# Learning Sparsity Inducing Analysis Operators for Discriminative Similarity Metrics

Çağdaş Bilen, Joaquin Zepeda and Patrick Perez
Technicolor, 35576, Cesson-Sévigné, France
email: Cagdas.Bilen Joaquin.Zepeda Patrick.Perez at technicolor.com

***Abstract*—We introduce a new form of similarity measure to be used for fast image comparison in large databases. The proposed approach makes use of sparse vectors obtained through a learned analysis operator to compute a computationally efficient and flexible image similarity metric.**

## I. Introduction and Proposed Method

A general image search algorithm can be seen as having two main goals: i) that of finding correctly matching images given a task-specific search criteria and ii) that of doing so in a time and resource efficient manner, particularly in the context of large image databases. In addressing the first goal, discriminative Mahalanobis metric learning methods have become an important part of the research toolbox. Such methods can be seen as applying an explicit linear transform to the image feature vector with the goal of making distance computations between transformed feature vectors better correspond to the search criteria. The linear transform can be learned using one of a variety of objectives in order to adapt it to various possible search criteria including image classification [4], [7], face verification [2], [5], or image ranking [1]. Common to all these methods is the fact that the learned linear transform is a complete or undercomplete matrix that is constant for all image feature vectors. Alternatively, similarity metrics in the form of correlation of two projected vectors have also been used for image and audio comparison and ranking [1], [3]. As opposed to approaches based on distance metrics, the correlation based approach also has the advantage of computationally benefitting from sparse representations.

In this abstract, given two feature vectors $\mathbf{y}_i, \mathbf{y}_j \in \mathbb{R}^N$ representing two images $i, j$, we present a similarity metric of the form

$$\mathrm{S}(\mathbf{y}_i, \mathbf{y}_j) = \mathbf{z}_i^\top \mathbf{B} \mathbf{z}_j \tag{1}$$

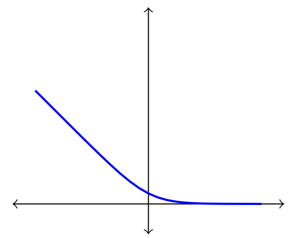
in which the sparse representations $\mathbf{z}_i, \mathbf{z}_j$ are obtained using the soft thresholding function acting on each entry of a vector with the set of thresholds $\boldsymbol{\lambda}$ as

$$\mathbf{z}_i = \mathrm{soft}_{\boldsymbol{\lambda}}(\mathbf{A}\mathbf{y}_i) \qquad , \qquad \mathbf{z}_j = \mathrm{soft}_{\boldsymbol{\lambda}}(\mathbf{A}\mathbf{y}_j). \tag{2}$$

The matrices $\mathbf{A}$, $\mathbf{B}$ and the parameter vector $\boldsymbol{\lambda}$ are all learned from a training dataset of pairs $(\mathbf{z}_{i_p}, \mathbf{z}_{j_p})$, $p = 1, \ldots, m$ selected from a training set of size $T$ with each pair labeled as similar or dissimilar ($\gamma_p = 1$ or $-1$) by minimizing an objective function of the form

$$\begin{aligned} \bar{\mathbf{A}}, \bar{\mathbf{B}}, \bar{\boldsymbol{\lambda}} = \quad & \mathrm{argmin}_{\mathbf{A},\mathbf{B},\boldsymbol{\lambda}} \textstyle\sum_{p=1}^{m} \ell(\gamma_p \mathbf{z}_{i_p}^\top \mathbf{B} \mathbf{z}_{j_p}) + \psi(\mathbf{A}) + \phi(\mathbf{B}) \\ & \text{s.t.} \quad \mathbf{z}_t = \mathrm{soft}_{\boldsymbol{\lambda}}(\mathbf{A}\mathbf{y}_t),\ t = 1, \ldots, T \end{aligned} \tag{3}$$

where the penalty function $\ell(.)$ can be selected as the continuous hinge function as shown in Figure 1. The functions $\psi(.)$ and $\phi(.)$ are regularization functions for the matrices $\mathbf{A}$ and $\mathbf{B}$. Even though the objective function in (3) is non-linear and non-convex, it can still be minimized using off the shelf optimization methods such as stochastic gradient descent.



$$\ell(x) = \log(\exp(-\alpha x) + 1)/\alpha$$

Fig. 1. Continuous hinge loss

Even though the use of correlation based similarity function with sparse vectors as in (1) is proposed in earlier works [1], [3], making use of analysis operators to obtain the sparse codes as in (2) to be used in this similarity function is a new idea with multiple advantages. Firstly, as compared to computing a sparse representation in a dictionary, the operation in (2) is computationally much more simple and faster. Secondly, the use of an analysis operator also enables an asymmetric system with a similarity function in the form of

$$\mathrm{S}(\mathbf{y}_i, \mathbf{y}_j) = \mathbf{y}_i^\top (\mathbf{A}^\top \mathbf{B} \mathbf{z}_j) \tag{4}$$

such that the comparison is even faster for a new item.

## II. Preliminary Results

Initial experiments have been performed using the proposed similarity metric in (1) and learning the parameters $\mathbf{A}$ and $\boldsymbol{\lambda}$ as in (3) assuming $\mathbf{B} = \mathbf{I}$ and enforcing normalized rows for the matrix $\mathbf{A}$. The image descriptors and the datasets given in [6] are used to train and test the learned similarity metric with difference being normalizing the descriptors. The resulting false positive rates for 95% recall are shown in Table I. For comparison, the performance of using the normalized descriptors without any learning has been shown. Even without training for the matrix $\mathbf{B}$, the proposed metric provides a good similarity estimation performance. The presentation will include further results also with the training of $\mathbf{B}$ and using additional datasets.

TABLE I
False positive rate corresponding to 95% recall

| *Trained Dataset* | *Tested Dataset* | *Proposed* | *Baseline* |
|---|---|---|---|
| Yosemite | Liberty | **17.83** | 23.82 |
| Yosemite | Notredame | **10.71** | 17.25 |
| Liberty | Yosemite | 20.20 | **17.36** |
| Liberty | Notredame | **11.25** | 14.77 |
| Notredame | Liberty | **16.87** | 21.59 |
| Notredame | Yosemite | 20.18 | **19.50** |

## REFERENCES


[1] G. Chechik, V. Sharma, U. Shalit, and S. Bengio. Large scale online learning of image similarity through ranking. *Journal of Machine Learning Research, JMLR*, pages 1109–1135, 2010.

[2] M. Guillaumin, J. Verbeek, and C. Schmid. Is that you? Metric learning approaches for face identification. In *International Conference on Computer Vision*, Sept. 2009.

[3] R. F. Lyon, M. Rehn, S. Bengio, T. C. Walters, and G. Chechik. Sound retrieval and ranking using sparse auditory representations. *Neural Comput.*, 22(9):2390–2416, Sept. 2010.

[4] T. Mensink, J. Verbeek, F. Perronnin, and G. Csurka. Metric Learning for Large Scale Image Classification : Generalizing to New Classes at Near-Zero Cost. *Pattern Analysis and Machine Intelligence*, 2012.

[5] K. Simonyan, O. M. Parkhi, A. Vedaldi, and A. Zisserman. Fisher Vector Faces in the Wild. In *British Machine Vision Conference*, 2013.

[6] K. Simonyan and A. Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition, Sept. 2014.

[7] K. Q. Weinberger and L. K. Saul. Distance Metric Learning for Large Margin Nearest Neighbor Classification. *The Journal of Machine Learning Research*, pages 207–244, 2009.