Image Tag Recommendation Based on Novel Tensor Structures and Their Decompositions

Panagiotis Barmpoutis, Constantine Kotropoulos, and Konstantinos Pliakos

Department of Informatics
Aristotle University of Thessaloniki
Thessaloniki, 54124, Greece
Email: panbar@auth.gr, costas@aiia.csd.auth.gr, kpliakos@aiia.csd.auth.gr

***Abstract*—In this paper, we address the problem of image tagging and we propose automatic methods for image tagging, using tensor decompositions. Tensors are a suitable way of mathematically representing multilink relations. Another, complementary structure that captures the aforementioned high-order relations is the hypergraph. More specifically, three different matrices are derived from the hypergraph, namely, the incidence, adjacency, and affinity matrices. The just mentioned matrices are used to create slices of a novel tensor structure, which combines users' and images' relations. Four methods are exploited to decompose the tensor, i.e., the Higher Order Singular Value Decomposition (HOSVD), the Canonical Decomposition/Parallel Factor Analysis (CANDECOMP/ PARAFAC, CP), the Non-negative Tensor Factor Analysis (NTF), and Tucker Decomposition (TD). Experiments conducted on a dataset retrieved from Flickr demonstrate the potential of the proposed approach.**

Keywords—Tensor Decompositions; Image Tagging;

I. INTRODUCTION

Social tagging is an important feature that allows users to annotate items, like images, songs, posts, etc. with keywords [*1*]. In recent years, due to the rapid growth of social media, image tag recommendation (or automatic image annotation) has emerged as a very interesting and hot research area. Indeed, a lot of effort has focused mainly on the image tag recommendation, enabling a quick and economic organization, browsing, and searching of user data. Most social media record user data (i.e., profiles), geographical data, and content-related descriptions of images. However, the assignment of a wide variety of labels to one particular image as well as the assignment of the same label to different images with respect to their visual appearance make image tag recommendation a difficult problem.

Undoubtedly image tagging is a multifaceted problem that is driven by multiple sources of information, such as visual similarity, the correlation between tags and visual features, the correlation between users, as is captured by friendships or group relations, and the user interests. Geo-location is another factor that heavily influences image tagging.

The aforementioned relations cannot be confined to pairs, yielding naturally to either hypergraphs [*2*] or hypermatrices commonly referred to as tensors in the electrical engineering/computer science jargon [*3*]. In the following, we shall use the term tensor to be on par with the existing literature. Tensors are a natural generalization of matrices into higher dimensions and could be exploited in the solution of image tagging problem, because they could: a) exploit the multiple correlations between the data, b) create groups of common characteristics, and c) capture the similarity between features [*4*], [*5*].

More specifically, this paper makes the following contributions:

- Starting with a hypergraph, three types of matrices are derived namely, the incidence, the adjacency, and the affinity ones that are used to define the slices of an appropriate tensor.
- A novel tensor structure is proposed that efficiently captures both user and image relations.
- High-order singular value decomposition (HOSVD) is demonstrated that outperforms other tensor decompositions, such as the Canonical decomposition/ Parallel factor analysis (CANDECOMP/PARAFAC, CP), the non-negative tensor factorization (NTF), or the Tucker decomposition (TD), when it is used to extract a latent semantic model to be exploited for image tagging.

The remainder of this paper is organized as follows. Related work is surveyed in Section II. Section III details the proposed method. Experimental results are discussed in Section IV. Finally, conclusions are drawn in Section V.

II. RELATED WORK

Many researchers have attempted to address the problem of image tagging, focusing mainly on the image content and the 2-way item representations. More specifically, Support Vector Machines [*6*], or artificial neural networks [*7*] were used for image classification. Automatic image annotation thanks to a multiple Bernoulli relevance model capturing the joint probability distribution of the possible annotations and the image feature vectors was proposed in [*8*]. Other techniques have focused on nearest neighbors [*9*], [*10*], structural group sparsity for feature selection to boost the annotation performance. The just described techniques distill the correlations among multiple tags [*11*] or semantic distance functions [*12*]. Another image annotation method based on image search was presented in [*13*]. Detailed surveys on

automatic annotation techniques can be found in [*5*], [*14*] and [*15*]. More recently, a new method for image tagging and geo-location prediction was proposed that was based on hypergraph learning [*2*].

On the other hand, in the field of recommendation systems, many methods have adopted tensors, because of the powerful properties of tensor decompositions. Symeonidis et al. proposed representing the data by 3rd-order tensors on which latent semantic analysis and dimensionality reduction was performed, using HOSVD [*16*]. A pairwise interaction tensor factorization model was proposed in [*17*], which is a special case of the TD model with linear runtime for both learning and prediction. The model was learned with an adaptation of the Bayesian personalized ranking criterion. Existing methods that use information from multi-type interrelated objects may also employ graphs [*18*], [*19*]. Multi-label image tagging within a sparse coding framework was addressed in [*20*].

In this paper, we propose a methodology for automatic image tagging, which aims to exploit the information from a hypergraph to build a tensor, capturing relations between images and users. The proposed methodology extends the work in [*5*] and is described in detail next.

III. METHODOLOGY

The proposed image tag recommendation scheme is shown in Fig. 1. It consists of three main steps: tensor creation, tensor decomposition, and tag recommendation. These steps are described in detail in the following subsections. In the paper, the notation used in [*4*] is adopted.

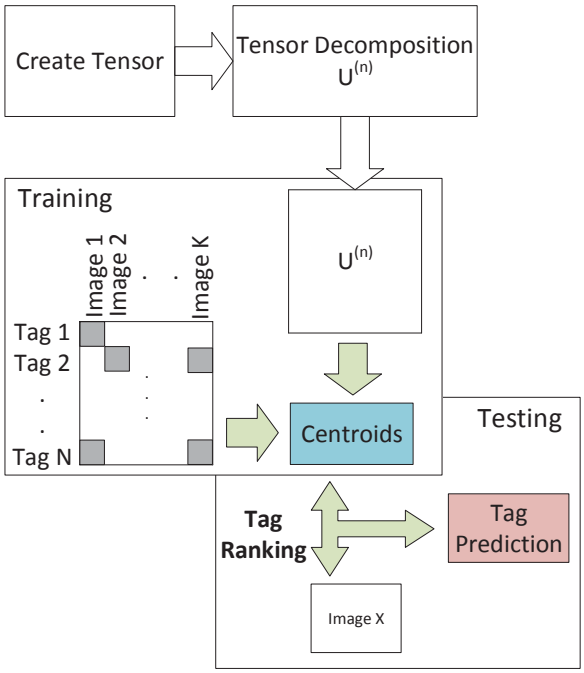

Fig. 1 Methodology of the proposed algorithm.

A. Dataset description and tensor creation

Motivated by work in [*5*], we are interested in creating the three 3rd-order tensors shown in Fig. 2. As can be seen, the slices of these tensors are square matrices.

The user tensor consists of $K_{us-slices} = 5$ slices, namely: a) user relationships induced by friendships, b) user relationships induced by their participation into groups, c) user relationships induced by image uploading, d) user relationships induced by the geo-tags of the uploaded images, and e) user relationships induced by the tags of the uploaded images.

The image tensor consists of $K_{im-slices} = 4$ slices, namely: a) image relationships induced by the users, b) image relationships induced by their geo-tags, c) image relationships induced by their tags, and d) visual image similarities.

The just-described tensors can be combined into one tensor, that shown in the right-hand side of Fig. 2. This tensor has $K_{us-slices \cup im-slices} = 6$ slices, which correspond to the types of the hyperedges of the underlying hypergraph, defined next.

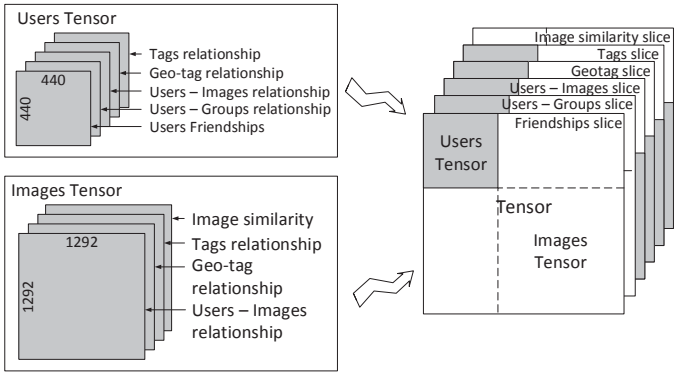

Fig. 2 Tensor representation.

To facilitate the discussion, let us describe the dataset used in the paper. This dataset was collected from Flickr [*21*] and contains both indoor and outdoor images of popular Greek landmarks. More specifically, it consists of $N_{im} = 1292$ images, uploaded by $N_{us} = 440$ users, who used a vocabulary of 2366 tags. The entire Greece was partitioned to 125 cells, which are defined in terms of 125 geo-tags. The users were participated into 1644 groups.

The aforementioned entities correspond to the vertices of a hypergraph. Various hyperedges were identified in the hypergraph, which define the incidence matrix H of the hypergraph, shown in Table 1 [*2*]. As can be seen in Table 1, six different sub-matrices, capturing various relations in the context of image tag recommendation, can be derived.

Table 1. Structure of the hypergraph incidence matrix H and its sub-matrices for image tag recommendation

$E^{(1)}$	$E^{(2)}$	$E^{(3)}$	$E^{(4)}$	$E^{(5)}$	$E^{(6)}$
$UE^{(1)}$	$UE^{(2)}$	$UE^{(3)}$	$UE^{(4)}$	$UE^{(5)}$	0
0	$GrE^{(2)}$	0	0	0	0
0	0	0	$GeoE^{(4)}$	0	0
0	0	0	0	$TaE^{(5)}$	0
0	0	$ImE^{(3)}$	$ImE^{(4)}$	$ImE^{(5)}$	$ImE^{(6)}$

In particular, $E^{(1)}$ has size 440×2276 and captures the pairwise friendship relations between users. $E^{(2)}$ captures the information of users groups. The incidence submatrix $E^{(2)}$ refers to both users and usergroups and has size 2048×1644. $E^{(3)}$ contains a user and an uploaded, representing a user-

image possession relation and has size 1732×1292. $E^{(4)}$ captures a geo-location relation, contains triplets of images, users and geo-locations and has size 1857×125. $E^{(5)}$ contains triplets of images, users and tags. Each hyperedge represents a tagging relation. Size of $E^{(5)}$ is 4098×19127. The weights of the hyperedges $E^{(1)} \dots E^{(5)}$ are set equal to one. $E^{(6)}$ contains pairs of vertices, which represent two images. The weight $w(e_{ij}^{(6)})$ is set as the normalized similarity between images i and j. Details can be found in [*3*]. The incidence submatrix $E^{(6)}$ has size 1292×6460.

Each hyperedge set $E^{(j)}$ drives a relationship that yields a slice of the user and image tensors denoted by $M_j, j = 1,2, \dots ,5$ and $\widetilde{M_j}, j = 3,4, \dots ,6$, respectively.

The first obvious choice for $M_j, j = 1,2, \dots ,5$ and $\widetilde{M_j}, j = 3,4, \dots ,6$ exploits the sub-matrices of the incidence matrix indicated in Table 1. User tensor is formed as follows:

$$\mathrm{M}_1 = \left[\mathrm{UE}^{(1)}\right]\left[\mathrm{UE}^{(1)}\right]^{\mathrm{T}} \quad (1)$$

$$\mathrm{M}_2 = \left[\mathrm{UE}^{(2)}\right]\left[\mathrm{GrE}^{(2)}\right]^{\mathrm{T}}\left[\mathrm{GrE}^{(2)}\right]\left[\mathrm{UE}^{(2)}\right]^{\mathrm{T}} \quad (2)$$

$$\mathrm{M}_3 = \left[\mathrm{UE}^{(3)}\right]\left[\mathrm{ImE}^{(3)}\right]^{\mathrm{T}}\left[\mathrm{ImE}^{(3)}\right]\left[\mathrm{UE}^{(3)}\right]^{\mathrm{T}} \quad (3)$$

$$\mathrm{M}_4 = \left[\mathrm{UE}^{(4)}\right]\left[\mathrm{GeoE}^{(4)}\right]^{\mathrm{T}}\left[\mathrm{GeoE}^{(4)}\right]\left[\mathrm{UE}^{(4)}\right]^{\mathrm{T}} \quad (4)$$

$$\mathrm{M}_5 = \left[\mathrm{UE}^{(5)}\right]\left[\mathrm{TaE}^{(5)}\right]^{\mathrm{T}}\left[\mathrm{TaE}^{(5)}\right]\left[\mathrm{UE}^{(5)}\right]^{\mathrm{T}} \quad (5)$$

And image tensor is formed as follows:

$$\widetilde{\mathrm{M}_3} = \left[\mathrm{ImE}^{(3)}\right]\left[\mathrm{UE}^{(3)}\right]^{\mathrm{T}}\left[\mathrm{UE}^{(3)}\right]\left[\mathrm{ImE}^{(3)}\right]^{\mathrm{T}} \quad (6)$$

$$\widetilde{\mathrm{M}_4} = \left[\mathrm{ImE}^{(4)}\right]\left[\mathrm{GeoE}^{(4)}\right]^{\mathrm{T}}\left[\mathrm{GeoE}^{(4)}\right]\left[\mathrm{ImE}^{(4)}\right]^{\mathrm{T}} \quad (7)$$

$$\widetilde{\mathrm{M}_5} = \left[\mathrm{ImE}^{(5)}\right]\left[\mathrm{TaE}^{(5)}\right]^{\mathrm{T}}\left[\mathrm{TaE}^{(5)}\right]\left[\mathrm{ImE}^{(5)}\right]^{\mathrm{T}} \quad (8)$$

$$\widetilde{\mathrm{M}_6} = \left[\mathrm{ImE}^{(6)}\right]\left[\mathrm{ImE}^{(6)}\right]^{\mathrm{T}} \quad (9)$$

Let $|V| = 5687$ be the total number of vertices and $|E| = 30924$ be the total number of hyperedges. The second choice is based on the adjacency matrix of the hypergraph, $A = HWH^T - D_v$ where W is the $|E| \times |E|$ diagonal matrix, $W = diag(I, I, I, I, I, W_i)$, with I denoting identity matrices of compatible dimensions and $D_v = diag(Hw)$ with $w \in \mathbb{R}^{|E| \times 1}$ made up of the elements in the main diagonal of W. A is a symmetric matrix. Its upper triangular part can be partitioned as follows:

$$A = \begin{bmatrix} A_{11} & A_{12} & A_{13} & A_{14} & A_{15} & A_{16} \\ & A_{22} & A_{23} & A_{24} & A_{25} & A_{26} \\ & & A_{33} & A_{34} & A_{35} & A_{36} \\ & & & A_{44} & A_{45} & A_{46} \\ & & & & A_{55} & A_{56} \\ & & & & & A_{66} \end{bmatrix} \quad (10)$$

Then, $M_j = A_{1j}{A_{1j}}^T, \ j = 1,2, \dots ,5$ and $\widetilde{M_j} = {A_{j6}}^T A_{j6}, \ j = 4,5, \dots ,6$.

The 3rd choice employs the Laplacian matrix of the hypergraph, $L = I - {D_v}^{-1/2} H W {D_e}^{-1} H^T {D_v}^{-1/2} = I - \Theta$, where $D_e = diag(1^T H)$ is the $|E| \times |E|$ hyperedge degree matrix. Θ admits the same structure with the adjacency matrix. Accordingly, $M_j = \Theta_{1j}{\Theta_{1j}}^T, \ j = 1,2, \dots ,5$ and $\widetilde{M_j} = {\Theta_{j6}}^T \Theta_{j6}, \ j = 4,5, \dots ,6$.

B. Tensor decompositions

Let $\boldsymbol{M} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$, denote the tensor created. Four decompositions were applied to $\boldsymbol{M}$, that are described next.

The HOSVD of $\boldsymbol{M}$ allows us to estimate its n- ranks [*22*], [*23*], yielding:

$$\boldsymbol{M} = \boldsymbol{S} \times_1 U^{(1)} \times_2 U^{(2)} \times_3 U^{(3)} \quad (11)$$

where $\boldsymbol{M}, \boldsymbol{S} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ and $U^{(i)} \in \mathbb{R}^{I_i \times I_i}, i = 1,2,3$ are orthogonal matrices. Clearly, I_1, I_2 equal N_{us}, N_{im}, or N_{us+im} and $I_3 = 5, 4, or\ 6$, for the three tensors created, respectively. The $tensor\ \boldsymbol{S}$ can be found by:

$$\boldsymbol{S} = \boldsymbol{M} \times_1 {U^{(1)}}^T \times_2 {U^{(2)}}^T \dots \times_N {U^{(N)}}^T \quad (12)$$

The CP decomposes $\boldsymbol{M}$ into a sum of rank-one tensors, where each rank-one tensor is an outer product of three vectors [*24*], [*25*], [*26*]:

$$\boldsymbol{M} = \boldsymbol{\mathcal{G}} \times_1 U^{(1)} \times_2 U^{(2)} \times_3 U^{(3)} = \sum_{j=1}^{R} u_j^{(1)} \otimes u_j^{(2)} \otimes u_j^{(3)} \quad (13)$$

where $U^{(1)} \in \mathbb{R}^{I_1 \times R}$, $U^{(2)} \in \mathbb{R}^{I_2 \times R}$ and $U^{(3)} \in \mathbb{R}^{I_3 \times R}$.
A common approach to fitting a CP decomposition is to use an alternating least-squares (ALS) algorithm. The algorithm solves the optimization problem:

$$\min_{u_j^{(l)}} \left\| \boldsymbol{M}_{i_1 i_2 i_3} - \sum_{j=1}^{R} u_j^{(1)}(i_1) u_j^{(2)}(i_2) u_j^{(3)}(i_3) \right\|_F^2 \quad (14)$$

for $j = 1,2, \dots ,R$ and $l = 1,2,3$, where $\| \ \|_F$ denotes the Frobenious norm of a tensor.

The NTF is the natural generalization of non-negative matrix factorization (NMF) to tensors [*27*], [*28*]. It decomposes:

$$\boldsymbol{M} = \sum_{j=1}^{k} u_1^{(1)} \otimes u_2^{(2)} \otimes u_N^{(3)} \quad (15)$$

where $u_j^i \in \mathbb{R}^{I_i} \geq 0$. The NTF model solves the minimization problem:

$$\min_{\boldsymbol{u}_i^{(j)}} \frac{1}{2} \left\| \boldsymbol{M} - \sum_{j=1}^{k} u_1^j \otimes u_2^j \otimes \dots \otimes u_N^j \right\|_F^2 \quad (16)$$

subject to the constraint that $\mathrm{u}_\iota^{\mathrm{j}} \geq 0$.

The TD is a higher order generalization of the principal component analysis. It differs from the CP decomposition, since the factorization provides a core tensor multiplied by a matrix along each dimension and the factor matrices have different dimensions [*29*], [*30*]. The Tucker decomposition of tensor $\boldsymbol{M}$ is given by:

$$\boldsymbol{M}_{\boldsymbol{i,j,k}}^{\boldsymbol{TUCKER}} = \sum_{r_1=1}^{R} \sum_{r_2=1}^{R} \sum_{r_3=1}^{R} g_{r_1 r_2 r_3} U_{i\, r_1}^{(1)} U_{j\, r_2}^{(2)} U_{k\, r_3}^{(3)} \quad (17)$$

where $\boldsymbol{M} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ and $\boldsymbol{\mathcal{G}} \in \mathbb{R}^{R \times R \times R}$. $\boldsymbol{\mathcal{G}}$ is the core tensor and determines the connections between the decomposed matrices:

$$\mathcal{G} = \boldsymbol{M} \times_1 U^{(1)} \times_2 U^{(2)} \times_3 U^{(3)}, \qquad (18)$$

where $U^{(1)} \in \mathbb{R}^{I_1 \times R}$, $U^{(2)} \in \mathbb{R}^{I_2 \times R}$ and $U^{(3)} \in \mathbb{R}^{I_3 \times R}$.

C. Tag recommendation

As we can see in Fig. 1, methodology includes training and testing of the algorithm. A training tensor is created by preserving 70% of the tagging relationships across all images. Images whose tags were omitted on purpose are referred to as test images next. During the training procedure, for each tag of interest, we detect all images that contain this tag. Subsequently, the centroids $g_{U^{(1)}} \in \mathbb{R}^R$ and $g_{U^{(2)}} \in \mathbb{R}^R$ are computed by using the rows of the matrices $U^{(1)}$ and $U^{(2)}$, which are associated to the training images containing this tag and averaging columnwise. In the last step, the similarity between the test image and the centroids associated to the tag under discussion is given by the corresponding element of the score vector [5] $s = \frac{1}{2}U^{(1)}g_{U^{(1)}} + \frac{1}{2}U^{(2)}g_{U^{(2)}}$, which is parameterized by the tags. By ranking the similarities across all tags, we propose the top ranking tags to annotate any test image.

IV. Experimental Results

The assessment of the proposed image tagging methodoology is detailed next. Image tagging is judged with respect of the recall-precision curves. Precision is defined as the number of correctly recommended tags divided by the number of total recommended tags. Recall is defined as the number of correctly recommended tags divided by the number of total tags the user has actually set.

The goal of the assessment is three-fold. First, the various tensor decompositions are assessed with respect to the recall-precision curves for image tagging. The decompositions are applied to the three tensors described in Section 2. Secondly, the most appropriate methods to compute the tensor slices is identified within the image tagging framework. Finally, we assess the impact of the user and image tensors separately.

A. Image tagging using different tensor decompositions

For CP decomposition, $R = 30$, so the factorized matrices have dimensions $N \times 30$, where $N = N_{us} + N_{im}$. The recall precision curves for the various decompositions are plotted in Fig. 3. It is seen that the HOSVD outperform the other decompositions, when it is applied to the tensor whose slices are based on hypergraph incidence sub-matrices.

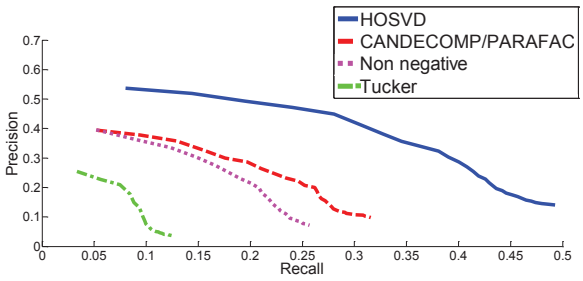

Fig. 3. Averaged Recall-Precision curves for different decompositions applied to the tensor whose slices are based on the incidence matrices.

HOSVD is found to outperform the other decompositions, when the tensors have slices that are created by exploiting the adjacency matrices (Fig. 4) or the affinity matrices (Fig. 5).

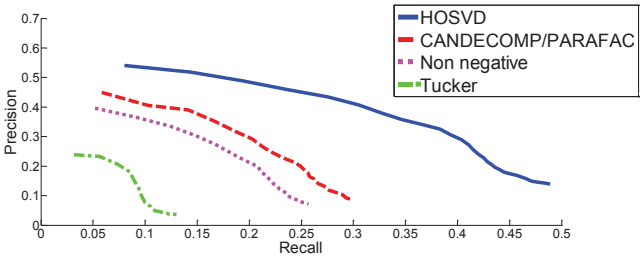

Fig. 4. Averaged Recall-Precision curves for different decompositions applied to the tensor whose slices are based on the adjacency matrices.

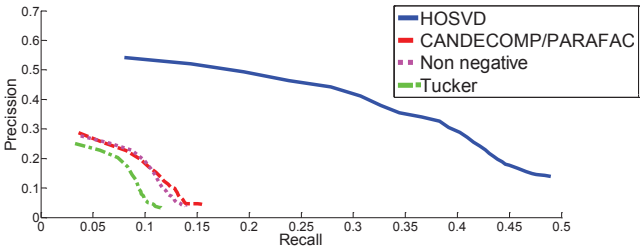

Fig. 5. Averaged Recall-Precision curves for different decompositions applied to the tensor whose slices are based on the affinity matrices.

B. Tag recommendation in user tensor and image tensor using different types of matrices

The recall-precision curves, when the HOSVD is applied to tensors whose slices are based on the incidence, adjacency, and affinity matrices, are overlaid in Fig. 6. The differences are marginal.

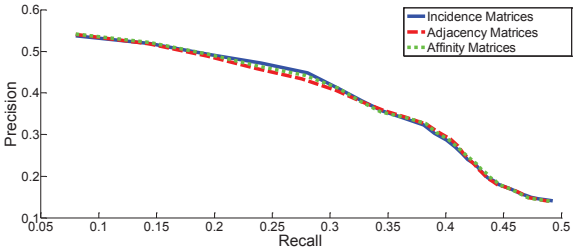

Fig. 6. Averaged Recall-Precision curves applying the HOSVD to tensors based on incidence, adjacency, and affinity matrices.

Next, we assess image tagging, when the HOSVD is applied to the user tensor and the image tensor, separately. These tensors are built by slices based on either the incidence, the adjacency or the affinity matrices.

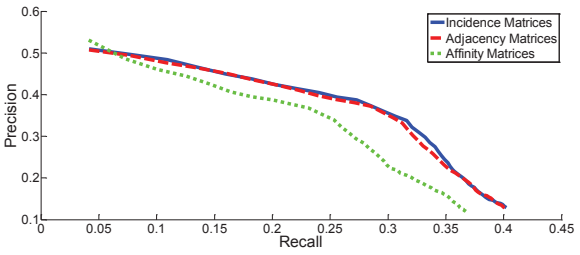

Fig. 7. Averaged Recall-Precision curves applying the HOSVD to user tensors based on incidence, adjacency, and affinity matrices.

The averaged recall-precision curves are plotted in Fig. 7 and Fig. 8, respectively. In both cases, when the tensors are built using the incidence matrices, the performance is the best.

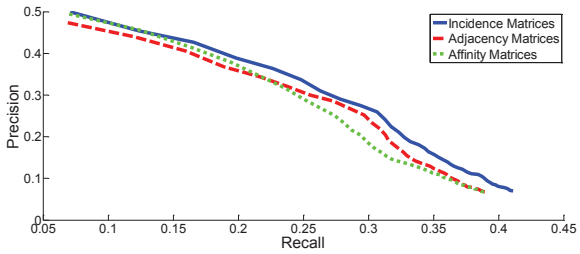

Fig. 8. Averaged Recall-Precision curves applying the HOSVD to image tensors based on incidence, adjacency, and affinity matrices.

C. Image tagging using different tensors

In Fig. 9, the top performing curves in Fig. 6, Fig. 7 and Fig. 8 are overlaid. It is seen that the full tensor, which combines both the user and the image tensors yields the best results.

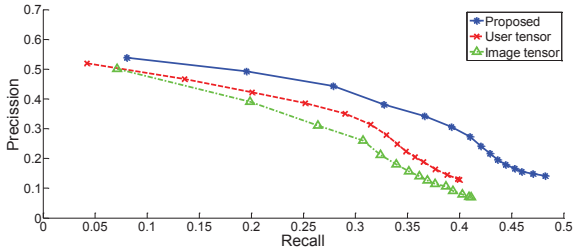

Fig. 9. Averaged Recall-Precision curves of user tensor, image tensor, and the full tensor, whose slices are based on the incidence matrices, and decomposed by the HOSVD.

V. Conclusion and Future Work

In this paper, we have proposed a new tensor structure that grasps both user and image relations. It has been demonstrated that when the HOSVD is applied to this tensor, the best averaged recall-precision curve is obtained for image tagging. The most descriptive tensor has been the one whose slices are based on the hypergraph incidence sub-matrices. The proposed method can accommodate tagging for any kind of multimedia. Future research may be focused on applying robust low-rank decomposition to the full tensor.

Acknowledgment

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: SYN-10-1730-ATLAS.

VI. References

[1] S. Golder and B. Huberman, “The structure of collaborative tagging systems”, in *Proc. Int. Conf.* RR, pp. 1-8, 2005.

[2] K. Pliakos and C. Kotropoulos, "Simultaneous image tagging and geo-location prediction wthin hypergraph ranking framework," in *Proc. IEEE Int. Conf. Accoustics, Speech and Signal Processing*, pp. 6944-6948, Florence, 2014.

[3] C. Kotropoulos, "Multimedia social search based on hypergraph learning," in I. Pitas (Eds), *Graph-Based Social Media Analysis*, pp. 277-349, 2015.

[4] T. G. Kolda and B. W. Bader, "Tensor decomposition and applications," *SIAM Rev.,* vol. 51, pp. 455-500, 2009.

[5] D. M. Dunlavy, T. G. Kolda, and W. P. Kegelmeyer, "Multilinear algebra for analyzing data with multiple linkages," in *J.* Kepner *and J.* Gilbert (Eds.) *Graph Algorithms in the Language of Linear Algebra, pp. 85-114, SIAM, 2011.*

[6] X. Qi and Y. Han, "Incorporating multiple SVMs for automatic image annotation," *Pattern Recognition*, vol. 40, no. 2, pp. 728-741, 2007.

[7] S. Kim, S. Park, and M. Kim, "Image classification into object/non-object classes," in *Proc. Int. Conf. Image and Video Retrieval*, vol. LNCS 3115, pp. 393-400, 2004.

[8] S. Feng, R. Manmatha, and V. Lavrenko, "Multiple Bernoulli relevance models for image and video annotation," in *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, vol. II, pp.1002-1009, 2004.

[9] R. Liu, Y. Wang, H. Yu, and S. Naoi, "A renewed image annotation baseline by image embedding and tag correlation," in *Proc. Int. Conf. Pattern Recognition*, pp. 3216-3219, 2012.

[10] X. Li, C. Snoek, and M. Worring, "Learning social tag relevance by neighbor voting," *IEEE Trans. Multimedia*, vol. 11, no. 7, pp. 1310-1322, 2009.

[11] Y. Han, "Multi-label boosting for image annotation by structural grouping sparsity," in *Proc. ACM Multimedia*, *Firenze, Italy,* pp. 15-24, 2010.

[12] T. Mei, Y. Wang, X. Hua, S. Gong, and S. Li, "Coherent image annotation by learning semantic distance," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, pp. 1-8, 2008.

[13] X. Wang, L. Zhang, X. Li, and W. Ma, "Annotating images by mining image search results," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 11, pp. 1919-1932, 2008.

[14] D. Zhang, M. Islam, and G. Lu, "A review on automatic image annotation techniques," *Pattern Recognition*, vol. 45, no. 1, pp. 346-362, 2012.

[15] A. Milicevic, A. Nanopoulos, and M. Ivanovic, "Social tagging in recommender systems: A survey of the state-of-the-art and possible extensions," *Artificial Intelligence*, vol. 33, no. 3, pp. 187-209, 2010.

[16] P. Symeonidis, A. Nanopoulos, and Y. Manolopoulos, "Tag recommendations based on tensor dimensionality reduction," in Proc. *ACM Recommender Systems*, pp. 43-50, 2008.

[17] S. Rendle and L. Schmidt-Thieme, "Pairwise interaction

tensor factorization for personalized tag recommendation," in Proc. *Web Search and Data Mining (WSDM)*, pp. 81-90, 2010.

[18] X. Zhang, X.Zhao, Z. Li, J. Xia, R. Jain, and W.Chao, "Social image tagging using graph-based reinforcment on multi-type interrelated objects," *Signal Processing*, vol. 93, no. 8, pp. 2178-2189, 2013.

[19] 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 ACM SIGIR Int. Conf. Research and Development in Information Retrieval*, pp. 540-547, 2009.

[20] C. Wang, S. Yan, and H. Zhang, "Multi-label sparse coding for automatic image annotation," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, Florida, USA, 2009, pp. 1643-1650.

[21] http://www.flickr.com/.

[22] G. Bergqvist and E. G. Larsson, "Higher order SVD: Theory and algorithms," IEEE Signal Processing Magazine, vol. 27, no. 3, pp. 151-154, May 2010.

[23] L. De Lathauwer, B. De Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM Journal Matrix Analysis*, vol. 21, no. 4, pp. 1253-1278, Apr. 2000.

[24] J. D. Carroll and J. J. Chang, "Analysis of individual differences in multidimensional scaling via an *N*-way generalization of 'Eckart-Young' decomposition," *Psychometrika*, vol. 35, pp. 283-319, 1970.

[25] R. A. Harshman, "Foundations of the PARAFAC procedure: Models and conditions for an 'explanatory' multi-modal factor analysis," *UCLA Working Papers in Phonetics*, vol. 16, pp. 1-84, 1970.

[26] R. Bro, "PARAFAC: Tutorial and applications," *Chemometrics and Intelligent Laboratory Systems*, vol. 38, pp. 149-171, 1997.

[27] P. O. Hoyer, "Non-negative matrix factorization with sparseness constraints," *Journal of Machine Learning Research*, vol. 5, pp. 1457-1469, 2004.

[28] M. Heiler and C. Schnorr, "Controlling Sparseness in Non-negative Tensor Factorization," *in Lecture Notes in Computer Science*, vol. 3951, pp. 1457-1469, 2006.

[29] L. R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, pp. 279-311, 1966.

[30] L. R. Tucker, R. F. Koopman, and R. L. Linn, "Evaluation of factor analytic research procedures by means of simulated correlation matrices," *Psychometrika*, vol. 34, pp. 421-459, 1969.