PERSONALIZED AND GEO-REFERENCED IMAGE RECOMMENDATION USING UNIFIED HYPERGRAPH LEARNING AND GROUP SPARSITY OPTIMIZATION

Konstantinos Pliakos, Constantine Kotropoulos

Department of Informatics, Aristotle University of Thessaloniki
Box 451, Thessaloniki, 54124, Greece
Email: {kpliakos, costas}@aiia.csd.auth.gr

ABSTRACT

The rapid development of social media has led to a surge of interest in multimedia recommendation. Several recommender systems have been developed, but achieving a satisfactory efficiency or accuracy still remains an open problem. In this paper, a novel multi-reference image recommendation system is proposed based on a unified hypergraph. Relevant images from a large pool are recommended to a reference user or a reference geo-location. In addition to that, the hypergraph ranking problem is enhanced by enforcing group sparsity constraints. By adjusting the different weights associated to the object groups, we control each object group effect in the recommendation process. Experiments on a dataset crawled from $Flickr$ demonstrate the merits of the proposed method.

***Index Terms*—** Image Retrieval, Recommender systems, Hypergraph, Group Sparsity Optimization

1. INTRODUCTION

Nowadays, we witness a growing research interest in multimedia recommendation. Various photo sharing websites like $Flickr$[1] or $PicasaWebAlbum$[2] have become popular, augmenting content-based image recommendation with social media information. Users of these sites, not only upload photos, but also create photo galleries, join into user communities, establish friendships, and insert tags and geo-tags indicating the geo-location of a specific photo. These additional sources of information are proved very useful, because ranking based solely on the similarity of image visual features usually yields unsatisfactory results due to the existing semantic gap. Clearly, the overwhelming number of uploaded images, that is growing exponentially, makes the need for efficient recommender systems indisputable.

During the past decade, many works were focused on image retrieval based on text and content [1]. In [2], retrieval was based on matching low-level features, then building clusters and finally finding the most similar images by comparing intra-cluster similarity distances. A randomized data mining method was proposed in [3], in order to find clusters of images with spatial overlap using the min-hash algorithm. By selecting an image region containing a specific object, James Philbin *et al.* [4] proposed a query-based method whose performance was enhanced by adding a spatial verification stage. Moreover, many works exploited the geo-location information contained in the metadata. In [5], images were first grouped geographically and then visually forming an image clustering scheme. Kennedy *et al.* [6] demonstrated image search results for landmarks using both context and content-based tools.

Here, a novel approach to image recommendation problem is proposed, using unified hypergraphs [7–12] and group sparsity. Hypergraphs consist of a set of vertices made by concatenating different kind of objects (images, users, social groups, geo-tags, tags) and hyperedges linking these vertices. This way, the existing multi-link relations between the vertices are represented. By fully exploiting the information provided by social media, hypergraphs are demonstrated to outperform existing methods that are based only on visual image features [2, 9] or graph-based methods, which model only the pairwise relations between the images. Motivated by these findings, we demonstrate that the inclusion of user friendship and user group relations on the top of image-related metadata and image similarity increases the efficiency of personalized image recommendation. The hypergraph ranking is also enhanced by enforcing group sparsity constraints. This way, the set of objects is split into different object groups (images, users, social groups, tags, geo-tags) and each object group effect in the recommendation process is controlled separately, by assigning them weights.

2. GROUP SPARSE REGULARIZATION FOR RANKING ON A HYPERGRAPH

In the following, $|\cdot|$ denotes set cardinality, $\|.\|_2$ the ℓ_2 norm of a vector and $\mathbf{I}$ is the identity matrix of compatible dimensions. A hypergraph is a generalization of a graph whose edges connect more than two vertices. Let $G(V, E, w)$ denote a hypergraph with set of vertices V and set of hyperedges E to which a weight function $w : E \rightarrow \mathbb{R}$ is assigned. The vertex set V is made by concatenating sets of objects of different type (images, users, social groups, geo-tags, tags). These vertices and hyperedges form a $|V| \times |E|$ incidence matrix with elements $H(v, e) = 1$ if $v \in e$ and 0 otherwise. Based on $\mathbf{H}$, the vertex and hyperedge degrees are defined as $\delta(v) = \sum_{e \in E} w(e)H(v, e)$ and $\delta(e) = \sum_{v \in V} H(v, e)$, accordingly. 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 $\mathbf{A} = \mathbf{D}_u^{-1/2}\mathbf{H}\mathbf{W}\mathbf{D}_e^{-1}\mathbf{H}^T\mathbf{D}_u^{-1/2}$, then $\mathbf{L} = \mathbf{I} - \mathbf{A}$ is the positive semi-definite Laplacian matrix of the hypergraph. The elements of $\mathbf{A}$, $A(u, v)$, indicate the relatedness between the objects u and v. In order to compute a real valued ranking vector $\mathbf{f} \in \mathbb{R}^{|V|}$, we minimize $\Omega(\mathbf{f}) = \frac{1}{2}\mathbf{f}^T\mathbf{L}\mathbf{f}$, requiring all vertices with the same value in the ranking vector $\mathbf{f}$ to be strongly connected [12]. The aforementioned optimization problem was extended by including the ℓ_2 regularization norm between the ranking vector $\mathbf{f}$ and the query vec-

[1] http://www.flickr.com
[2] http://picasaweb.google.com

tor $\mathbf{y} \in \mathbb{R}^{|V|}$ in music recommendation [11]. The function to be minimized is expressed as

$$\tilde{Q}(\mathbf{f}) = \Omega(\mathbf{f}) + \vartheta \, ||\mathbf{f} - \mathbf{y}||_2^2 \tag{1}$$

where ϑ is a regularizing parameter. The ranking vector $\mathbf{f}^* = \arg\min_{\mathbf{f}} \tilde{Q}(\mathbf{f})$ is [11]:

$$\mathbf{f}^* = \frac{\vartheta}{1+\vartheta} \left(\mathbf{I} - \frac{1}{1+\vartheta}\mathbf{A}\right)^{-1} \mathbf{y}. \tag{2}$$

Hereafter, each vertex subset is referred to *object group* to avoid confusion with social groups. Indisputably, each object group contributes differently to the ranking procedure. According to [13], a Group Lasso regularizing term is more appropriate than the ℓ_2 norm in this kind of problems. The hypergraph vertices are split into S non-overlapping object groups (images, users, social groups, geo-tags, tags) and different weights γ_s, $s = 1, 2, \ldots, S$ are assigned to each object group, yielding the following objective function to be minimized:

$$Q(\mathbf{f}) = \Omega(\mathbf{f}) + \vartheta \sum_{s=1}^{S} \sqrt{\gamma_s \, (\mathbf{f}-\mathbf{y})^T \mathbf{K}_s (\mathbf{f}-\mathbf{y})}. \tag{3}$$

In (3), ϑ is also a regularizing parameter and $\mathbf{K}_s$ is the $|V| \times |V|$ diagonal matrix with elements equal to 1 for the vertices, which belong to the s-th object group. The minimization problem can be expressed as:

$$\mathbf{f}^* = \underset{\mathbf{f}}{\operatorname{argmin}} \, Q(\mathbf{f}). \tag{4}$$

Let $\mathbf{x} = \mathbf{f} - \mathbf{y}$. The auxiliary variable $\mathbf{z} = \mathbf{x}$ is introduced and (4) is rewritten as:

$$\underset{\mathbf{x}}{\operatorname{argmin}} \frac{1}{2}(\mathbf{x}+\mathbf{y})^T \mathbf{L} (\mathbf{x}+\mathbf{y}) + \vartheta \sum_{s=1}^{S} \sqrt{\gamma_s \, \mathbf{z}^T \mathbf{K}_s \mathbf{z}}$$
$$\text{s.t. } \mathbf{z} = \mathbf{x}. \tag{5}$$

The solution of (5) can be obtained by minimizing the augmented Lagrangian function

$$\mathcal{L}(\mathbf{x}, \mathbf{z}, \boldsymbol{\lambda}) = \frac{1}{2}(\mathbf{x}+\mathbf{y})^T \mathbf{L}(\mathbf{x}+\mathbf{y}) + \vartheta \sum_{s=1}^{S} \sqrt{\gamma_s \mathbf{z}^T \mathbf{K}_s \mathbf{z}}$$
$$+ \boldsymbol{\lambda}^T(\mathbf{z}-\mathbf{x}) + \frac{\mu}{2}||\mathbf{z}-\mathbf{x}||_2^2, \tag{6}$$

where $\boldsymbol{\lambda}$ is the vector of the Lagrange multipliers, which is updated at each iteration and μ is a parameter regularizing the violation of the constraint $\mathbf{x} = \mathbf{z}$. (6) can be solved by the Alternating Directions Method [14], as shown in Algorithm 1. Solving for $\mathbf{x}^{t+1}$ in line 3 yields

$$\mathbf{x}^{t+1} = (\mathbf{L} + \mu^t \mathbf{I})^{-1}(\boldsymbol{\lambda}^t + \mu^t \mathbf{z}^t - \mathbf{L}\mathbf{y}). \tag{7}$$

The minimization problem described in line 4 of Algorithm 1 can be expressed as

$$\min_{\mathbf{z}} \mu^t \left\{ \frac{\vartheta}{\mu^t} \sum_{s=1}^{S} \sqrt{\gamma_s} \sqrt{\mathbf{z}^T \mathbf{K}_s \mathbf{z}} + \frac{1}{2}||\mathbf{z} - (\mathbf{x}^{t+1} - \frac{1}{\mu^t}\boldsymbol{\lambda}^t)||_2^2 \right\}. \tag{8}$$

By applying the soft-thresholding operator [15], we obtain

$$z_j = \frac{r_j}{||\mathbf{r}_s||_2} \max\left(0, ||\mathbf{r}_s||_2 - \vartheta \mu^t \frac{1}{\sqrt{\gamma_s}}\right) \tag{9}$$

where $r_j = x_j^{t+1} - \frac{1}{\mu^t}\lambda_j^t$, s is the object group where the j-th element belongs, and $\mathbf{r}_s$ indicates the segment of $\mathbf{r}$ associated to the s-th object group.

Algorithm 1 Alternating Directions Method

1: Given $\mathbf{x}^t$,$\mathbf{z}^t$ and $\boldsymbol{\lambda}^t$.
2: Set tolerance ϵ and initialize μ.
3: $\mathbf{x}^{t+1} \leftarrow \underset{\mathbf{x}}{\operatorname{argmin}} \mathcal{L}(\mathbf{x}, \mathbf{z}^t, \boldsymbol{\lambda}^t)$
4: $\mathbf{z}^{t+1} \leftarrow \underset{\mathbf{z}}{\operatorname{argmin}} \mathcal{L}(\mathbf{x}^{t+1}, \mathbf{z}, \boldsymbol{\lambda}^t)$
5: **if** $\|\mathbf{z} - \mathbf{x}\|_2^2 > \epsilon$ **then**
6: $\quad \boldsymbol{\lambda}^{t+1} \leftarrow \boldsymbol{\lambda}^t + \mu^t(\mathbf{z}^{t+1} - \mathbf{x}^{t+1})$
7: $\quad \mu^{t+1} = \min(1.1\mu^t, 10^6)$
8: **else**
9: $\quad$ return $\mathbf{x}^{t+1}$, $\mathbf{z}^{t+1}$.
10: $\quad \mathbf{f} = \mathbf{x}^{t+1} + \mathbf{y}$
11: **end if**

3. DATASET DESCRIPTION AND HYPERGRAPH CONSTRUCTION

3.1. Dataset description

For evaluation purposes, an image dataset was collected from $Flickr$. It contains both indoor and outdoor medium size photos of popular Greek landmarks, various city scenes and landscapes. Using $FlickrApi$ [3], a large set of ”geotagged” images was downloaded along with valuable information related to them (id, title, owner, latitude, longitude, tags, image views). Then, the dataset was filtered based on image views (times that the specific image has been seen in $Flickr$) and owner's uploading statistics. At this point, it was assumed that images with many views normally depict important content and owners (users) with many uploaded images are active ones, possessing many social relations (friends, social groups). The owners of these images were the users in the dataset. Then, corresponding social information (friends, social groups) was crawled and only the groups that had at least 5 owners from the dataset as members were kept. The specific cardinalities are summarized in Table 1.

In order to form a proper set of tags, all characters were converted to lower case, unreadable symbols and redundant information were removed. Next, a dictionary of unique words was generated along with their frequencies. Then, terms with frequency 1 or 2 were deemed as trash and were removed from the set of tags and the vocabulary. Finally, spelling mistakes were corrected and any morphological variations merged using the Edit Distance [16].

Geo-tags were clustered into 125 different clusters using hierarchical clustering after computing pairwise distances with “Haversine formula”[4].

Table 1. Dataset objects, notations, and counts.

Object	Notation	Count
Images	Im	1292
Users	U	440
User Groups	Gr	1644
Geo-tags	Geo	125
Tags	Ta	2366

[3] http://www.flickr.com/services/api/
[4] http://www.movable-type.co.uk/scripts/latlong.html

3.2. Hypergraph construction

The vertex set is defined as $V = Im \cup U \cup Gr \cup Geo \cup Ta$. The hypergraph, $\mathbf{H}$ is formed concatenating the 6 hyperedge sets as reflected in Table 2 and it has a size of 5867×30924 elements. In the following, the weights of the hyperedge sets $E^{(1)}$–$E^{(5)}$ are set equal to one. The dataset has captured 2276 friendship relations and 19127 tagging ones.

$E^{(1)}$ represents a pairwise friendship relation between users. The incidence matrix of the hypergraph $UE^{(1)}$ has a size of 440×2276 elements.

$E^{(2)}$ represents a user group and it contains all the vertices of the corresponding users as well as the ones corresponding to the user group. The incidence matrix of the hypergraph $UE^{(2)} - GrE^{(2)}$ has a size of 2084×1644 elements.

$E^{(3)}$ contains a user and an uploaded image, representing a user-image possession relation. The incidence matrix of the hypergraph $UE^{(3)} - ImE^{(3)}$ has a size of 1732×1292 elements. Evident as each image has only one owner.

$E^{(4)}$ captures a geo-location relation. This hyperedge set contains triplets of Im, U and Geo. The incidence matrix of the hypergraph $ImE^{(4)} - UE^{(4)} - GeoE^{(4)}$ has a size of 1857×125 elements.

$E^{(5)}$ also contains triplets, Im, U and Ta. Each hyperedge represents a tagging relation. The $ImE^{(5)} - UE^{(5)} - TaE^{(5)}$ has a size of 4098×19127 elements.

$E^{(6)}$ contains pairs of vertices, which represent two images, with its weight $w(e_{ij}^{(6)})$ set as the similarity between images i and j, normalized as follows to eliminate the bias, $w(e_{ij}^{(6)})' = \frac{w(e_{ij}^{(6)})}{max(w(e_i^{(6)}))}$. In order to form this part of the hypergraph, both global and local features were used. Firstly, the 100 nearest neighbors to each image were identified using GIST descriptors [17] and they were reduced to the 5 most similar images to the reference image, by using scale-invariant feature transform (SIFT) [18]. The size of $ImE^{(6)}$ is 1292×6460.

The query vector $\mathbf{y}$ is initialized by setting the entry corresponding to the target user u to 1 and all others objects Im, Gr, Geo, Ta connected to the specific user to $A(u, v)$. Similarly, in geo-referenced image recommendation, the entry corresponding to the referenced geo-location (geo-cluster) geo is set to 1, users u and images im connected to this geo-location are set to $A(geo, u)$ and $A(geo, im)$ respectively. It is underlined, that $A(i, j)$ is the element of $\mathbf{A}$ which correspond to the objects i and j and it is a relatedness measure of the 2 connected objects. The query vector $\mathbf{y}$ has a length of 5867 elements.

The ranking vector $\mathbf{f}^*$ is derived by solving either (2) or (4), after setting the values of the query vector $\mathbf{y}$, the regularization parameter ϑ, and the group of objects weights γ_s in the case of (4). It has the same size and structure as $\mathbf{y}$. The values corresponding to images are used for personalized or geo-referenced image recommendation with the top ranked images being recommended to the user or recommended for the referenced geo-location (geographical cluster).

3.3. Experiments

The averaged Recall-Precision and F_1 measure are used as figures of merit. Precision is defined as the number of correctly recommended images divided by the number of all recommended images. Recall is defined as the number of correctly recommended images divided by the number of all images the user has actually uploaded. The F_1 measure is the weighted harmonic mean of precision and recall, which measures the effectiveness of recommendation when treating precision and recall as equally important. We refer to the ranking obtained by (2) and (4) as Image Recommendation on Hypergraph (IRH) and Query Group Sparse Optimization (QGSO), respectively.

Table 2. The structure of the hypergraph incidence matrix $\mathbf{H}$ and its sub-matrices.

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

For evaluation purposes, a test set containing the 25% of the images and a training set containing the remaining 75% are defined. The test set is not included in the training procedure. The results of the personalized image recommendation (IRH) are demonstrated in Fig. 1 and the ones of the geo-referenced recommendation (Geo-IRH) in Fig. 2. In Fig. 1, the averaged Recall-Precision curves are plotted by averaging the Recall-Precision curves over 110 users with at least 3 uploaded images. To calculate the recall and precision, the 10 top ranked images are being recommended to the user. Similarly, in the Fig. 2, Recall-Precision curves for 60 geo-locations having at least 3 associations with images are averaged. The 10 top ranked images are being recommended as relevant to the referenced geo-location.

In Fig. 1, IRH1 corresponds to a reduced hypergraph defined as $\tilde{H} = E^{(3)} \cup E^{(4)} \cup E^{(5)} \cup E^{(6)}$ without any information about user groups and friendships. The IRH exploits the complete hypergraph, yielding better results. It is clearly seen that the information provided by social media improves the quality of image recommendation. By enforcing group sparsity in the ranking problem, the performance is further improved. The weights for the 5 different object groups (images, users, user groups, geo-tags and tags) were set to 0.9, 0.9, 0.6, 0.2, 0.8 respectively. This choice was made empirically. Preliminary results for the Geo-IRH are promising, as it is shown in Fig. 2.

In Table 3, the averaged F_1 measure is listed for IRH, IRH1, Geo-IRH and QGSO, corresponding to 5 different ranking positions. It is evident that the IRH, Geo-IRH and QGSO have efficient results. The IRH and the QGSO yield a higher F_1 measure than IRH1 and it is also clearly indicated, that the QGSO outperforms the IRH, especially at lower ranking positions.

Table 3. F_1 measures for all compared algorithms at ranking positions 1, 2, 5, 8 and 10.

	F_1@1	F_1@2	F_1@5	F_1@8	F_1@10
IRH1	0.450	0.640	0.527	0.414	0.359
IRH	0.470	0.685	0.544	0.420	0.358
QGSO	0.494	0.734	0.589	0.457	0.401
Geo-IRH	0.423	0.644	0.597	0.530	0.487

4. CONCLUSION AND FUTURE WORK

In this paper, a novel and efficient recommendation method is proposed, which fully exploits image content, context and social me-

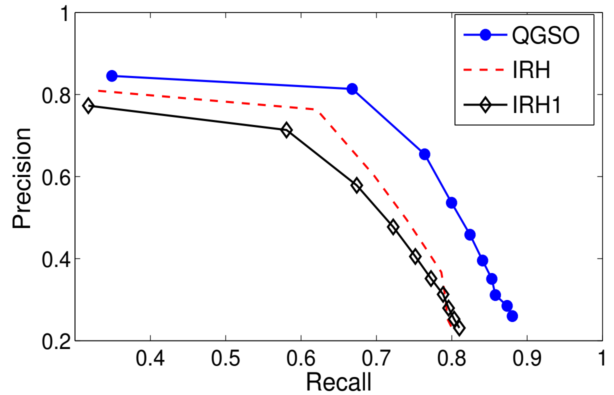

Fig. 1. Averaged Recall-Precision curves for IRH1, IRH and QGSO.

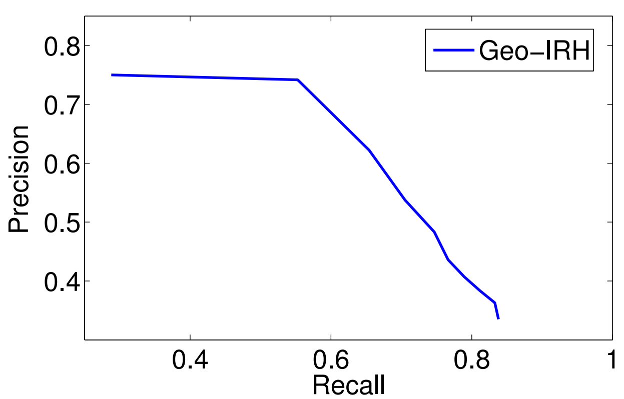

Fig. 2. Averaged Recall-Precision curves for the Geo-IRH method.

dia information. Thanks to hypergraph learning, personalized and geo-referenced image recommendation have been suggested and further improved by enforcing group sparsity constraints. The proposed methods can also accommodate friend recommendation, geo-tag prediction, or image annotation. They can also support the fusion of different types of multimedia, such as audio, video, and text.

5. ACKNOWLEDGMENTS

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

6. REFERENCES

[1] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age," *ACM Comput. Surv.*, vol. 40, no. 2, pp. 5:1–5:60, 2008.

[2] 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 Pattern Recognition*, 2009, pp. 1085–1092.

[3] O. Chum and J. Matas, "Large-scale discovery of spatially related images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 32, no. 2, pp. 371–377, 2010.

[4] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, "Object retrieval with large vocabularies and fast spatial matching," in *Proc. IEEE Computer Vision Pattern Recognition*, 2007, pp. 1–8.

[5] Y. Avrithis, Y. Kalantidis, G. Tolias, and E. Spyrou, "Retrieving landmark and non-landmark images from community photo collections," in *Proc. Int. Conf. Multimedia*, 2010, pp. 153–162.

[6] L. S. Kennedy and M. Naaman, "Generating diverse and representative image search results for landmarks," in *Proc. 17th Int. Conf. World Wide Web*, 2008, pp. 297–306.

[7] C. Berge and Edward Minieka, *Graphs and Hypergraphs*, vol. 7, North-Holland, Amsterdam, 1973.

[8] D. Zhou, J. Huang, and B. Schölkopf, "Learning with hypergraphs: Clustering, classification, and embedding," in *Advances in Neural Information Processing Systems*, 2006, pp. 1601–1608.

[9] 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*, 2012, pp. 779–789.

[10] J. Yu, D. Tao, and M. Wang, "Adaptive hypergraph learning and its application in image classification," *IEEE Trans. Image Processing*, vol. 21, no. 7, pp. 3262–3272, 2012.

[11] 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*, 2010, pp. 391–400.

[12] S. Agarwal, K. Branson, and S. Belongie, "Higher order learning with graphs," in *Proc. 23rd Int. Conf. Machine Learning*, 2006, pp. 17–24.

[13] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, no. 1, pp. 49–67, 2006.

[14] Z. Lin, R. Lui, and Z. Su, "Linearized alternating direction method with adaptive penalty for low-rank representation," in *Proc. Neural Information Processing Systems*, 2011, pp. 612–620.

[15] Z. Qin and K. Scheinberg, "Efficient block-coordinate descent algorithms for the Group Lasso," *Industrial Engineering*, pp. 1–21, 2010.

[16] E. S. Ristad and P. Yianilos, "Learning string-edit distance," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no. 5, pp. 522–532, 1998.

[17] A. Oliva and A. Torralba, "Building the GIST of a scene: The role of global image features in recognition," *Progress in Brain Research*, vol. 155, pp. 23–36, 2006.

[18] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. Journal Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.