Proceedings of International Joint Conference on Neural Networks, Atlanta, Georgia, USA, June 14-19, 2009

Optimal Dimension Reduction for Image Retrieval with Correlation Metrics

Yuhua Zhu, Washington Mio and Xiuwen Liu

***Abstract*— We investigate content-based image retrieval employing a representation of images based on the statistics of their spectral components and a new linear dimension reduction technique. This linear dimension reduction technique is designed to optimize class separation with respect to metrics derived from cross-correlation of spectral histograms. Our approach to retrieval involves a preliminary classification step to index images in a database followed by a class-by-class retrieval step. We carry out several experiments with the Corel database and compare the outcome with several results previously reported in the literature.**

I. Introduction

The theme of this paper is content-based image retrieval based on: (i) image representations obtained from the statistics of spectral components; (ii) a metric derived from cross-correlation of spectral features; (iii) a dimension reduction technique to optimize class separation with respect to the cross-correlation metric. We demonstrate that the proposed representation and metric have enough discriminative power to allow us to completely bypass any form of explicit image segmentation and yet obtain higher retrieval performance than many previously proposed systems. We take a two-step approach to image retrieval. First, we employ a new learning technique to classify and index all images in a database. Subsequently, given a query image, we use the classification tool to rank the classes according to their compatibility with the query image, and then retrieve images from one class at a time according to this ranking. One of the challenges in developing general image retrieval methods is that the breadth of the semantic categories can vary a great deal. Classification problems may range from discerning landscapes and indoor scenes to categorizing specific objects such as cars or books. Thus, an important component in the proposed approach is an effective technique to simultaneously learn representations that optimize class separation with respect to the cross-correlation metrics and reduce the dimension of the spectral-feature space for robustness and computational efficiency.

Zhu *et al.* studied the statistics of spectral components for texture analysis and synthesis [12]. In particular, they demonstrated that marginal distributions of spectral components characterize homogeneous textures; other studies include [6] and [10]. In [4], an Euclidean representation of histograms of spectral components was used for image classification and retrieval in combination with a dimension reduction technique called Optimal Factor Analysis (OFA). The technical details and application of OFA are revised in [5]. In this paper, we develop a variant of OFA for metrics based on cross-correlation and use it for both classification and retrieval. Our viewpoint is that one can enhance the retrieval performance by adopting a metric based on correlation of histograms of spectral components and optimizing the dimension reduction strategies. To corroborate this view, we carry out several experiments on the same data sets used in [9], [5], as this allow us to compare the results objectively. Image retrieval strategies employing a variety of methods have been investigated in [9], [1], [7], [8], [11], [2]; further references can be found in these papers. Some of these proposals also employ a relevance feedback mechanism in an attempt to progressively improve the retrieval performance.

The paper is structured, as follows. In Section II, we describe the features and metrics used in classification and retrieval. Section III describes the dimension reduction techniques and Section IV is devoted to a discussion of the numerical aspects of dimension reduction which is based on stochastic optimization. In Sections V and VI we employ the methods developed to image categorization and retrieval, discuss the results of several experiments and compare them with results previously reported in the literature.

II. Image Representation and Metrics

Let I be a monochrome image and F a convolution filter. The spectral component of I associated with F is the image I_F obtained by applying F to the image I. If I is a color image, we apply the filter to its R,G,B channels to obtain 3 filtered images. For a given set of bins, which is assumed fixed, we let $h(I,F)$ denote the corresponding histogram of I_F. We refer to $h(I,F)$ as the spectral histogram (SH) feature of I associated with F. If the number of bins is b, then $h(I,F)$ can be viewed as a vector in $\mathbb{R}^b$. Figure 1 illustrates the process of obtaining SH-features from an image. Frames (a) and (b) show a color image and its green channel response to a Gabor filter, respectively. The last panel shows an 11-bin histogram of the filtered image.

If $\mathcal{F} = \{F_1, \ldots, F_r\}$ is a bank of filters, we often combine the SH-features $h(I,F_i)$, $1 \leqslant i \leqslant r$, into a single m-vector

$$h(I,\mathcal{F}) = (h(I,F_1), \ldots, h(I,F_r)), \tag{1}$$

where $m = rb$. For a color image, $m = 3rb$. In [5], this representation of images was used in conjunction with linear dimension reduction techniques for content-based image classification and retrieval based on the usual Euclidean metric. One of our goals is to demonstrate that, in addition to

Yuhua Zhu and Xiuwen Liu are with the Department of Computer Science, Florida State University(email: {zhu, liux}@cs.fsu.edu). Washington Mio is with the Department of Mathematics, Florida State University(email: {mio}@math.fsu.edu).

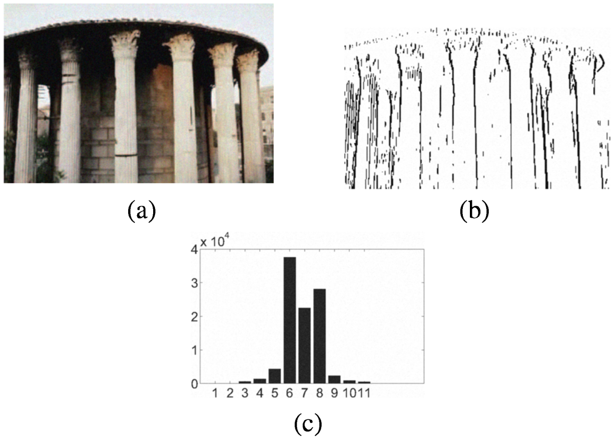


Fig. 1. (a) A color image; (b) Its green channel response to a Gabor filter; (c) 11-bin histogram of the filtered image.

the representation used, the choice of metric is very important in the classification of image content. In this section, we offer preliminary evidence of this fact, by showing that a significant improvement already can be achieved by simply replacing the Euclidean metric with a (pseudo) metric derived from correlation of SH-features.

A. *Cross-Correlation Metrics*

Given $x = (x_1, \ldots, x_b)$ and $y = (y_1, \ldots, y_b)$, denote their mean values by $\hat{x} = \sum x_i/n$ and $\hat{y} = \sum y_i/n$ and let $\bar{x} = x - \hat{x}$ and $\bar{y} = y - \hat{y}$ be the corresponding centered vectors. If $\bar{x}$ and $\bar{y}$ are nonzero, the correlation coefficient $(\bar{x}/\|\bar{x}\|) \cdot (\bar{y}/\|\bar{y}\|)$ can be viewed geometrically as the cosine of the angle between $\bar{x}$ and $\bar{y}$, which can be turned into a (pseudo) metric d by defining

$$d(x, y) = \arccos\left(\frac{\bar{x}}{\|\bar{x}\|} \cdot \frac{\bar{y}}{\|\bar{y}\|}\right). \qquad (2)$$

Given a bank of filters $\mathcal{F}$ and images I_1 and I_2, the quantity

$$\delta_{\mathcal{F}}(I_1, I_2) = \; d(h(I_1, \mathcal{F}), h(I_2, \mathcal{F})) \qquad (3)$$

defines a pseudo-metric on image space based on cross-correlation of the full SH-feature associated with $\mathcal{F}$.

III. Optimal Dimension Reduction

In [5], a dimension reduction technique called Optimal Factor Analysis (OFA) was developed by Mio *et al.*, which seeks to find a linear mapping that reduces the dimension of feature space and optimizes the discriminative ability of the K-nearest neighbor (KNN) classifier with respect to the Euclidean metric, as measured by the performance on training data. We propose a variant of this dimension reduction technique modified for the cross-relation metrics of Section II-A. Since we no longer will have a family of histograms once we map the feature vector (1) to a lower dimensional space, instead of $\delta_{\mathcal{F}}$, we use the more general metric d defined in (2).

Suppose that the training data consists of a collection of labeled m-dimensional feature vectors representing K different classes of objects. For each class c, $1 \leq c \leq K$, we let $x_1^c, \ldots, x_{t_c}^c$ be the training vectors in class c. If $A \colon \mathbb{R}^m \to \mathbb{R}^k$ is a linear mapping, the quantity

$$\rho(x_i^c; A) = \frac{\min_{c \neq b, j} d^2(Ax_i^c, Ax_j^b)}{\min_{j \neq i} d^2(Ax_i^c, Ax_j^c) + \epsilon} \qquad (4)$$

measures how well the nearest-neighbor (NN) classifier applied to the transformed data, with respect to the cross-correlation metric (2), identifies the element x_i^c as a member of its class ($\epsilon > 0$ is a small number, generally irrelevant in applications, used to disallow vanishing denominators). A large value of $\rho(x_i^c; A)$ indicates that, after the transformation A is applied, x_i^c lies much closer to a training sample in its own class than to samples in other classes. A value of $\rho(x_i^c; A)$ near 1 indicates a transitional behavior between correct and incorrect decisions by the NN classifier. Expression (4) can be easily modified to quantify the performance of the KNN classifier. Note that if A maps any of the feature vectors in the training set to zero, then $\rho(x_i^c; A)$ is not defined. However, in practice, we can easily circumvent the problem by perturbing A slightly.

The goal is to choose a transformation A that maximizes the average value of $\rho(x_i^c; A)$ over the training set. In other words, to maximize the performance function

$$F(A) = \frac{1}{K} \sum_{c=1}^{K} \left(\frac{1}{t_c} \sum_{i=1}^{t_c} \rho(x_i^c; A) \right). \qquad (5)$$

Scaling a matrix A does not affect decisions made by the NN classifier in the reduced feature space and this is reflected in the (near) scale invariance of F, as $F(A) \approx F(\alpha A)$, for $\alpha > 0$. Equality does not hold because of ϵ, but in applications ϵ is negligible. Thus, if we identify a linear map $A \colon \mathbb{R}^m \to \mathbb{R}^k$ with a $k \times m$ matrix, we can treat F as a function defined on $\mathbb{R}^{k \times m}$ and scale-invariance implies that it suffices to maximize F over matrices of unit norm. Letting

$$\mathbb{S} = \left\{ A \in \mathbb{R}^{k \times m} : \|A\|^2 = \mathrm{tr}\,(AA^T) = 1 \right\} \qquad (6)$$

be the unit sphere in $\mathbb{R}^{k \times m}$, the goal is to find

$$\hat{A} = \operatorname*{argmax}_{A \in \mathbb{S}} F(A). \qquad (7)$$

IV. Numerical Optimization

We take a stochastic gradient approach with simulated annealing to the numerical estimation of $\hat{A}$. The optimization strategy is similar to that used in the OCA search of [3], but the implementation is simpler because the domain is a sphere instead of a Grassmann manifold.

To estimate the deterministic spherical gradient $\nabla_{\mathbb{S}} F(A)$, we first calculate $\nabla_{\mathbb{R}^{k \times m}} F(A)$, the gradient of F as a function on $\mathbb{R}^{k \times m}$. Since F is nearly scale invariant, $\nabla_{\mathbb{R}^{k \times m}} F(A) \approx \nabla_{\mathbb{S}} F(A)$, as the component normal to the sphere is almost negligible. In practice, we do subtract the normal component to minimize errors. The calculation of the gradient is based on estimations of the partial derivatives. For $1 \leqslant i \leqslant k$, $1 \leqslant j \leqslant m$, let E_{ij} be the $k \times m$ matrix whose

(i, j) entry is 1 and all others vanish. The partial derivative of F in the direction E_{ij} is estimated as

$$\partial_{ij} F(A) \approx \frac{F(A + \delta E_{ij}) - F(A)}{\delta},$$

with $\delta > 0$ small. Since only one entry of A is changed by adding δE_{ij}, the estimation of $\nabla_{\mathbb{S}} F$ through $\nabla_{\mathbb{R}^{k \times m}} F$ is a manageable calculation.

Our next goal is to add a stochastic component to the deterministic gradient field $\nabla_{\mathbb{S}} F$ on $\mathbb{S}$. This can be accomplished by adding a component obtained from a stochastic process on the sphere. To simplify the model, we employ a stochastic component on $\mathbb{R}^{k \times m}$ projected onto the tangent space of the sphere. Let $\Pi_A : \mathbb{R}^{m \times k} \to T_A \mathbb{S}$ be the orthogonal projection onto the tangent space of the sphere at A. We proceed, as follows:

1) Choose $A \in \mathbb{S}$, a cooling factor $\gamma < 1$, an initial temperature $T_0 > 0$, a step size $\delta > 0$, and a positive integer N to control the number of iterations.
2) Set $t = 0$ and initialize the search with $A_t = A \in \mathbb{S}$.
3) Calculate $\nabla_{\mathbb{S}} F(A_t)$ as described above.
4) Generate samples $w_{ij}(t) \in \mathbb{R}$, $1 \le i \le m$, $1 \le j \le k$, from the standard normal distribution and construct the tangent vector

$$f_t = \delta \, \nabla_{\mathbb{S}} F(A_t) + \sqrt{2\delta T_t} \, \prod\nolimits_{A_t} \left(\sum_{i,j} w_{ij}(t) E_{ji} \right)$$

where $\prod_{A_t}$ is orthogonal projection onto the tangent space of the sphere at A_t and $\sum_{i,j} w_{ij}(t) E_{ji}$ is the stochastic component.

5) Generate a candidate $B \in \mathbb{S}$ by moving a small step on the great circle of unit sphere according to

$$B = A_t \cos(\|f_t\|) + \frac{f_t}{\|f_t\|} \sin(\|f_t\|) \, .$$

where $\|f_t\|$ is approximate the length of arc.

6) Calculate $F(B)$, $F(A_t)$, and the increment $\Delta F = F(B) - F(A_t)$.
7) Accept B with probability $\min\{e^{\Delta F / T_t}, 1\}$. If B is accepted, set $A_{t+1} = B$. Else, $A_{t+1} = A_t$.
8) If $t \le N$, set $T_{t+1} = \gamma T_t$ and $t = t + 1$, and go to Step 3. Else, let $\hat{A} = A_t$ and stop.

Remark. In our experiments, we usually initialize A with an orthogonal projection onto a subspace obtained from PCA or linear discriminant analysis.

V. Image Classification

Targeting applications to image retrieval, we first report the results of several image categorization experiments with the same subset of the Corel database used in [9], [5]. The data consists of 10 semantic categories with 100 images each; we refer to this data set as Corel-1000. The specific categories are: (1) African people and villages; (2) beach scenes; (3) buildings; (4) buses; (5) dinosaurs; (6) elephants; (7) flowers; (8) horses; (9) mountains and glaciers; (10) food. Sample images from some categories are shown in Figure 2 to illustrate the within-class variations observed in the data. In each experiment, we placed an equal number of images from each class in the training set and used the remaining ones as query images to be indexed by the nearest neighbor classifier (based on the cross-correlation metric (2)) applied to a reduced feature learned with dimension reduction algorithm of Section III. For a fair comparison with [5], an image is represented by an SH-feature vector $h(I, \mathcal{F})$ of dimension 165 obtained from the 11-bin histograms associated with 5 filters (including intensity, Gabor, Laplacian, Gradxx and Gradyy) applied to the R, G, and B channels and the dimension was reduced from $m = 165$ to $k = 12$. Figure 3 shows a plot of the categorization performance against the number T of training images. Categorization performance refers to the fraction of images indexed correctly using all $1,000 - T$ images outside the training set as queries. Compared to the categorization performance listed in table I [5] , the categorization performance is considerably improved.

TABLE I

Results of categorization experiments with OFA via Euclidean metrics. T is the number of training images.

T	Categorization Performance
200	71.7%
400	84.5%
600	85.5%

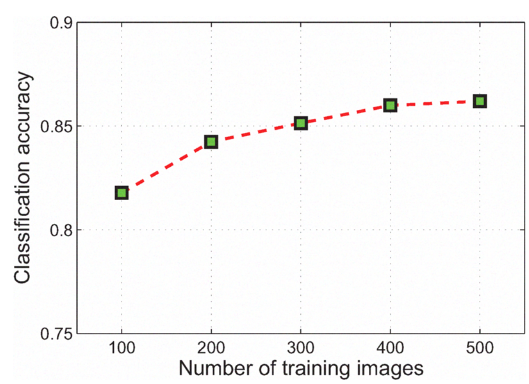


Fig. 3. Classification performance × size of training set.

VI. Image Retrieval

We now use the low-dimensionl representation learned for optimal image categorization to retrieve images from the database. We first remark that the classifier was optimized to categorize images, but not necessarily to rank matches to a query image correctly. Thus, instead of simply ranking retrievals using distance in reduced feature space, we exploit the strengths of the image categorization strategy in a more essential way.

We first index all images in the database using an optimal low-dimensional representation learned with the dimension reduction algorithm and the NN classifier based on the cross-correlation metric. Given a query image I and a positive

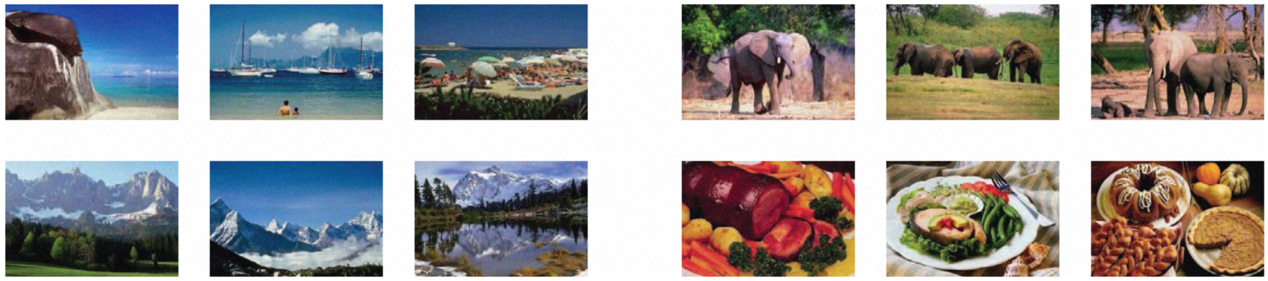


Fig. 2. Samples from 4 of ten classes of the Corel-1000 data set.

integer ℓ, the goal is to retrieve a ranked list of ℓ images from the database. Let T_c, $1 \leqslant c \leqslant K$, denote the set of training images in class c. Calculate the distances $d(I; T_c)$ from I to T_c in reduced feature space and rank the classes according to these distances. We retrieve images as follows: (i) select as many images as possible from the top ranked class; (ii) once that class is exhausted, we proceed similarly with the class ranked second and iterate the procedure until ℓ images are obtained. Within each class, the images are ranked using the distance d, defined in (2), applied to the reduced SH-features of the images. With this retrieval strategy, even if the within-class variation is large, we are likely to first retrieve more images from the correct class provided that the image classification step is accurate.

We use the same performance metrics adopted in [9] to compare the retrieval results objectively. The precision for the top ℓ returns is defined as n_ℓ/ℓ, where n_ℓ is the number of correct matches. The weighted precision for a query image I is

$$p(I) = \frac{1}{100} \sum_{\ell=1}^{100} \frac{n_\ell}{\ell}. \tag{8}$$

For each image I, rank order all 1,000 images in the database, as described above. The average rank $r(I)$ is the mean value of the ranks of all images that belong to the same class as I. The average values

$$\bar{p}_i = \frac{1}{100} \sum_{I \in C_i} p(I) \quad \text{and} \quad \bar{r}_i = \frac{1}{100} \sum_{I \in C_i} r(I) \tag{9}$$

of the weighted precision and average rank within each class C_i, $1 \leqslant i \leqslant 10$ will be used to quantify retrieval accuracy. A high performance retrieval system yields high mean precision and low mean rank.

A. *Experimental Results*

We report the results of retrieval experiments with the Corel-1000 dataset and a spectral representation based on 11-bin histograms associated with 5 filters applied to the R,G,B channels, as described in Section V. Since each class contains 100 images, the maximum possible number of matches to a query image is 100. We compare our results with those obtained with the retrieval system SIMPLIcity [9], color histograms with the earth mover's distance (EMD) investigated in [7], and OFA-400 [5] with OFA via Euclidean metrics. The results for SIMPLIcity and color histograms have been reported in [9]. We use the notation CC-400 to indicate our method based on cross-correlation metrics with 400 training images. For an objective comparison, we only use query images that are part of the database and we place 40 images from each class in the training set and reduce the dimension from 165 to 12. We calculated the average values $\bar{p}_i$ and $\bar{r}_i$ of the weighted precision and average rank, as defined in (9). The plots shown in Figure 4 show a moderate improvement in retrieval performance. The cross correlation metrics and Euclidean metrics are related. The cross correlation metric between two feature vectors is the radian θ while the Euclidean metric between the two feature vectors is $2\sin(\frac{\theta}{2})$. When θ is small, cross correlation metrics and Euclidean metrics are very close. The performance function defined in equation 5 via cross correlation metrics can be very different from that via Euclidean metrics when transformed feature vectors are far away seperated. To show the consistency of the results reported in Figure 4 for retrieval based on dimension reduction with cross-correlation metrics, we carried out 10 experiments with 400 randomly selected training images and the mean values of the average weighted precision and average rank within each class are shown in Figures 5 and 6. The performance is stable and consistent compared with the results shown in Figure 4. To illustrate

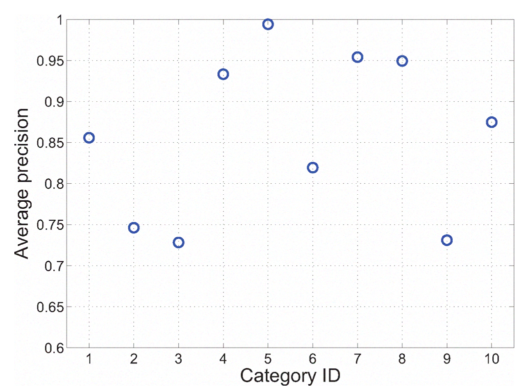
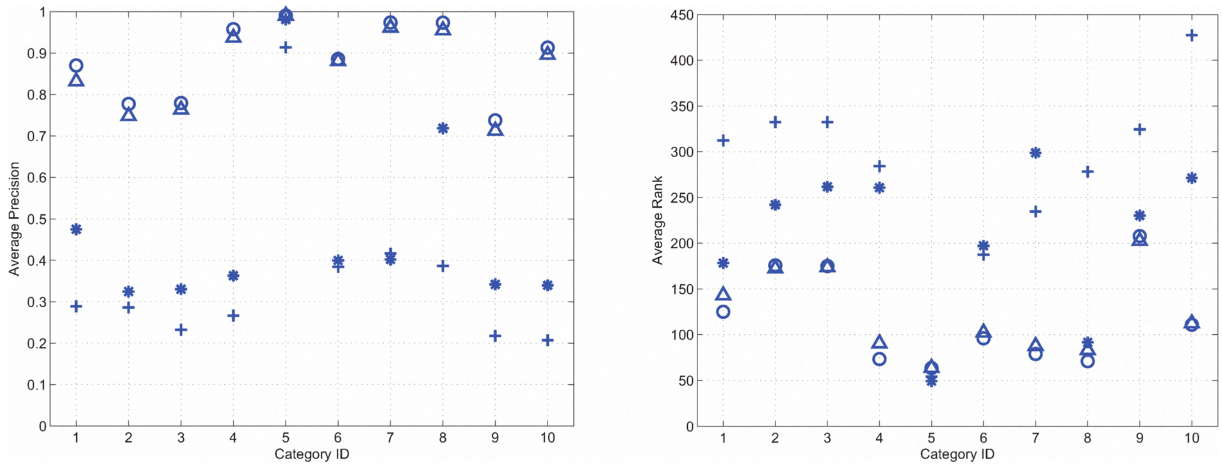


Fig. 5. Average precision over the entire data set × size of training set.

the dependence of the retrieval performance on the size of training set, Figures 7 and 8 show the mean values of the weighted precision and average rank over the entire data set as a function of the number of training images.



(a) (b)

Fig. 4. Mean values over 10 experiments of the (a) average precision and (b) average rank within each class. The methods are labeled as follows: (∘) CC-400; (∗) SIMPLIcity; (△) OFA-400; (+) color histogram.

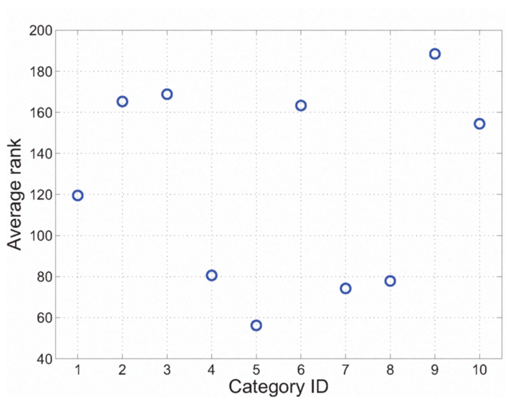


Fig. 6. Average rank over the entire data set × size of training set.

Fig. 7. Mean value of the average precision over the entire data set × size of training set. Retrieval based on dimension reduction with cross-correlation metrics.

We further quantify retrieval performance with a precision-recall curve. For an image I and a positive integer ℓ, let m_ℓ be the number of matching images among the top ℓ returns. Define

$$p_\ell(I) = \frac{m_\ell(I)}{\ell} \quad \text{and} \quad r_\ell(I) = \frac{m_\ell(I)}{100}, \tag{10}$$

which are the precision and recall rates for ℓ returns for image I. The average precision and average recall for the top ℓ returns are defined as

$$p_\ell = \frac{\sum_I p_\ell(I)}{1000} \quad \text{and} \quad r_\ell = \frac{\sum_I r_\ell(I)}{1000}, \tag{11}$$

respectively. Here, the sum is taken over all 1,000 images in the database. For a perfect retrieval system, $p_\ell = 1$, for $1 \leqslant \ell \leqslant 100$, and gradually decays to $p_{1000} = 0.1$. Similarly $r_\ell = 1$, for $\ell \geq 100$, decaying to $r_1 = 0.01$.

The average-precision-recall plots for a 12-dimensional representation obtained with dimension reduction are shown in Figure 9 for 100, 200, 300, 400 and 500 training images. It is obvious that the curve is approaching to the ideal situation as the number of training images increases.

VII. Conclusion

We represented images using histograms of their spectral components for content-based image categorization and retrieval. A learning technique was developed to reduce the dimension of the representation and optimize the discriminative ability of the nearest-neighbor classifier based on metrics derived from cross-correlation of histograms. Several experiments were carried out and the results indicate a significant improvement in retrieval performance over a number of existing retrieval systems and moderate improvement over OFA via Euclidean metrics. In the future we will do experiments on other image datasets of bigger variance to further explore the potential of cross correlation OFA. Learning based on local spectral features will also be investigated to mitigate the negative effects of the background on retrieval performance.

VIII. Acknowledgements

This work was supported in part by the National Science Foundation grant CCF-0514743 and by the National Institutes of Health through the NIH Roadmap for Medical

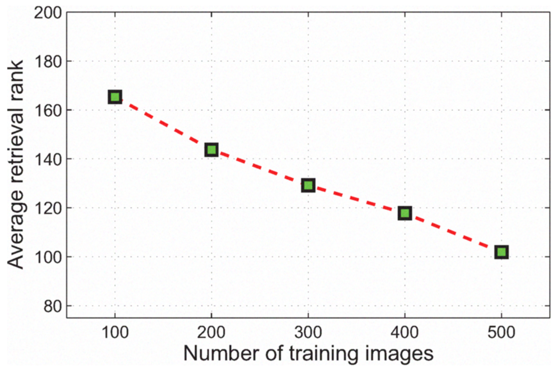


Fig. 8. Mean value of the average rank over the entire data set × size of training set. Retrieval based on dimension reduction with cross-correlation metrics.

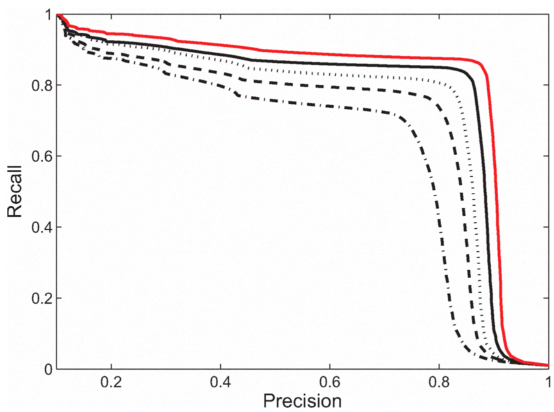


Fig. 9. Corel-1000: plots of average-precision × average-recall for 100(-.), 200(–), 300(..), 400(black solid line) and 500(red solid line) training images.

Research, Grant U54 RR021813. Information on the National Centers for Biomedical Computing can be obtained from http://nihroadmap.nih.gov/bioinformatics.

References

[1] C. Carson, M. Thomas, S. Belongie, J. Hellerstein, and J. Malik. "Blobworld: a system for region-based image indexing and retrieval", In *Proc. Visual Information Systems*, pp. 509-516, 1999.

[2] S. Hoi, W. Liu, M. Lyu, and W. Ma. "Learning distance metrics with contextual constraints for image retrieval", In *Proceedings of Computer Vision and Pattern Recognition*, pp. 2072-2078, 2006.

[3] X. Liu, A. Srivastava, and K. Gallivan. "Optimal linear representations of images for object recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 662-666, 2004.

[4] W. Mio, Y. Zhu, and X. Liu. "A learning approach to content-based image categorization and retrieval", In *2nd International Conference on Computer Vision Theory and Applications*, vol. 2, pp. 36-43, 2007.

[5] Y. Zhu, W. Mio and X. Liu. "Optimal Factor Analysis and Applications to Content-Based Image Retrieval", *Communications in Computer and Information Science*, vol. 21, pp. 164-176, 2008.

[6] J. Portilla and E. Simoncelli. "A parametric texture model based on joint statistics of complex wavelet coeficients", *International Journal of Computer Vision*, vol. 40, pp. 49-70, 2000.

[7] Y. Rubner, L. Guibas, and C. Tomasi. "The earth mover's distance, multi-dimensional scaling, and color-based image databases", *Proceedings of ARPA Image Understanding Workshop*, pp. 661-668, 1997.

[8] J. Smith and C. Li. "Image classification and querying using composite region templates", *Computer Vision and Image Understanding*, vol. 75, no. 9, pp. 165-174, 1999.

[9] J. Wang, J. Li, and G. Wiederhold. "SIMPLIcity: Semantics-sensitive integrated matching for picture libraries", *IEEE transcations on pattern analysis and machine intelligence*, vol. 23, no. 9, pp. 947-963, 2001.

[10] Y. Wu, S. Zhu, and X. Liu. "Equivalence of Julesz ensembles and FRAME models", *International Journal of Computer Vision*, vol. 38, pp. 247-265, 2000.

[11] P. Yin, B. Bhanu, K. Chang, and A. Dong. "Integrating relevance feedback techniques for image retrieval using reinforcement learning", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1536-1551, 2005.

[12] S. Zhu, Y. Wu, and D. Mumford. "Filters, random fields and maximum entropy (FRAME)", *International Journal of Computer Vision*, vol. 27, pp. 1-20, 1998.