On Inferring Image Label Information Using Rank Minimization for Supervised Concept Embedding

Dmitriy Bespalov[1], Anders Lindbjerg Dahl[2], Bing Bai[3] and Ali Shokoufandeh[1]

[1]Department of Computer Science, Drexel University
[2]DTU Informatics, Technical University of Denmark
[3]NEC Labs America

Abstract. Concept-based representation — combined with some classifier (e.g., support vector machine) or regression analysis (e.g., linear regression) — induces a popular approach among image processing community, used to infer image labels. We propose a supervised learning procedure to obtain an embedding to a latent concept space with the pre-defined inner product. This learning procedure uses rank minimization of the sought inner product matrix, defined in the original concept space, to find an embedding to a new low dimensional space. The empirical evidence show that the proposed supervised learning method can be used in combination with another computational image embedding procedure, such as bag-of-features method, to significantly improve accuracy of label inference, while producing embedding of low complexity.

1 Introduction

Inferring label information from image data has a wide range of applications in computer vision and image processing. A common approach to tackle this problem is to first build a descriptor for every image, for instance bag-of-features (BOF) histogram, followed by the label inference formulated as classification or regression analysis problem. Computationally, this process is congruent to embedding of every image to multi-dimensional vector space. The vector space defined by the embedding is sometimes referred to as the *concept space*, and the embedding itself is said to give rise to *concept-based representation* of the data.

Concept-based representation is a popular approach in information retrieval community. The representation describes an item — e.g., text document, image, etc — using sparse vectors of concepts. Construction of the item's concept-based representation can be viewed as category labeling procedure, and corresponding concept set as collection of categories. Furthermore, weights associated with every concept relates information content of the item with the concept's category. Intuitively, two related items should receive similar category labels and formally, relatedness between two items is computed with cosine similarity (i.e., cosine of the angle between two vectors) in the concept space. Naturally, it may be desirable to control concept weights to suit specific category labeling task. We refer to this process as computing *context* for concept-based representation. The context controls which concepts are promoted during construction of the item's concept-based representation.

In this paper, we propose supervised learning procedure that finds an embedding to concept space with the inner product operator tuned for the specific label inference task. Our procedure can be used in conjunction with another label inference method to improve its accuracy. Among all possible concept spaces that can describe a training set, the proposed method estimates the one of lowest dimensionality. The labels from the training data are used to construct "ground truth" similarity matrix, which in turn will be used to compute the inner product operator that defines the embedding. Our method requires concept-based representation for the input data. Hence, it can be applied to any popular unsupervised image embedding procedure used for label inference. The proposed supervised learning framework is applied to the concept space obtained with a variant of unsupervised embedding, the so-called bag-of-features representation. We show that significant improvements in accuracy of label inference is achieved over the original embedding, when supervised learning is used. In addition, the dimensionality of the new concept space is much lower than the original.

Classical examples of linear concept space embedding include Singular Value Decomposition (SVD) based methods, Principal Component Analysis (PCA) [13] and Latent Semantic Indexing (LSI) [6]. LSI finds orthogonal loading vectors that minimizes the reconstruction error of the matrix of interest. These methods have been very successful in pattern recognition and information retrieval. In recent years, different criteria or constraints have been proposed to address specific needs. For instance, Independent Component Analysis (ICA) enforces statistical independence of loading vectors, and sparse coding (SC) promotes sparse components. These methods are unsupervised and consequently independent of the classification task. This can result in unnecessary information loss in the feature extraction stage. Therefore, a large effort has been put into generating the loading vectors directly from the targeted task, using supervised learning. A well known example is supervised metric distance learning [16]. Recently, applications of such methods were reported on large scale problems like information retrieval [1] and image annotations [17]. These supervised methods usually achieve better performance when training labels are available.

A common problem to concept-based representation is selecting the dimensionality of the concept space (i.e., the correct number of concepts). The dimensionality can affect performance of computational tasks in that concept space. Hence, it is usually desirable to find a concept space of low dimensionality. Most methods that achieve this require expensive validation or experience, which might not be available. Components can be ranked in methods like PCA or LSI, but this ranking is not necessarily optimal for the targeted task. Rank Minimization (RM) [10] on the other hand, when combined with the optimization for the supervised embedding, can estimate the optimal dimensionality for the specific label inference task.

In this paper we use supervised learning with RM to obtain concept space embedding tuned for the specific label inference task. We show that our method can significantly improve accuracy of a label inference procedure that uses concept-based representation of images. We experiment with several label inference tasks relating images with meat spoilage, defined on a set of hyper-spectral images of minced meat. The improved accuracy is demonstrated on meat spoilage measured as bacterial count (regression analysis) and sensor panel assessment (classification). In addition only few features are selected, which is highly relevant for interpretation of the obtained result.

2 Related work

Using embedding into concept spaces has a long history in information retrieval. The first and perhaps most well-known method is Latent Semantic Indexing (LSI) [6]. LSI applies SVD on term-document matrix, and simultaneously computes loading and document embedding vectors. Unseen documents can be represented in the concept space by projecting to the loading factors. LSI create an index structure for the documents using concepts instead of "terms", thus can match documents with "synonyms", which was absent in term-based indexing models. LSI is considered the pioneering work that inspired methods such as probabilistic LSI (pLSI) [11] and Latent Dirichlet Allocation (LDA) in a probabilistic framework [2]. There are also alternative methods for generating concepts directly from labeled data. Such approaches have been applied to conceptual embedding and learning in a variety of information retrieval tasks such as link prediction, cross-lingual retrieval, and image annotations [1, 17]. Despite their success these methods suffer from lack of clear strategy for setting the dimensionality of embedding space.

Concept space embedding has also been successfully used in pattern recognition: e.g., simple clustering approach to obtain BOF [5, 15] for images or video. Compared to raw feature matching used in early work [12, 14], the BOF represents an image as a histogram of "visual words", giving rise to image embedding to a vector space where retrieval or classification can take place.

Furthermore, observe that RM is a generalization of sparse representation for matrices. Various pursuit methods for minimizing the L_0 norm attain remarkable results within problems such as image denoising, compression, inpainting and upscaling [4, 9]. In addition, RM methods have found a number of successful applications for problems such as visual tracking [8] and video inpainting [7].

3 Method

Assume, we are given a set of Q training items $d_i \in \mathbb{R}^M$, $i \in [1, Q]$ arranged in columns of matrix $D \in \mathbb{R}^{M \times Q}$. Each element in d_i can be a raw feature or a concept, which can be viewed as an extracted feature. Let $\tilde{S} \in \mathbb{R}^{Q \times Q}$ denote the "Ground-truth Similarity matrix (GSM)" of D. That is, $\tilde{S}_{i,j}$ is assigned with "ground truth" similarity for the corresponding items d_i and d_j. The GSM can be generated in different ways from labeled data, or from unsupervised learning, depending on the task of interest. For example, in classification tasks, we can have

$$\tilde{S}_{i,j} = \begin{cases} 1 & \text{items } i \text{ and } j \text{ in same category,} \\ -1 & \text{otherwise.} \end{cases}$$

In inferring continuous measurement values, we can have element

$$\tilde{S}_{ij} = 1 - 2|c_i - c_j|/R,$$

where c_i and c_j are measurement values for items i and j, respectively; $R = r_1 - r_2$ and r_1 and r_2 are maximum and minimum values of the measurement.

We then define "Contextualized Similarity Matrix"(CSM) $S = D^\top WD$, where W is what we call "Context Matrix". Every element $S_{ij} = d_i^\top W d_j$ describes the "Contextualized Similarity" between sample d_i and d_j. The formulation of contextualized similarity was used in [1] for the problem of text retrieval. Intuitively speaking, the element W_{ij} models the relevance of the i-th element of d_1 and j-th element of d_2. Note that the CSM S is not necessarily $D^T D$, in which $W = I$ and no learning is required.

In this paper, we try to find the optimal context matrix W that minimizes the reconstruction error: $\|S - \tilde{S}\|$. From a supervised learning point of view, the W will catch the essential information that best describes our target task.

We observe that W should be a symmetric matrix, so the similarity within the context maintains the commutative property. Moreover, W will be diagonally dominant matrix, since each item should be ranked most similar to itself. As a result, we obtain that W is a positive semi-definite matrix, so it defines an inner product in concept space. In addition, W gives rise to new concept space embedding. Indeed, since W is positive semi-definite, we obtain $W = P^T P$ where $P \in \mathbb{R}^{r \times M}$ and $r = \text{rank}(W)$. Matrix P transforms any item $d \in \mathbb{R}^M$ defined in the original concept space, to the new concept space $Pd \in \mathbb{R}^r$. Among all possible solutions W to the norm minimization, selecting the one with minimum rank, describes the new concept space with the least number of concepts. Clearly, if we minimize the rank of W so that $r < M$, we also obtain a lower-dimensional embedding of the original concept space. It should also be noted that rank minimization can be considered a generalization of vector sparsity for matrices [10]. Indeed, minimizing a rank of a diagonal matrix is equivalent to minimizing number of non-zero elements of the diagonal vector.

The problem of *learning minimum rank context embedding* W for the given set of training items D can be formulated as optimization problem (1):

$$\begin{aligned} &\min_{W} \|D^T W D - \tilde{S}\| + \gamma \, \text{rank}(W) \\ &s.t. \quad W \succeq 0. \end{aligned} \tag{1}$$

Here, γ is regularization parameter. In general, optimizations with rank(.) can not be solved directly. However, Fazel *et al.* [10] showed that approximate solution to (1) can be obtained by replacing rank(W) objective by its smooth surrogate function $\log\det(W + \delta I)$. This results in the following iterative method for approximating rank(W):

$$W_{k+1} = \operatorname*{argmin}_{W \in \mathcal{C}} \text{Trace}(W_k + \delta I)^{-1} W \tag{2}$$

where $\mathcal{C}$ is a set of optimization constraints from (1). Computing each iteration of W_{k+1} requires solving a semi-definite program (SDP), and its initial estimate W_0 can be obtained with another SDP: e.g., replacing rank(W) objective with Trace(W) (the so-called trace heuristic). Empirical evidence in [10] suggests that $\log\det$ heuristic produces better approximations of rank(W) (i.e., lower rank solutions can be found) than trace heuristic. Our procedure, detailed in Algorithm 1, uses $\|D^T W D - \tilde{S}\|$ objective to find W_0 followed by few iterations of (2). The value for parameter γ was empirically chosen to ensure the rank is not optimized at the expense of the norm $\|D^T W D - \tilde{S}\|$ minimization. Parameter β ensures the low rank solution W results in the norm value very close to the minimum.

Algorithm 1 Minimum Rank Context Embedding

Input: $\{d_i\}$, set of training items
$D \in \mathbb{R}^{M \times Q}$, concept based representation for d_i arranged in columns

Output: P, minimum rank embedding for learned context W

$\gamma = 0.0001$
$\beta = 1 + 10^{-6}$
$\tilde{S} = \text{getGSM}(\{d_i\})$ # build Ground-truth Similarity Matrix
$W_0 = \arg\min_{W \in \mathcal{C}} \|D^T W D - \tilde{S}\|$ # Initialize W_0 (solve SDP)
$\kappa = \|D^T W_0 D - \tilde{S}\|$
$\mathcal{C} = \left\{ \begin{array}{l} W \succeq 0 \\ \|D^T W D - \tilde{S}\| \leq \beta\kappa \end{array} \right\}$ # create set of constraints
for $k = 1$ to 3 **do**
$W_k = \arg\min_{W \in \mathcal{C}} \left[\|D^T W D - \tilde{S}\| + \gamma \,\text{Trace}(W_{k-1} + \delta I)^{-1} W \right]$
end for # log det approx. (solve 3 SDPs)
decompose $W_3 = P^T P$
return P

Due to high complexity of SDP solvers, the size of problems that can be handled is very limited. On the other hand, a large number of concepts Q is needed to describe even moderately sized datasets (e.g., tens to hundreds of thousands). Hence, handling optimizations of that size by Algorithm 1 is impractical. We employ Latent Semantic Indexing [6] (LSI), a well known dimensionality reduction technique that is widely used in information retrieval community for efficient document indexing and retrieval. For the given set of training items $\{d_i\}$ and their concept-based descriptors arranged in columns of D, SVD of $D = U \Sigma V^T$ is computed using SlepC library [1]. Matrix $\Sigma_l^{-1} U^T$ forms an embedding for every concept-based item descriptor in l-dimensional space, where Σ_l is a matrix formed with l most significant singular values of D. Computing $\hat{D} = \Sigma_l^{-1} U^T D$ yields low dimensional (e.g., $15 \leq l \leq 100$) representation of training items, and matrix $\hat{D}$ is passed to Algorithm 1 for learning minimum rank context embedding.

Alternatively, one may choose to use first-order method described in [18] and [19] to approximate optimization (1) using convex relaxation. Each iteration of their method requires only a SVD, that can handle large matrices (on the order of hundreds of million non-zero entries) using efficient libraries such as Slepc. Using such an approximation to (1) in Algorithm 1 will enable learning context embedding for larger datasets de-

[1] http://www.grycap.upv.es/slepc/

scribed in higher dimensional concept spaces. As a result, no dimensionality reduction such as LSI or PCA will be required.

The primary objective of this work was to empirically validate the viability of the proposed framework, described in Algorithm 1, to improve the accuracy of label inference and reduce dimensionality of concept space. The results of our empirical validations are presented in Section 4. Having obtained initial validation for the minimum rank context embedding framework, the most prominent direction of our future work is re-formulating optimization (1) so a first order approximation method can be used instead of SDP.

4 Experiments

Non-destructive methods for food inspection is important in industrial manufacturing, and image processing can be one way of measuring such quality parameters. We present results for label inference experiments on multi-spectral images of minced meat with ground truth labels for storage degradation. The rest of this section is organized as follows. Section 4.1 describes dataset of meat images used in our empirical validations. Algorithm 1 assumes the input data is represented in concept space. Hence, Section 4.2 presents a variant of popular BOF-based image representation used to embed images to a concept space. Finally, results of label inference experiments performed in both concept spaces, can be found in Section 4.3.

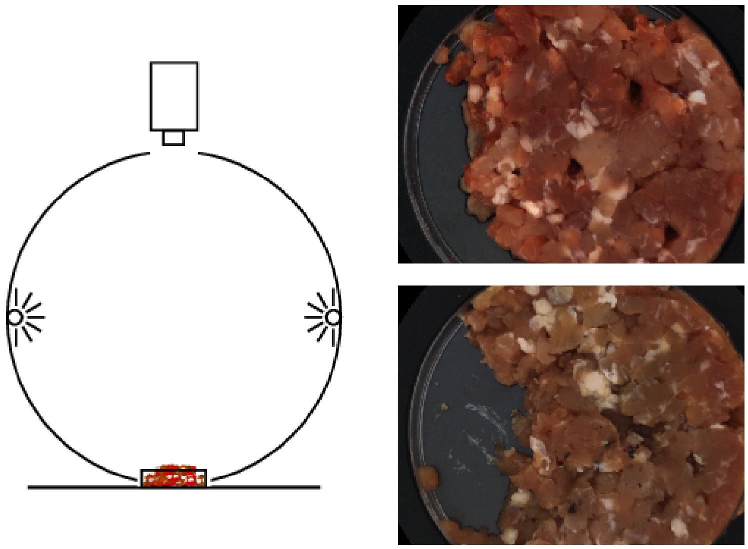
4.1 Data

Our multi-spectral images are acquired from device called VideometerLab[2], which employs wavelength specific LED illumination placed in an integrating sphere, see Figure 1. Hereby the meat sample is illuminated by narrow spectral bands of diffuse light spanning the spectrum at 18 wavelengths from 405 – 970 nm. An image is acquired from each spectral band using a normal CCD camera. The resolution of the sample images is 1280×960 pixels. The minced meat samples have been stored at different temperature and under different package conditions, and the spoilage has been assessed with six bacteria count methods and a sensory panel assessment into three categories – *fresh* (F), *semi-fresh* (SF), and *spoiled* (S). Figure 1 shows an example of the image data, and Table 4.1 gives an overview of the measured parameters.

4.2 Explicit Unsupervised Feature Extraction

The proposed *concept-based representation for images* treats each image as a collection of *terms;* i.e., pre-defined descriptors for extracted image features. Given a set of images, we construct their concept-based representation as follows. First, a subset of images is randomly selected and all of their terms are recorded as *concepts*. Then, the remaining images are interpreted in terms of these concepts. Every term associated with an image votes for k nearest concepts that are identified using approximate nearest

[2] http://www.videometer.com/



(a) (b)

Fig. 1. Principle sketch of the VideometerLab integrating sphere (a), with camera at the top, LED's along the equatorial rim, and a petri dish at the bottom. The wavelengths are 405, 435, 450, 470, 505, 525, 570, 590, 630, 645, 660, 700, 850, 870, 890, 910, 940 and 970 nm. Image examples in a petri dish (b). Top is a fresh meat sample and bottom is a sample stored for 67 hours at 20°C. Note the black spots of bacteria growth on the bottom sample. Color images are made from three spectral band (R – 630nm, G – 525nm, and B – 450nm).

Table 1. Data parameters. Bacteria is measured by blending the meat sample and placing it on an agar medium and counting the number of bacteria colonies after an incubation period.

Storage				
AIR (normal atmosphere)		MAP (modified atmosphere)		
Temperature				
0°C	5°C	10°C	15°C	20°C
Storage time				
0 – 590 hours				
Incubation methods for bacteria count				
PCA (Plate Count Agar – 30°C for 48 hours)				
PAB (Pseudomonas Agar Base – 30°C for 48 - 72 hours)				
STAA (Streptomycin Thallous Acetate-Actidione Agar – 25°C for 72 hours)				
RBC (Rose Bengal Chloramphenicol Agar – 25°C for 72 hours)				
VRBGA (Violet Red Bile Glucose Agar – 37°C for 18 - 24 hours)				
MRS (Man-Rogosa-Sharp medium – 30°C for 48 hours)				
Acidity measurement				
pH				
Sensory panel				
F – fresh		*SF* – semi-fresh		*S* – spoiled

neighbor data structure. A histogram bin, associated with every concept, accumulates weights received from each term. The weight $\chi(t, q)$ between term t and concept q is computed using

$$\chi(t,q) = \begin{cases} 1 - \|t-q\| & \text{when } \|t-q\| \leq 1 \\ 0 & \text{otherwise} \end{cases},$$

where $\|t - q\|$ is Euclidean distance between term t and concept q. Once all terms are processed, histogram bins are normalized to form concept-based descriptor for the image. Formal overview of this procedure is presented in Algorithm 2.

Algorithm 2 Concept-based Image Interpretor

Input: $q_i \in \mathbb{Q}$, concepts terms
$t_j \in \mathbb{I}$, image terms
k, nearest neighbor parameter
Output: d, concept-based representation of image $\mathbb{I}$

```
d = 0
for all t_j ∈ I do
  {q_{n_i}} = kNN(Q, t_j, k)                  # select k closest concepts q_{n_i} for term t_j
  for all {q_{n_i}} do
    d[q_{n_i}] = d[q_{n_i}] + χ(t_j, q_{n_i})  # accumulate weight for concept q_{n_i}
  end for
end for
d = d/||d||                                   # normalize concept-based image descriptor
return d
```

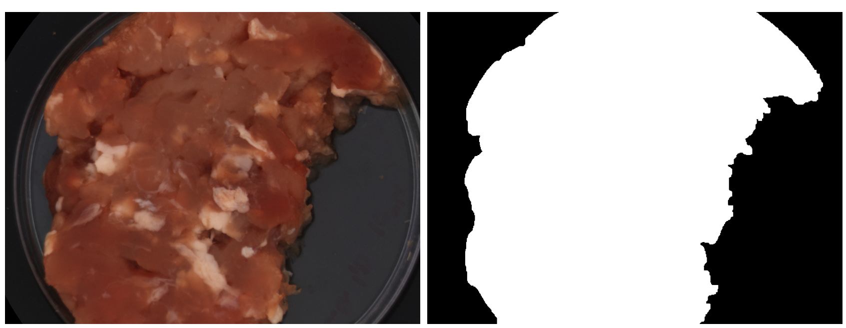


Fig. 2. Illustration of minced meat region mask.

Note that the aforementioned method is a variant of well-known bag-of-features representation. However, it does not perform any clustering or other feature quantization when concepts are created. In the spirit of recent empirical evidence presented by Boiman *et al.* [3], we decided against feature quantization.

Images of minced meat samples contain intensities in 18 spectral bands. Image features that are used as terms in the proposed concept-based representation are defined as follows. The proposed image feature descriptor combines histograms of intensity changes in neighboring spectral bands. Each feature descriptor is constructed for a rectangular image window. For every band b_i, we compute the difference of intensity values with the two neighboring bands b_{i-1} and b_{i+1}. Then, each pixel in band b_i is characterized with the angle formed by the two corresponding changes in intensities. A normalized orientation histogram is computed for each of the 16 bands (two extrema bands are not used), and the orientation histograms are concatenated to form a feature descriptor for the rectangular window. We use 9 orientation bins, which results in $9 \times 16 = 144$ dimensional feature descriptors. Image features are computed for 17×17 pixel windows, sampled every 11 pixels.

We extract image features for regions that depict minced meat only, while ignoring regions that depict plate or table. Mask for minced meat region is computed for every sample image. The mask is constructed by training a foreground-background classifier from one hand-labeled image. From the training image a random subset of pixels was selected and clustered to 30 clusters using k-means algorithm. Each cluster was labeled foreground or background, depending on the majority of the labels of the pixels in the cluster. For an unknown image the nearest cluster center was found, using Euclidean distance, and the pixel was labeled with the label of the cluster. Finally, morphological opening and closing was applied to smooth the result, using 11 pixels squared structuring element. Only features residing entirely within the meat region are selected as terms for the sample. Please refer to Figure 2 for the illustration of a minced meat region mask.

We now present findings for label inference experiments using the aforementioned concept-based representation for meat samples. We measure accuracy of SVM classifier and linear regression, trained in concept space computed with Algorithm 2. We then use Algorithm 1 to obtain embedding to new concept space, and compare label inference performance in the new space with the original.

4.3 Label Inference Results

We collected a total of 141 pork meat samples. In all of the experiments, 10 meat samples were randomly selected as terms, while 30 samples were used for training, and the rest of the samples were used for testing. Every experiment was repeated 100 times and average prediction accuracy (or error) was recorded. We retained 15 most significant singular values in LSI procedure and used them in the construction of matrix $\hat{D}$, that is then passed to Algorithm 1. Inferring sensory panel scores is a classification problem, since only three labels (fresh, semi-fresh and spoiled) are available. However, predicting bacteria count values for the meat samples requires regression analysis, since the bacteria count values are continuous. We used LibSVM tool[3] for both classification and regression problems: SVM classifier with linear kernel and parameter $C = 32$, and linear regression with parameters $C = 32$ and $\epsilon = 0.1$ (used in loss function).

[3] http://www.csie.ntu.edu.tw/ cjlin/libsvm/

For inference of sensory panel score, the average classification accuracy for unsupervised embedding (Algorithm 2) was 69.2%, while for supervised embedding with RM (Algorithm 2 followed by 1) average accuracy was 73.6%. In addition, supervised embedding always resulted in two- or three- dimensional concept spaces, while the dimensionality of the original concept space obtained with unsupervised BOF-based embedding was 15.

The average mean squared error for predicting six various bacteria count measurements can be found in Table 4.3. In addition to significant improvement in accuracy, the dimensionality of the concept space, obtained with the supervised embedding procedure, was always between two and five.

Table 2. Predicting bacteria counts. Mean squared error.

Method	PCA	PAB	STAA	RBC	VRBGA	MRS
Algorithm 2	1.28	2.32	1.19	3.72	3.55	1.34
Algorithm 2 followed by 1	0.44	0.74	0.38	0.70	0.99	0.46
Measurement min	5.1	4.9	3.7	2.3	2.2	3.2
Measurement max	9.9	9.9	8.2	6.8	8.8	8.1

5 Conclusion

We presented a supervised learning procedure to compute embedding in concept space, where inner product operator is tuned for specific task. We show that the proposed framework can significantly improve label inference performance of prior art methods that rely on concept-based representation. However, current formulation of the procedure (i.e. Algorithm 1) requires solving several SDP's, which significantly limits the complexity of the problems that can be tackled. Our imminent goals for future work is to introduce first-order approximation method for (1). Another promising direction for future work is introduction of novel image features used as terms in Algorithm 2 that can improve label inference on the meat dataset.

6 Acknowledgments

The first and fourth authors of this work are supported by National Science Foundation grant #0803670 under the IIS Division, and Office of Naval Research grant ONR-N000140410363. The second author is financed by the Centre for Imaging Food Quality project which is funded by the Danish Council for Strategic Research (contract no 09-067039) within the Programme Commission on Health, Food and Welfare.

References

1. B. Bai, J. Weston, R. Collobert, D. Grangier, K. Sadamasa, Y. Qi, O. Chapelle, and K. Weinberger. Supervised semantic indexing. In *Proceeding of the 18th ACM Conference on Information and Knowledge Management*, pages 187–196. ACM, 2009.
2. D.M. Blei, A.Y. Ng, and M.I. Jordan. Latent dirichlet allocation. *The Journal of Machine Learning Research*, 3:993–1022, 2003.
3. O. Boiman, E. Shechtman, and M. Irani. In defense of nearest-neighbor based image classification. In *IEEE Conference on Computer Vision and Pattern Recognition, 2008.*, pages 1–8. IEEE, 2008.
4. A.M. Bruckstein, D.L. Donoho, and M. Elad. From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM review*, 51(1):34–81, 2009.
5. G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray. Visual categorization with bags of keypoints. In *Workshop on Statistical Learning in Computer Vision, ECCV*, volume 1, page 22. Citeseer, 2004.
6. S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman. Indexing by latent semantic analysis. *Journal of The American Society for Information Science*, 41(6):391–407, 1990.
7. Tao Ding, Mario Sznaier, and Octavia I. Camps. A rank minimization approach to video inpainting. *Computer Vision, IEEE International Conference on*, 0:1–8, 2007.
8. Tao Ding, Mario Sznaier, and Octavia I. Camps. Receding horizon rank minimization based estimation with applications to visual tracking. In *Proceedings of the 47th IEEE Conference on Decision and Control, CDC 2008, December 9-11, 2008, Cancún, México*, pages 3446–3451, 2008.
9. M. Elad. *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. Springer, 2010.
10. M. Fazel, H. Hindi, and S. P. Boyd. Log-det heuristic for matrix rank minimization with applications to hankel and euclidean distance matrices. In *Proceedings American Control Conference*, pages 2156–2162, 2003.
11. T. Hofmann. Probabilistic latent semantic indexing. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, pages 50–57. ACM Press New York, NY, USA, 1999.
12. D.G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.
13. K. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2(6):559–572, 1901.
14. C. Schmid and R. Mohr. Local grayvalue invariants for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(5):530–535, 1997.
15. J. Sivic and A. Zisserman. Video Google: Efficient visual search of videos. *Lecture Notes in Computer Science*, 4170:127, 2006.
16. K. Q. Weinberger, J. Blitzer, and L. K. Saul. Distance metric learning for large margin nearest neighbor classification. In *NIPS*. MIT Press, 2006.
17. J. Weston, S. Bengio, and N. Usunier. Large scale image annotation: Learning to rank with joint word-image embeddings. *Machine learning*, 81(1):21–35, 2010.
18. J. Yang and X. Yuan. An Inexact Alternating Direction Method for Trace Norm Regularized Least Squares Problem. *Report, Department of Mathematics, Nanjing Uinversity*, 2010.
19. X. Yuan. Alternating Direction Methods for Sparse Covariance Selection. In *20th International Symposium of Mathematical Programming (ISMP)*, 2009.