ is an unobserved class variable representing the topics. As is indicated in (2), the document specific term distribution $P(t|d)$ is a convex combination of the n topic dependent distributions $P(t|z)$. In order to determine $P(d)$, $P(z|d)$, and $P(t|z)$, the log-likelihood function

$$\mathcal{L} = \sum_{d \in D} \sum_{t \in T} \tilde{n}(d,t)\ \log P(d,t) \tag{3}$$

has to be maximized with respect to all the aforementioned probabilities. In eq.(3), $\tilde{n}(d,t)$ denotes the term-document frequency. That is, the number of times t occurs in d. The estimation of $P(d)$ can be carried out independently resulting in $P(d) = \frac{\tilde{n}(d)}{\sum_{d' \in D} \tilde{n}(d')}$. The conditional probabilities $P(z|d)$ and $P(t|z)$ are estimated by means of the EM algorithm [34,35], which alternates between the *Expectation (E)-step:*

$$\hat{P}(z|d,t) = \frac{P(t|z)P(z|d)}{\sum_{z' \in Z} P(t|z')P(z'|d)}. \tag{4}$$

and the *Maximization (M)-step:*

$$P(t|z) = \frac{\sum_{d \in D} \tilde{n}(d,t)\hat{P}(z|d,t)}{\sum_{d \in D} \sum_{t' \in T} \tilde{n}(d,t')\hat{P}(z|d,t')} \tag{5}$$

$$P(z|d) = \frac{\sum_{t \in T} \tilde{n}(d,t)\hat{P}(z|d,t)}{\tilde{n}(d)}. \tag{6}$$

By alternating eq.(4) with eqs.(5)-(6), a convergent procedure is obtained to a local maximum of the log-likelihood. The annotation procedure is performed as follows:

1 The PLSA is applied to the term-document matrix $\mathbf{A} \in \mathbb{R}^{k \times m}$. Here, the documents are formed by concatenating any terms in the tags or the title of

the images that belong to a geo-cluster, forming the columns of **A**. That is, a document $d \in D$ is represented by a vector of size k, having as elements the frequency of occurrence of each term in d.

2 For each document to be annotated, the most related topic is chosen, that with the highest probability, i.e., $z^* = \arg\max_{z \in Z} P(z|d)$.

3 The $k' << k$ most related terms to z^* are identified by sorting $P(t|z^*)$ in decreasing order of magnitude. k' is the same for any hidden topic.

Here, the number of hidden topics is set empirically. However, there are techniques, such as those based on information criteria (e.g., the Bayesian Information Criterion or the method in Ref. 36) that could be adapted to PLSA. The term document matrix **A** is of size 1901×500. Among the most descriptive terms of a document, those providing geographical information are identified using geo-gazetteers[e]. Thus, a complete annotation model is built, which provides geographic information in addition to the semantic information.

3.2. *PLSA Initialization*

The PLSA depends on proper initialization method. In addition to the common random initialization, there are many other schemes, e.g., the Random C (RC) or the Random Acol (RAcol)[24]. A variant of RC is the Dense Random C (DRC) summarized in Algorithm 1. The DRC treats the columns of **A** unequally. Only the most dense columns are chosen, as they provide more valuable information. The reduction of the number of the columns makes the method less time consuming. The DRC was found to be more effective than the RC and the RAcol in the experiments conducted.

Algorithm 1 Dense Random C Initialization

Input: matrix $\mathbf{A} \in \mathbb{R}^{k \times m}$ with $A(i,j) \geq 0$.
Output: matrix $\mathbf{S} \in \mathbb{R}^{k \times n}$, containing the conditional probabilities $P(t|d)$.

1 Count the non-zero elements of each column of **A**.
2 Compute the mean document vector $\overline{\mu}$.
3 Find the ξ columns of **A**, having more non-zero elements than $\overline{\mu}$.
4 Average $\tilde{\zeta}$ randomly chosen columns out of the ξ and set the average column vector as a column of **S**. Repeat 3-4 for all columns of **S**.

3.3. *Classification-based Visual Content Annotation*

The visual features of an image provide valuable, complementary, information about its content. Image annotation is strongly related to image classification, considering

[e]http://www.geonames.org

the class label as a global description of the image, while the tags are treated as local description of the individual image parts. In order to construct a proper visual dataset, an image subset made of images without occlusion or unwanted noise was manually extracted from the aforementioned dataset of 31814 images and annotated using l different class labels. Here, the different visual classes have been manually defined, capturing the various themes, pertaining the image dataset. Each visual topic (class), was assigned manually one label and a few representative tags, e.g., clouds, sky, sea, sunset, defining the image visual content.

The GIST descriptor[28], was extracted from any image $\phi \in \Phi = \{\phi_1, \phi_2, \cdots, \phi_N\}$, where Φ is a smaller image subset, made of images without occlusion or unwanted noise. Φ is manually extracted and annotated using the l class labels. The GIST descriptors are global descriptors initially proposed in Ref. 28. Each image ϕ is segmented by a 4-by-4 grid for which orientation histograms are extracted, producing a final vector of low dimensionality, the GIST descriptor. The GIST descriptors have recently demonstrated good results for image search tasks on large databases[37]. Local descriptors, such as SIFT[38] represented in a bag of features approach, were not used due to their heavy computational complexity and their memory needs.

An SVM classifier was trained using the GIST descriptors G as input vectors. Each test image was classified into one among the l classes by applying the SVM classifier fed by the GIST descriptors of the test images. Here, the RBF kernel was employed. It was found that it outperforms the other commonly used kernels (i.e., the linear, the polynomial, and the sigmoid). The multiclass problem was treated by applying the "one against one" strategy, also known as pairwise coupling. It is based on the construction of one SVM for each pair of classes. Thus, for a problem with l classes, $l(l-1)/2$ SVMs are trained to recognize the samples of one class from the samples of its rivals. Each SVM votes for one class. The classification of a test sample is achieved by majority voting. Representative images assigned to several classes are shown in Fig. 2.

4. Tourism Recommendation

The second part of the proposed system consists of a hypergraph model representing the multi-link relations between terms of the vocabulary, documents (geo-clusters), and topics as they were computed in Sec. 3.1.

4.1. *Hypergraph Construction*

A hypergraph $\mathbf{H}$ having size of 2751×1000 elements was formed by concatenating 500 documents associated to the geo-clusters, 350 topics, z, and 1901 vocabulary terms, t. The vertex set is defined as $V = D \cup Z \cup T$. The structure of the hypergraph incidence matrix is summarized in Table 1. For each document d_j associated to a geo-cluster, a hyperedge e_1 is inserted as a column of $\mathbf{De}_1, \mathbf{Ze}_1$, and $\mathbf{Te}_1$, containing 1 in the j-th entry of $\mathbf{De}_1$, 1 for the most related topic to d_j, z^* in $\mathbf{Ze}_1$, and 30 ones for the 30 most descriptive terms t for z^*, in $\mathbf{Te}_1$. The weight for this hyperedge

Fig. 2. A sample of 16 images, representing several visual classes. Top raw: First image belongs to visual class sea, Second image belongs to visual class mills, Third image belongs to visual class sky, Fourth image belongs to visual class White Tower, Second raw: First image belongs to visual class meteora, Second image belongs to visual class Byzantine, Third image belongs to visual class artifacts, Fourth image belongs to visual class Parthenon, Third raw: Second image belongs to visual class city scene, Fourth raw: Second image belongs to visual class plants, and Fourth image belongs to image class sunset.

is $w(e_1) = P(z^*|d_j)$. To capture the geographical proximity, hyperedges $e_2 \in E_2$ are created. For each d_j corresponding to a specific geo-cluster, one hyperedge e_2 is inserted as a column of $\mathbf{De}_2$ having 1 to its j-th entry, as well as to the entries corresponding to geo-clusters being at a geographical distance less than 150 km. The weight for this hyperedge is set to 1.

Table 1. Structure of $\mathbf{H}$.

	e_1	e_2
D	$\mathbf{De}_1$	$\mathbf{De}_2$
Z	$\mathbf{Ze}_1$	0
T	$\mathbf{Te}_1$	0

4.2. *Hypergraph model*

Hereafter, set cardinality is denoted by $|\cdot|$, the ℓ_2 norm of a vector appears as $\|.\|$ and $\mathbf{I}$ is the identity matrix of compatible dimensions. $\Psi(V, E, w)$ denotes a hypergraph with set of vertices V and set of hyperedges E to which a weight function $w : E \to \mathbb{R}$

is assigned. V consists of sets of objects of different type (i.e., documents, topics, terms). A $|V| \times |E|$ incidence matrix $\mathbf{H}$ is formed having as elements:

$$H(v,e) = \begin{cases} 1, & \text{if } v \in e \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

Based on $\mathbf{H}$, the vertex and hyperedge degrees are defined as:

$$\delta(v) = \sum_{e \in E} w(e) H(v,e) \tag{8}$$

$$\delta(e) = \sum_{v \in V} H(v,e). \tag{9}$$

The following diagonal matrices are defined: the vertex degree matrix $\mathbf{D}_u$ of size $|V| \times |V|$, the hyperedge degree matrix $\mathbf{D}_e$ of size $|E| \times |E|$, and the $|E| \times |E|$ matrix $\mathbf{W}$ containing the hyperedge weights.

Let $\mathbf{\Theta} = \mathbf{D}_u^{-1/2}\mathbf{H}\mathbf{W}\mathbf{D}_e^{-1}\mathbf{H}^T\mathbf{D}_u^{-1/2}$. $\mathbf{\Theta}$ is a symmetric matrix as the diagonal matrices $\mathbf{W}$ and $\mathbf{D}_e^{-1}$ commute in multiplication. $\mathbf{\Theta}$ has size $|V| \times |V|$ and its elements, $\Theta(j,i)$, indicate the relatedness between the vertices j and i. Considering j, i as two vertices of the same type (e.g., document(j), document(i)) the element $\Theta(i,j)$ indicates the similarity of the two documents. In case j, i corresponds to two vertices of different type (e.g., document(j), term(i)) then the element $\Theta(j,i)$ indicates how much the document(j) is related to the term(i). Then, $\mathbf{L} = \mathbf{I} - \mathbf{\Theta}$ is known as Zhou's normalized Laplacian of the hypergraph[39]. To perform clustering on a hypergraph one is seeking for a real-valued ranking vector $\mathbf{f} \in \mathbb{R}^{|V|}$, minimizing the cost function:

$$\Omega(\mathbf{f}) = \mathbf{f}^T\mathbf{L}\mathbf{f}. \tag{10}$$

in terms of the ranking vector $\mathbf{f}$. That is, one requires all vertices with the same value in the ranking vector $\mathbf{f}$ to be strongly connected[40]. $\Omega(\mathbf{f})$ is small, if vertices with high affinities have the same label. For instance, two documents are probably similar, if they are linked with many common topics and textual terms. The aforementioned optimization problem was extended to solve a recommendation problem by including the ℓ_2 regularization norm between the ranking vector $\mathbf{f}$ and a query vector $\mathbf{y} \in \mathbb{R}^{|V|}$ in Ref. 14. This guarantees that the ranking vector does not differ too much from the initial query vector $\mathbf{y}$. The function to be minimized is then expressed as

$$Q(\mathbf{f}) = \Omega(\mathbf{f}) + \vartheta \, ||\mathbf{f} - \mathbf{y}||^2 \tag{11}$$

where ϑ is a regularizing parameter. The best ranking vector, $\mathbf{f}^* = \arg\min_{\mathbf{f}} Q(\mathbf{f})$, is found to be[14]:

$$\mathbf{f}^* = \frac{\vartheta}{1+\vartheta}\left(\mathbf{I} - \frac{1}{1+\vartheta}\mathbf{\Theta}\right)^{-1}\mathbf{y}. \tag{12}$$

Let d'_j be the geo-cluster where the test image ϕ_{test} belongs to with respect to its geo-tag. The query vector $\mathbf{y} \in \mathbb{R}^{|V|}$ is defined as:

$$y(v) = \begin{cases} 1, & \text{if } v = d'_j \\ \Theta(d'_j, v), & \text{otherwise} \end{cases} \tag{13}$$

treating $\Theta(d'_j, v)$ as a measure of relatedness between the vertices of the hypergraph.

The ranking vector $\mathbf{f}^* \in \mathbb{R}^{|V|}$ has the same size and structure as $\mathbf{y}$. It corresponds to a significant query of the system. In other terms, for every query vector $\mathbf{y}$ the optimization procedure presented in (11) is performed and the values of the ranking vector $\mathbf{f}$ corresponding to the entries associated to geo-cluster documents are used as rankings for tourist destination recommendation. The top ranked geo-cluster documents are recommended as tourism POIs to the user, who has imported the test image ϕ_{test}.

4.3. *Group Sparse Regularization*

The hypergraph vertices are split into Δ non-overlapping object groups (geo-cluster derived documents, topics, terms). Undoubtedly, each object group contributes differently to the ranking procedure. A Group Lasso regularizing term is more appropriate than the ℓ_2 norm in this kind of problems[31]. Accordingly, different weights γ_δ, $\delta = 1, 2, \ldots, \Delta$ are assigned to each object group, yielding the following objective function to be minimized:

$$\tilde{Q}(\mathbf{f}) = \Omega(\mathbf{f}) + \vartheta \sum_{\delta=1}^{\Delta} \sqrt{\gamma_\delta \, (\mathbf{f} - \mathbf{y})^T \mathbf{K}_\delta (\mathbf{f} - \mathbf{y})}. \tag{14}$$

In (14), ϑ is a positive regularization parameter and $\mathbf{K}_\delta$ is the $|V| \times |V|$ diagonal matrix with elements equal to 1 for the vertices, which belong to the δ-th object group.

4.4. *FISTA Solution*

In order to solve the minimization problem (14) by applying the FISTA algorithm, one needs to transform the cost function $\Omega(\mathbf{f})$ to the following ℓ_2 norm[33]:

$$\Omega(\mathbf{f}) = \left|\left| \tilde{\mathbf{A}}\mathbf{x} - \mathbf{b} \right|\right|^2. \tag{15}$$

By introducing the auxiliary variable $\mathbf{x} = \mathbf{f} - \mathbf{y}$, the cost function (10) is rewritten as:

$$\Omega(\mathbf{f}) = (\mathbf{x} + \mathbf{y})^T \mathbf{L} (\mathbf{x} + \mathbf{y}). \tag{16}$$

The fact that $\mathbf{L}$ is a symmetric and positive definite matrix implies that its Cholesky factor $\mathbf{L}^{\frac{1}{2}}$ is also a symmetric positive definite matrix. Thus, the cost function (16)

is rewritten as:

$$\Omega(\mathbf{f}) = (\mathbf{x}+\mathbf{y})^T \mathbf{L}^{\frac{1}{2}} \mathbf{L}^{\frac{1}{2}} (\mathbf{x}+\mathbf{y}) = \left[\mathbf{L}^{\frac{1}{2}}(\mathbf{x}+\mathbf{y})\right]^T \left[\mathbf{L}^{\frac{1}{2}}(\mathbf{x}+\mathbf{y})\right]$$
$$= \left|\left|\mathbf{L}^{\frac{1}{2}}(\mathbf{x}+\mathbf{y})\right|\right|^2. \quad (17)$$

From Equations (15) and (17), it is deduced that:

$$\tilde{\mathbf{A}} = \mathbf{L}^{\frac{1}{2}}, \quad (18)$$

$$\mathbf{b} = -\mathbf{L}^{\frac{1}{2}}\mathbf{y}. \quad (19)$$

The optimization problem (14) takes the form:

$$\tilde{Q}(\mathbf{x}) = g(\mathbf{x}) + q(\mathbf{x}) \quad (20)$$

where $g(\mathbf{x}) = \left|\left|\mathbf{L}^{\frac{1}{2}}(\mathbf{x}+\mathbf{y})\right|\right|^2$, $q(\mathbf{x}) = \vartheta \sum_{\delta=1}^{\Delta} \sqrt{\gamma_\delta}||\mathbf{x}_\delta||$, and $\mathbf{x}_\delta$ refers to the entries of $\mathbf{x}$ related to group δ.

Given a point $\mathbf{x}$, the ISTA approach applied to $\tilde{Q}(\mathbf{x})$ minimizes the sum of $q(\mathbf{x})$ and a quadratic approximation of $g(\mathbf{x})$ at each iteration, i.e.

$$\mathbf{x}_{new} = \arg\min_{\tilde{\mathbf{z}}} \left\{ g(\mathbf{x}) + \nabla^T g(\mathbf{x})(\tilde{\mathbf{z}} - \mathbf{x}) + \frac{1}{2\varrho}||\tilde{\mathbf{z}} - \mathbf{x}||^2 + q(\tilde{\mathbf{z}}) \right\}$$
$$= \arg\min_{\tilde{\mathbf{z}}} \left\{ \sum_{\delta} \left\{ \frac{1}{\varrho}||\tilde{\mathbf{z}}_\delta - \tilde{\mathbf{d}}(\mathbf{x})_\delta||^2 + \vartheta\sqrt{\gamma_\delta}||\tilde{\mathbf{z}}_\delta|| \right\} \right\} \quad (21)$$

where ϱ is the step-length and $\tilde{\mathbf{d}}(\mathbf{x}) = \mathbf{x} - 2\varrho\nabla g(\mathbf{x})$.

The optimization problem (21) is separable. The solution of each of the Δ sub-problems given a soft-thresholding operation

$$\mathbf{x}_{new} = \frac{\tilde{\mathbf{d}}(\mathbf{x})_\delta}{||\tilde{\mathbf{d}}(\mathbf{x})_\delta||} \max\left\{0, ||\tilde{\mathbf{d}}(\mathbf{x})_\delta|| - \vartheta\sqrt{\gamma_\delta}\varrho\right\} \quad (22)$$

The step-length ϱ satisfies the inequality $\varrho \leq \frac{1}{\eta}$, where η is the maximal eigenvalue of $\mathbf{L}$ (i.e., the Lipschitz constant for $\nabla g(\mathbf{x})$).

Algorithm 2 summarizes the steps of the FISTA for $\varrho = \frac{1}{\eta}$. $\mathbf{x}^\infty$ denotes the vector $\mathbf{x}$ upon the convergence of the while loop in Algorithm 2.

5. System Outline

Given an image as input, the distances between image geo-location captured using the GPS technology and the geo-cluster centers are computed. The input image is then assigned to the nearest geo-cluster. Simultaneously, the image visual content is classified by means of an SVM classifier fed by the image GIST descriptor. Next, the class label and the predefined representative tags offer a visual content annotation. Simultaneously, the vocabulary terms assigned to the closest geo-cluster derived document by the PLSA, offer geographic and semantic annotation for the image,

Algorithm 2 Solving (14) with FISTA.

Input: The Laplacian matrix $\mathbf{L} \in \mathbb{R}^{|V|\times|V|}$, the query vector $\mathbf{y} \in \mathbb{R}^{|V|}$, $\gamma_\delta, \delta = 1, 2, \cdots, \Delta$, parameter ϑ, and maximum eigenvalue of $\mathbf{L}$, η.
Output: The ranking vector $\mathbf{f} \in \mathbb{R}^{|V|}$.

Initialize $\tau = 1$, $\zeta^{(1)} = 1$, $\mathbf{x}^{(0)}$, $\tilde{\mathbf{z}}^{(1)} = \mathbf{x}^{(0)}$.
while not converged **do**
  $\tilde{\mathbf{d}}(\tilde{\mathbf{z}}^{(\tau)}) = \tilde{\mathbf{z}}^{(\tau)} - \frac{2}{\eta}\mathbf{L}(\tilde{\mathbf{z}}^{(\tau)} + \mathbf{y})$
  for $\delta = 1 \cdots \Delta$ do
    $\mathbf{x}_\delta^{(\tau)} = \frac{\tilde{\mathbf{d}}(\tilde{\mathbf{z}}^{(\tau)})_\delta}{||\tilde{\mathbf{d}}(\tilde{\mathbf{z}}^{(\tau)})_\delta||} \max\left\{0, ||\tilde{\mathbf{d}}(\tilde{\mathbf{z}}^{(\tau)})_\delta|| - \vartheta\frac{\sqrt{\gamma_\delta}}{\eta}\right\}$
  end
  $\zeta^{(\tau+1)} = \frac{1+\sqrt{1+4(\zeta^{(\tau)})^2}}{2}$
  $\tilde{\mathbf{z}}^{(\tau+1)} = \mathbf{x}^{(\tau)} + \frac{\zeta^{(\tau)}-1}{\zeta^{(\tau+1)}}(\mathbf{x}^{(\tau)} - \mathbf{x}^{(\tau-1)})$
  $\tau \leftarrow \tau + 1$
end while
$\mathbf{f} = \mathbf{x}^{\infty} + \mathbf{y}$

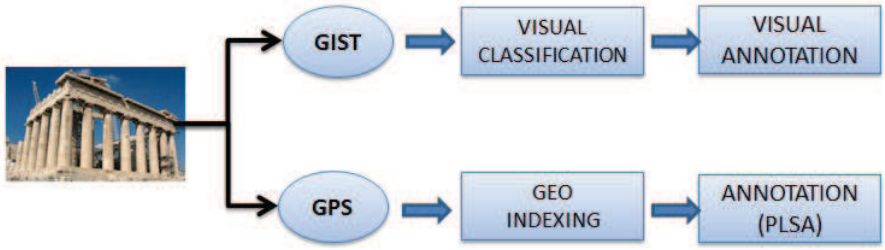

Fig. 3. Annotation system.

as was described in Sec. 3.1. Fig. 3 demonstrates the proposed annotation system. Proceeding to tourism recommendation, the query vector $\mathbf{y}$ is initialized, as was suggested in Sec. 4.2. Hypergraph ranking is applied and the top ranked geo-clusters are recommended as tourist destinations.

6. Experimental Results

The averaged Recall-Precision curve, the F_1 measure, and the Mean Average Precision (MAP) are used as figures of merit. Precision is defined as the number of correctly recommended objects divided by the number of all recommended objects. Recall is defined as the number of correctly recommended objects divided by the number of all objects. The F_1 measure is the weighted harmonic mean of precision and recall, which measures the effectiveness of retrieval when treating precision and

Table 2. F_1 measure at 7 ranking positions.

	F_1@1	F_1@2	F_1@5	F_1@10	F_1@15	F_1@20	F_1@25
PLSA	0.1704	0.2987	0.4926	0.5707	**0.5546**	**0.5213**	**0.4816**
LDA	0.1601	0.2829	0.4583	0.5275	0.5204	0.4884	0.4510
TF-IDF	0.1641	0.2867	0.4529	0.5261	0.5260	0.4993	0.4613
maxPLSA	**0.1882**	**0.3202**	**0.4996**	**0.5749**	0.5487	0.5175	0.4779

recall as equally important, i.e.,

$$F_1 = 2\frac{Precision \cdot Recall}{Precision + Recall}. \tag{23}$$

The MAP is the mean value of the Average Precision (AP) of all the queries. The AP is defined as the average of precisions computed at the point of each correctly retrieved item, as is shown below:

$$AP = \frac{\sum_{i=1}^{Num} Precision@i \cdot true_i}{cNum} \tag{24}$$

where $Precision@i$ is the precision at ranking position i, Num is the number of retrieved items, $cNum$ is the number of correctly retrieved items, and $true_i = 1$ if the item at position i is correctly retrieved.

For evaluation purposes, a test set containing 310 images was randomly chosen and removed from the training set along with their text information. The PLSA performance in semantic image annotation has been compared to that of the LDA[26] and the term frequency-inverse document frequency (TF-IDF)[41]. As is demonstrated in Fig. 4, the PLSA outperforms both the LDA and the TF-IDF method. The recall-precision curves were obtained by averaging recall-precision pairs in 100 repetitions using different initializations. The curve corresponding to the best results at the top ranking positions obtained by using the PLSA is denoted as *maxPLSA*. An average precision of 92% at 10% recall is reported, using the PLSA. It is worth noting, that the PLSA is much simpler than the LDA. In Table 2, the averaged F_1 measure corresponding to 7 different ranking positions is listed for the PLSA, the LDA, and the TF-IDF, for semantic image annotation. It is seen that the PLSA outperforms the LDA and the TF-IDF.

In Fig. 5, recall-precision curves are plotted for the PLSA, having been initialized by the DRC, the RC, the RAcol, and the Random initialization coined as (Rnd) for 2 and 4 iterations of the Expectation Maximization (EM) algorithm. The average recall precision curves over 100 repetitions of the PLSA training for each initialization are shown. The results indicate that the precision falls with a slower rate for the DRC.

As is depicted in Fig. 6, the MAP improves as the number of the topics increases. A plausible compromise between the annotation accuracy and the additional complexity caused by large number of topics, suggests fixing the number of topics to 350. In Fig. 7, recall-precision curves are demonstrated for several number of topics. The recall-precision curves were obtained by averaging recall-precision pairs in 100 repetitions.

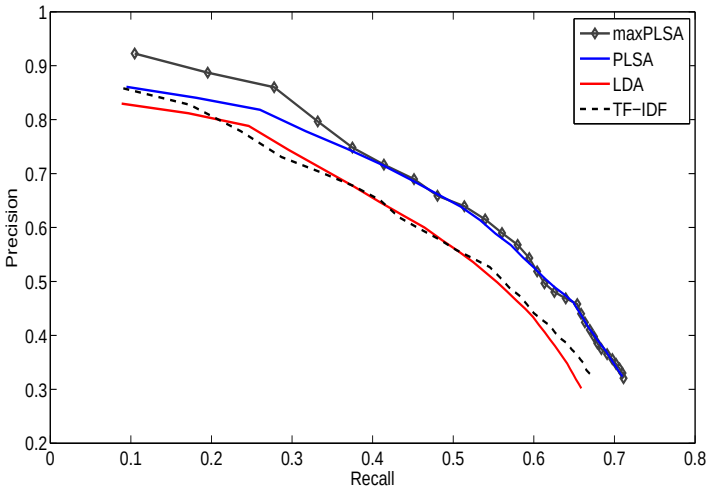

Fig. 4. Recall-precision for the PLSA, the LDA, and the TF-IDF.

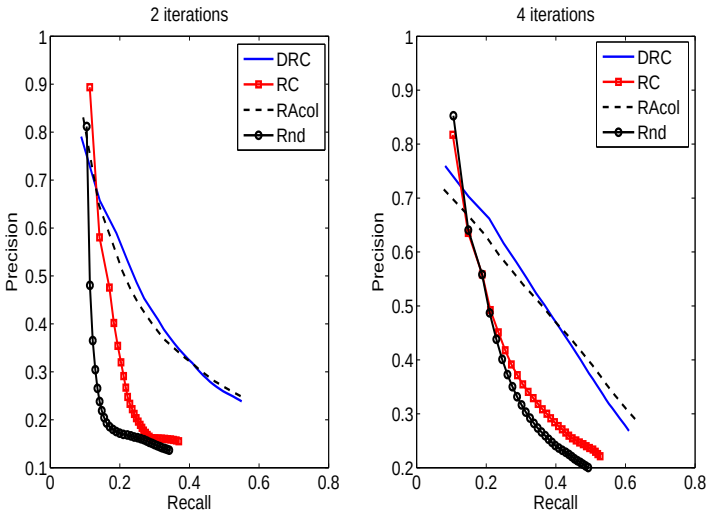

Fig. 5. DRC, RC, RAcol, and Rnd recall-precision curves for 2 and 4 iterations of the EM.

For visual image classification, the same test set was used. Each test image was assigned into one of 14 representative classes manually in order to form the ground truth. For each image one GIST descriptor is extracted. Visual classification accuracy for the proposed approach (GS), the approach presented in Ref. 23 (GK) and a simple approach employing GIST descriptors and majority voting (GMJ) is shown in Table. 3. The best results were obtained by using the SVM classifier.

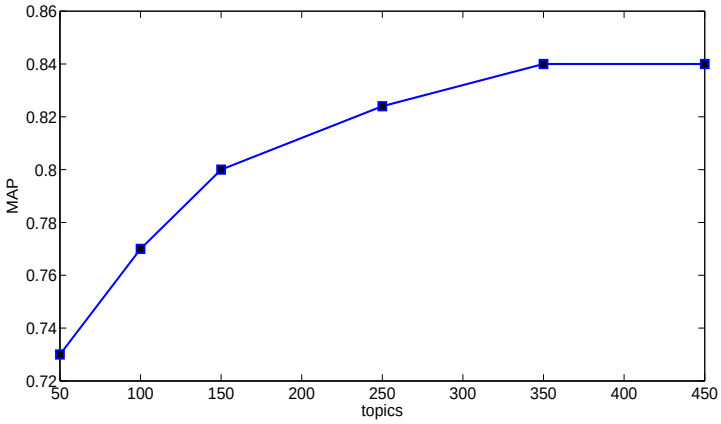

Fig. 6. *MAP* vs. the number of topics.

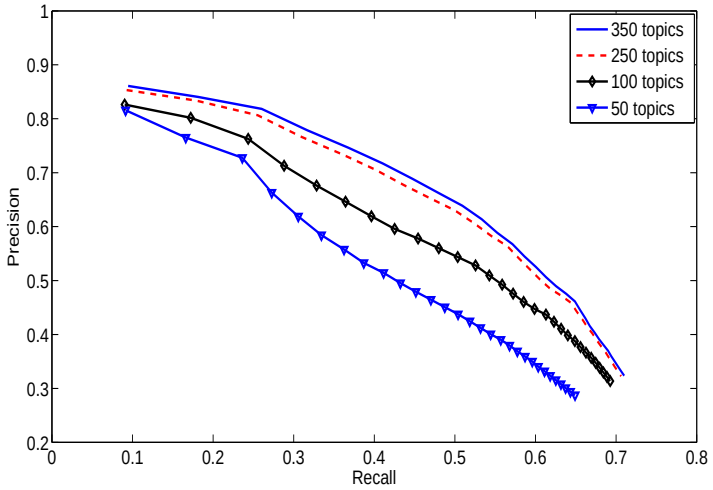

Fig. 7. PLSA recall-precision curves for several number of topics.

Across the 310 test set images, the accuracy of content-based image classification over 14 classes is $82,6\%$. In Table 4, the classification accuracy is demonstrated for 4 different SVM kernels, the linear, the polynomial, the sigmoid, and the RBF. The RBF kernel clearly outperforms the other 3. The width parameter $\tilde{\gamma}$ of the RBF kernel function $\exp(-\tilde{\gamma}||\chi_i - \chi_j||^2)$ was set equal to 1. In Fig. 8, the classification rates over the regularization parameter C that weighs the sum of slack variables[42] are displayed. The classification accuracy is reduced greatly for small values of the cost parameter C. A small value of C will cause the optimizer to look for a larger-

Table 3. Accuracy results for the 3 compared methods.

	GS	GK	GMJ
classification accuracy	**0.826**	0.700	0.607

Table 4. Accuracy results for 4 different SVM kernels.

	linear	*polynomial*	*sigmoid*	*RBF*
classification accuracy	0.784	0.784	0.439	**0.826**

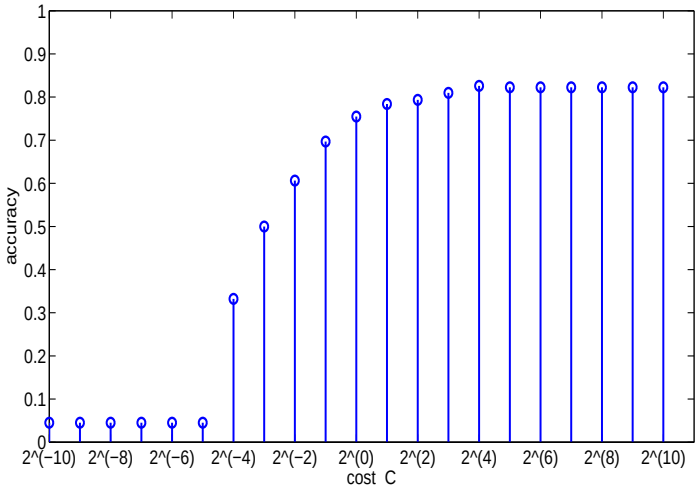

Fig. 8. Classification accuracy results per cost parameter C.

margin separating hyperplane, even if that hyperplane misclassifies more points. Both C and $\tilde{\gamma}$ were estimated in a grid search, performing 5-fold cross validation on the training set. In Fig. 9, the confusion matrix is shown, demonstrating the classification rates across the classes.

Moreover, experiments were conducted to assess tourism POI recommendation. In order to form the ground truth, relations were established manually among the geo-clusters, taking into account the distance, common geographical entities (e.g., mainland, island) and leisure activities. For this, various tourist related web sources were exploited, such as `TripAdvisor`[f] and `TravelMuse`[g]. Firstly, only the hyperedges $e_1 \in E_1$ were taken into account in hypergraph creation, as in Ref. 23. Secondly, all the hyperedges were considered. Let us denote the first approach as *HR1* and the one exploiting also the geographic information as *HR*. The associated recall-precision curves are plotted in Fig. 10. It is seen that the results are increased when

[f]http://www.tripadvisor.com.gr/
[g]http://www.travelmuse.com/

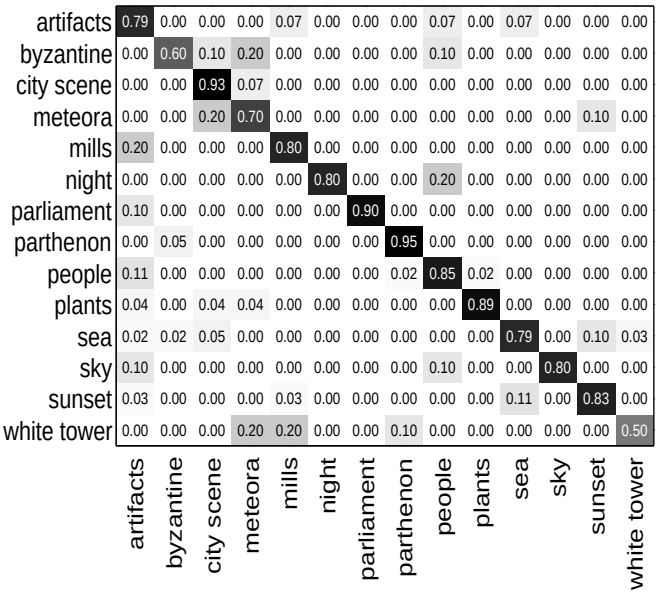

Fig. 9. Confusion matrix for the visual image classification.

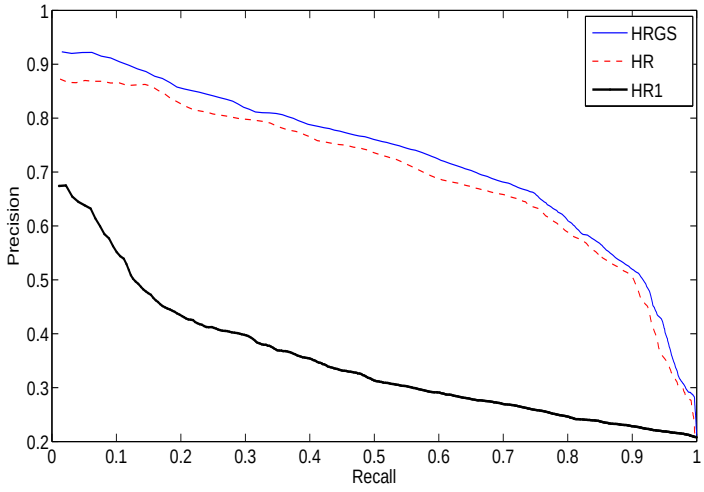

Fig. 10. Averaged Recall-precision curves for tourism POI recommendation.

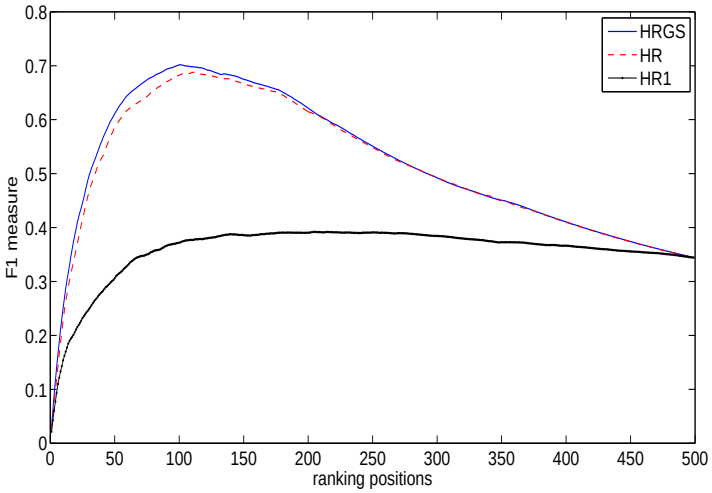

Fig. 11. Averaged F_1 measure at several ranking positions.

both types of hyperedges are considered.

Furthermore, the recommendation is improved by enforcing group sparsity constraints. The tourism POI recommendation enforcing group sparsity is coined as *HRGS*. In Fig. 10, the recall-precision curves for tourism POI recommendation are plotted revealing the superiority of the approach proposed here. As is demonstrated in Fig. 10, a precision rate of 92% and 91% is achieved for 1% and 10% recall, respectively. The complete curves of the averaged F_1 measure per ranking position for all algorithms compared are displayed in Fig. 11. The maximum average F_1 measure equals 0.7019. It is obtained at the ranking position 101. The MAP results for the HRGS, the HR, and the HR1 are 0.741, 0.710, and 0.3647, respectively. Here, the objects were split in 2 object groups, the first is formed by the geo-cluster derived documents and the second consists of the concatenation of both topics and terms. This choice was made by observing the sparsity distribution over the incidence matrix $\mathbf{H}$ of the hypergraph. Indeed, by applying spectral analysis to the hypergraph Laplacian matrix $\mathbf{L}$, these 2 object groups are clearly discriminated as can be seen in Fig.12. The columns of $\mathbf{L}$, are projected into the 2 dimensional space by employing spectral analysis and taking the 2 eigenvectors corresponding to the 2 non-zero smallest eigenvalues of $\mathbf{L}$. The weights for the 2 different object groups were empirically set to 0.9 for the geo-cluster derived documents and to 0.2 for the other group. In Fig. 13, a visual example of the proposed recommendation method on the Google maps is demonstrated where the user has inserted as a query an image of Parthenon and the proposed system recommended the Zappeion and the temple of Zeus.

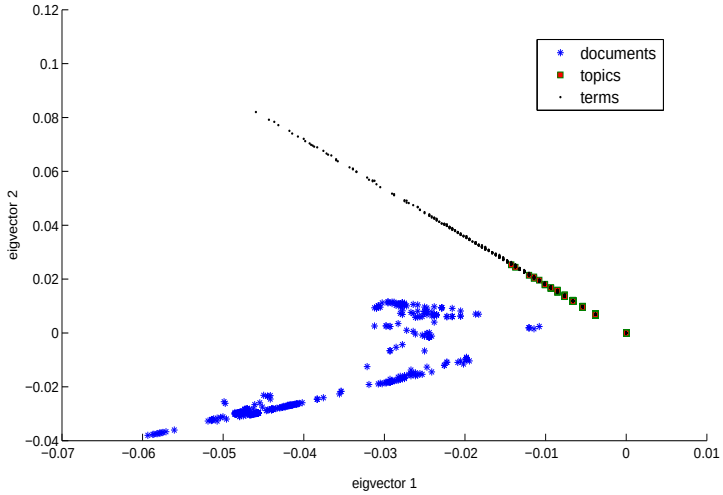

Fig. 12. Spectral analysis on the hypergraph Laplacian.

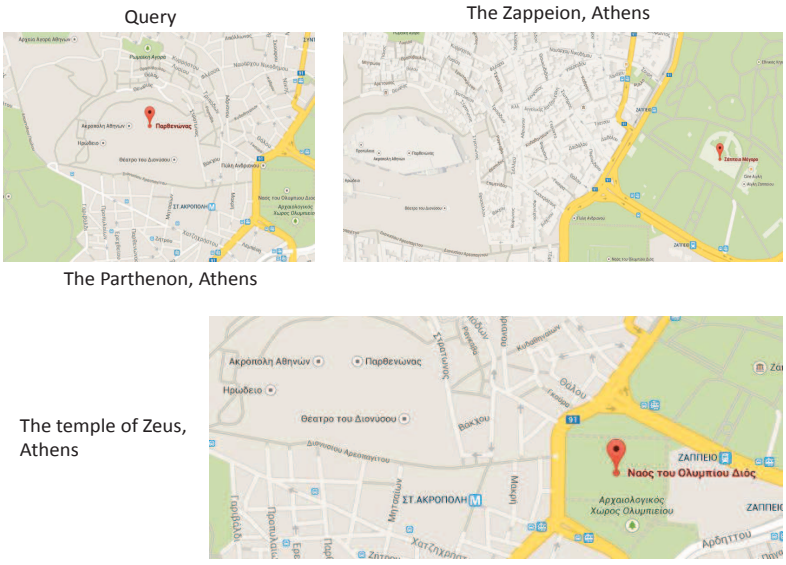

Fig. 13. A visual example of the recommendation method (Google maps).

7. Conclusions and Future Work

An efficient image annotation and tourism POI recommender system has been proposed. A method to organize large collections of images has been developed based on

clustering and classification. Tourism related images have been annotated geographically, semantically, and visually, harnessing visual attributes and text information. The PLSA has been enhanced by an effective initialization method and used in order to extract semantic information from image metadata. The annotation procedure has been complemented by exploiting the image visual attributes. Furthermore, a tourism recommendation approach has been proposed based on hypergraph ranking and enforcing group sparsity constraints. The solution of the optimization problem was significantly accelerated via the gradient accelerated method. Subjects of future research might be the exploitation of personalized user information and any other social media information and the on-line updating of the hyperedge weights. Another interesting approach would be the employment of the majorization-minimization algorithm proposed in Ref 43, enhancing the efficiency of the ranking in the experiments.

Acknowledgments

This research has been co-financed by the European Union (European Regional Development Fund - ERDF) and Greek national funds through the Operation Program "Competitiveness-Cooperation 2011" - Research Funding Program: 11SYN-10-1730-ATLAS.

References

1. Y. T. Zheng, M. Zhao, Y. Song, H. Adam, U. Buddemeier, A. Bissacco, F. Brucher, T. S. Chua, and H. Neven. Tour the world: building a web-scale landmark recognition engine, in Proc. IEEE Computer Vision and Pattern Recognition, pp. 1085–1092, 2009.
2. S. Papadopoulos, C. Zigkolis, Y. Kompatsiaris, and A. Vakali, Cluster-based landmark and event detection on tagged photo collections, IEEE Multimedia, pp. 52–63, 2011
3. L. Yunpeng, D. J. Crandall, and D. P. Huttenlocher, Landmark classification in large-scale image collections, in Proc. Int. Conf. Computer Vision, pp. 1957–1964, 2009.
4. S. L. Feng, R. Manmatha, and V. Lavrenko, Multiple Bernoulli relevance models for image and video annotation, in Proc. IEEE Conf. Computer Vision and Pattern Recognition, vol. 2, pp. 1002-1009, 2004.
5. F. Monay and D. Gatica-Perez, On image auto-annotation with latent space models, in Proc. ACM Int. Conf. Multimedia, pp. 275–278, 2003.
6. F. Monay and D. Gatica-Perez, PLSA-based image auto-annotation: constraining the latent space, in Proc. ACM Int. Conf. Multimedia, pp. 348–351, 2004.
7. T. Hofmann, Probabilistic latent semantic analysis, in Proc. 15th Conf. Uncertainty in Artificial Intelligence, pp. 289–296, 1999.
8. D. Tian, X. Zhao, and Z. Shi, Refining image annotation by integrating PLSA with random walk model, in Advances in Multimedia Modeling, vol. 7732, pp. 13–23, 2013
9. C. Wang, D. M. Blei, and L. Fei-Fei, Simultaneous image classification and annotation, in Proc. IEEE Computer Vision and Pattern Recognition, pp. 1903–1910, 2009.
10. D. M. Blei and J. D. McAuliffe, Supervised topic models, in Proc. Conf. Neural Information Processing Systems, vol. 7, pp. 121-128, 2007.
11. X. Zhang, X. Zhao, Z. Li, J. Xia, R. Jain, and W. Chao, Social image tagging using graph-based reinforcement on multi-type interrelated objects, Signal Processing, vol. 93, no. 8, pp. 2178-2189, 2013.
12. Z. Guan, J. Bu, Q. Mei, C. Chen, and C. Wang, Personalized tag recommendation using graph-based ranking on multi-type interrelated objects, in Proc. 32nd Int. ACM SIGIR Conf. Research and Development in Information Retrieval, pp. 540-547, 2009.
13. J. Tang, R. Hong, S. Yan, T. S. Chua, G. J. Qi, and R. Jain, Image annotation by K-NN-sparse graph-based label propagation over noisily tagged web images, ACM Trans. Intelligent Systems and Technology, vol. 2, no. 2, pp 111–126, 2014.
14. J. Bu, S. Tan, C. Chen, C. Wang, H. Wu, Z. Lijun, and X. He, Music recommendation by unified hypergraph: Combining social media information and music content, in Proc. ACM Conf. Multimedia, pp. 391–400, 2010.
15. K. Pliakos and C. Kotropoulos, Personalized and geo-referenced image recommendation using unified hypergraph learning and group sparsity optimization, in Proc. IEEE Int. Symp. Communications, Control, and Signal Processing, pp. 323–326, 2014
16. K. Pliakos and C. Kotropoulos, Simultaneous image tagging and geo-location prediction within hypergraph ranking framework, in Proc. IEEE Int. Conf. Audio, Speech, and Signal Processing, pp. 6944–6948, 2014.
17. G. H. Golub and C. F. Van Loan, Matrix computations, Johns Hopkins studies in mathematical sciences, 1996.
18. L. Li, and T. Li, News recommendation via hypergraph learning: encapsulation of user behavior and news content, in Proc. ACM Int. Conf. Web Search and Data Mining, pp. 305–314, 2013
19. Y. Huang, Q. Liu, S. Zhang, and D. Metaxas, Image retrieval via probabilistic hypergraph ranking, in Proc. IEEE Int. Conf. Computer Vision Pattern Recognition, pp. 3376-3383, 2010.

20. Z. Yu, S. Tang, Y. Zhang, and J. Shao, Image ranking via attribute boosted hypergraph, in Proc. 13th Pacific-Rim Conf. Advances in Multimedia Information Processing, pp. 779-789, 2012.
21. J. Xu, V. Singh, Z. Guan, and B. S. Manjunath, Unified hypergraph for image ranking in a multimodal context, in Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, pp. 2333-2336, 2012.
22. L. Cao, J. Luo, and A. Gallagher, A worldwide tourist recommendation system based on geotagged web photos, in Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, pp. 2274–2277, 2010.
23. K. Pliakos and C. Kotropoulos, Simultaneous image clustering, classification and annotation for tourism recommendation, in Proc. 8th Hellenic Conf. Artificial Intelligence, pp. 630–640, 2014.
24. N. Bassiou and C. Kotropoulos, On-line PLSA: Batch updating techniques including out of vocabulary words, IEEE Trans. Neural Networks and Learning Systems, 2014.
25. N. Bassiou and C. Kotropoulos, RPLSA: A novel updating scheme for probabilistic latent semantic analysis, Computer, Speech, and Language, vol. 25, no. 4, pp. 741–760, 2011.
26. D. M. Blei, A. Y. Ng, and M. I. Jordan, Latent Dirichlet Allocation, Journal of Machine Learning Research, vol. 3, pp. 993-1022, 2003.
27. C. J. Burges, A tutorial on support vector machines for pattern recognition, Data mining and knowledge discovery, vol. 2, no.2, pp. 121–167, 1998.
28. A. Oliva and A. Torralba, Modeling the shape of the scene: a holistic representation of the spatial envelope, Int. Journal Computer Vision, vol. 42, no. 3, pp. 145–175, 2001.
29. C. Berge and E. Minieka, Graphs and Hypergraphs, vol. 7, North-Holland, Amsterdam, 1973.
30. J. Friedman, T. Hastie, and R. Tibshirani, A note on the group LASSO and a sparse group LASSO, Arxiv preprint arXiv:1001.0736, 2010.
31. M. Yuan and Y. Lin, Model selection and estimation in regression with grouped variables, Journal of the Royal Statistical Society: Series B (Statistical Methodology), vol. 68, no. 1, pp. 49–67, 2006.
32. J. Huang and T. Zhang, The benefit of group sparsity, The Annals of Statistics, vol. 38, no. 4, pp. 1978–2004, 2010
33. A. Beck and M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, SIAM Journal Imaging Sciences, vol. 2, no. 1, pp.183-202, 2009
34. T. Hofmann, Unsupervised learning by probabilistic latent semantic analysis, Machine Learning, vol. 42, no. 12, pp. 177-196, 2001.
35. A. Dempster, N. Laird, and D. Rubin, Maximum likelihood from incomplete data via the EM algorithm (with discussion), Journal of the royal statistical society, Series B, vol. 39, no. 1, pp. 138, 1977.
36. G. Tzortzis and A. Likas, The Global Kernel k-Means Clustering Algorithm, IEEE Trans. Neural Networks, vol. 20, no. 7, pp. 1181–1194, 2009.
37. M. Douze, H. Jgou, H. Sandhawalia, L. Amsaleg, and C. Schmid, Evaluation of gist descriptors for web-scale image search. in Proc. ACM Int. Conf. Image and Video Retrieval, pp. 19–26, 2009.
38. D. G. Lowe, Distinctive image features from scale-invariant keypoints, Int. Journal Computer Vision, vol. 60, no. 2, pp.91–110, 2004.
39. D. Zhou, J. Huang, and B. Schokopf, Learning with hypergraphs: Clustering, classification, and embedding, in Proc. Advanced Neural Information Processing Systems, pp.1601–1608, 2007
40. S. Agarwal, K. Branson, and S. Belongie, Higher order learning with graphs, in Proc.

23rd Int. Conf. Machine Learning, pp. 17–24, 2006.
41. G. Salton, A. Wong, C.S. Yang, A vector space model for automatic indexing, Communications ACM, vol. 18, no. 11, pp. 613–620, 1975.
42. C. Cortes and V. Vapnik, Support-vector networks, Machine learning, vol. 20, no. 3, pp. 273–297, 1995.
43. J. Bioucas-Dias, M. Figueiredo, and J. Oliveira, Adaptive total-variation image deconvolution: A majorization-minimization approach, in Proc. European Signal Processing Conf., pp. 1–4, 2006.