

Image Retrieval Based on Intrinsic Spectral Histogram Representation

Yuhua Zhu, Xiuwen Liu, and Washington Mio

Abstract—The spectral histogram features are not invariant to images' scale transformation. We investigate in the technique of scale-invariant feature extraction. An approach is proposed to get the characteristic scales based on the reliable keypoints which are detected as local extrema in combination of normalized derivatives. Making use of characteristic scale of image content, which reflects characteristic length of a corresponding image structure, we are able to contribute in eliminating the effect of image transformation.

In our content based image retrieval process, images are firstly resized by the characteristic scale and then represented based on the statistics of their spectral components and a linear dimension reduction technique optimizing class differentiation with respect to cross-correlation metrics of spectral histograms. Our retrieval consists of a preliminary classification step to index images in dataset and a following step of class by class retrieval. Experiments are performed on the Corel database and the outcome is compared with those of some existing work.

I. INTRODUCTION

Along with the availability of large collections of images and videos, organizing and retrieving relevant images based on their contents has become a challenging problem in many application areas. The performance of a content based image retrieval (CBIR) system highly relies on effective and efficient features. In prior work, spectral-histograms associated with a bank of filters were proposed and proved to be effective in representing images' characteristics [7]. However, the spectral-histogram is sensitive to image scales and consequently affects the retrieval results. For example, if an image from corel-1000 dataset is scaled down to 0.5×0.5 of its original size, and then used as query image, the retrieval result will be worse than query by the image with original size. In this paper we introduce a method making use of characteristic scale to enhance the robustness to images' scale transformation, in other words, to make spectral histogram intrinsic. The scale invariant approach involved in this method is inspired by Lowe's research [3], [4], in which scale-invariant keypoints are selected to extracting distinctive features for image matching. It is proved that the selected points are robust to scale, rotation and translation. In our method, the keypoints are detected as local extrema in the scale space pyramid built with difference of Gaussian (DoG) filters. And then the characteristic scale of a certain image is achieved by averaging the characteristic scales on some of these keypoints which are identified as reliable. Based on the ratios of average characteristic scale among all the members

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).

in dataset to every one of them, the images in the dataset are rescaled. The spectral histograms computed from the resized dataset are intrinsic representation of images, which are used in categorization and retrieval.

As described in [6], an Euclidean representation of histograms of spectral components is used for image classification and retrieval. In order to maximize the effectiveness of the spectral histogram features and reduce the complexity of training and retrieval, a dimension reduction technique called Optimal Factor Analysis (OFA) is used to find the optimal combination of features for categorization. The technical details and application of OFA are reported in [7]. In this paper, a variant of OFA based on cross-correlation is used for both classification and retrieval.

This paper is arranged as follows: In Section II, the features and metrics used in classification and retrieval are described. Description of the theory of characteristic scale of image content and its implementation is also given in this section. We present the dimension reduction techniques in Section III and the numerical aspects of dimension reduction based on stochastic optimization in Section IV. In Sections V and VI, by employing the methods developed to image categorization and retrieval, we discuss the results of several experiments and make a comparison among them.

II. INTRINSIC IMAGE REPRESENTATION AND METRICS

Let I be a gray-scale image and F a convolution filter. Then I_F is the spectral component of I associated with F , which is obtained through the convolution of I and F as the summation taken over all the pixel locations

$$I_F(p) = F * I(p) = \sum_q F(q)I(p - q), \quad (1)$$

where $I_F(p)$ is the spectral component at pixel p . If I is a color image, the filter is applied to its R,G,B channels. For a given set of bins, $h(I, F)$ is used to denote the corresponding histogram of I_F . The SH-feature $h(I, F)$ can be considered as a vector in \mathbb{R}^b if the number of bins is b . Figure 1 illustrates the process of obtaining SH-features. 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. Frame (c) shows an 11-bin histogram of the filtered image.

Let $\mathcal{F} = \{F_1, \dots, F_r\}$ be a bank of filters, the SH-features $h(I, F_i)$ $1 \leq i \leq r$ are combined into a single m -vector

$$h(I, \mathcal{F}) = (h(I, F_1), \dots, h(I, F_r)), \quad (2)$$

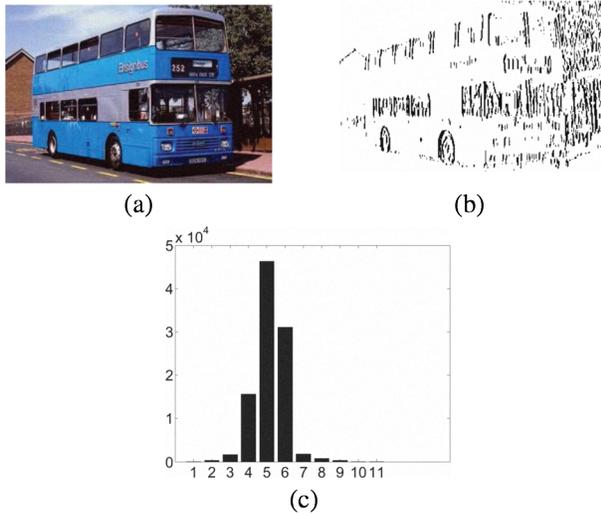


Fig. 1. (a) A color image; (b) Its green channel response to a Gabor filter; (c) 11-bin histogram of the filtered image.

where $m = rb$. $m = 3rb$ for a color image. The SH-features were used with dimension reduction techniques for content-based image classification and retrieval via usual Euclidean metric in [7]. In addition to feature representation, the choice of metric is also important to image classification. Here in this paper we use cross correlation metric.

A. Cross-Correlation Metrics

$x = (x_1, \dots, x_b)$ and $y = (y_1, \dots, y_b)$ are the given 2 vectors whose mean values are denoted by $\hat{x} = \sum x_i/n$ and $\hat{y} = \sum y_i/n$ respectively. $\bar{x} = x - \hat{x}$ and $\bar{y} = y - \hat{y}$ represent 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 value of the angle between \bar{x} and \bar{y} . The metric d can be then defined in

$$d(x, y) = \arccos \left(\frac{\bar{x}}{\|\bar{x}\|} \cdot \frac{\bar{y}}{\|\bar{y}\|} \right). \quad (3)$$

Given 2 images I_1 and I_2 and a bank of filters \mathcal{F} , the similarity degree between the 2 images can be estimated by the pseudo cross-correlation metric of the SH-features associated with \mathcal{F}

$$\delta_{\mathcal{F}}(I_1, I_2) = d(h(I_1, \mathcal{F}), h(I_2, \mathcal{F})). \quad (4)$$

When the dimension reduction machine learning algorithms are applied onto SH-features, expression 3 is used instead of expression 4.

B. Characteristic Scale of Image Content

As described in Section I, spectral histogram is sensitive to scale transformation. In order to make spectral histogram more intrinsic, we propose a method making use of characteristic scale.

Much work has been done for extracting scale invariant features [5]. These kinds of techniques have the assumption that the scale change is the same in every direction, which is similar to resizing an image with same scale along both

horizontal and vertical direction. In [4], an algorithm is proposed based on local 3D extrema in scale-space pyramid built with DoG filters. The scale space L of an image is defined as $L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)$, where G is a variable-scale Gaussian and I is an image. The DoG representation is obtained from the difference of two successive smoothed images with different scale

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) \\ &= L(x, y, k\sigma) - L(x, y, \sigma). \end{aligned}$$

where k represents the index of layer in the scale-space. The local extrema in the scale-space pyramid are detected as keypoints by comparing sample point to 26 neighbors in its current, above and below smoothed images. Characteristic scale of a keypoint is then selected as extremum over DoG-scale curve.

The idea of selecting characteristic scale was studied by Lindeberg in 1998 [1]. Characteristic scale, for which a function is given to achieve local extremum over scales from combination of normalized derivatives, reflects a characteristic length of a corresponding image structure. The selected characteristic scale is characteristic because it gives a measure of scale where the similarity between local image structure and feature detection operator can be maximized.

In our experiment, we use the same method as in [4] to get keypoints, except that we have 1 octave and 22 layers of DoG. A scale selection operator shows its response over a set of scales σ_n upon a given point in Fig. 2. The local extremum of DoG corresponds to the characteristic scale. There are several extrema among which we select the first one. According to Fig. 2 we can see obviously that characteristic scale is independent of image size and only relates to the content or structure of image. A conclusion is drawn based on the observation that the ratio of scales where the extrema are detected for corresponding points represents the scale factor between the point neighborhoods. The characteristic scale of an image is computed as the average of characteristic scale of reliable keypoints detected. We define reliable keypoints as those keypoints detected in pyramid of both original size image and corresponding downsampled image, e.g., the point of left image in Fig. 2.

After obtaining the characteristic scale of every image, the average characteristic scale among the whole dataset is computed and all the images in the dataset are resized according to ratio between the average and their own characteristic scale. Spectral Histogram features are then extracted from the resized images for future retrieval.

III. OPTIMAL FACTOR ANALYSIS

The problem of how to reduce the features' dimension and maximize the discriminative ability of classifier is proposed after obtaining the intrinsic representation of SH features. A machine learning technique, Optimal Factor Analysis (OFA), is developed in [6] to learn an optimal linear mapping of SH features that maximizes the discrimination based on K-nearest-neighbor classifier via Euclidean metrics. The

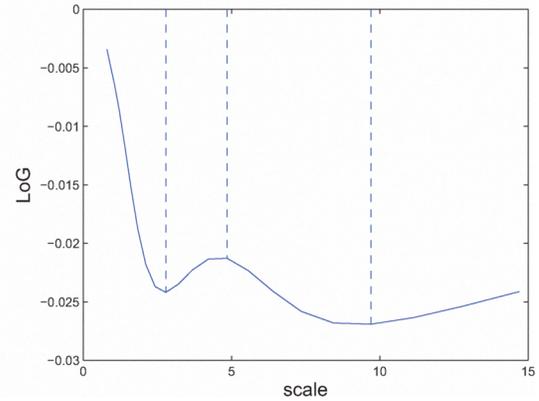
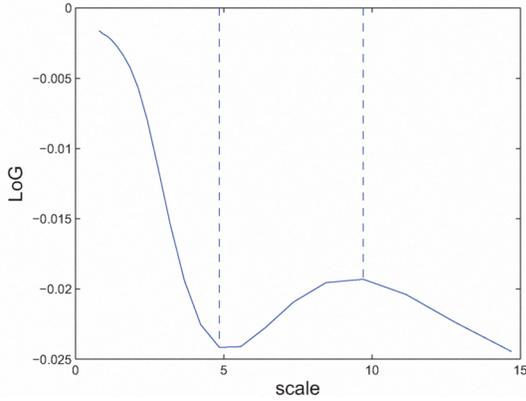


Fig. 2. Example of characteristic scales. Top left image is 2×2 of the top right image. The bottom row shows the response $D(x, y, \sigma_n)$ over scales. The selected characteristic scales are 4.8503 and 2.7858 for left and right image respectively. The ratio of the two scales corresponds to the scale factor 2 of the two images. The radius of the displayed round region in the 2 images of top row is equal to 4 times the characteristic scale.

technical details and application of OFA are revised in [7]. In this paper, a variant version of OFA with respect to cross-correlation metrics is adopted. The general cross correlation d is defined by Equation 3 in Section II-A to give the pseudo metric.

We suppose that the feature vectors' dimension is m and there are K different categories of images in the dataset. Let $x_1^c, \dots, x_{t_c}^c$ be a collection of training feature vectors belonging to an arbitrary class c , $1 \leq c \leq K$, and A be a $k \times m$ transformation matrix where $k \ll m$. Then Ax_i^c is a $\mathbb{R}^m \rightarrow \mathbb{R}^k$ linear mapping of x_i^c . In order to quantify how well the nearest neighbor (NN) classifier functions on the dimension reduced feature vectors with respect to cross-correlation metric, we define a value

$$\rho(x_i^c; A) = \frac{\min_{c \neq b, j} \|d(Ax_i^c, Ax_j^b)\|^2}{\min_{j \neq i} \|d(Ax_i^c, Ax_j^c)\|^2 + \epsilon} \quad (5)$$

where ϵ is a very small value used to avoid vanishing denominators. Since the denominator represents the minimum within-class distance and the numerator indicates the minimum between-class distance, $\rho(x_i^c; A)$ identifies how well the element x_i^c belonging to its class. The larger the value of $\rho(x_i^c; A)$ is, the closer x_i^c transformed by A lies to a training element in the class it belongs to than to elements in other classes. $\rho(x_i^c; A) \approx 1$ indicates a transitional behavior between correct and incorrect decisions by the NN classifier. Equation 5 can also be modified to measure how well the

KNN classifier performs. In order to avoid the problem of A mapping any feature vector in training set to zero, we can slightly perturb A .

Now the aim to find a linear mapping optimizing the discriminative ability of NN is converted into choosing a transformation A that maximizes the average value of $\rho(x_i^c; A)$ over the training set. To further quantify the average value of $\rho(x_i^c; A)$ over the whole training set, we define the performance function

$$F(A) = \frac{1}{K} \sum_{c=1}^K \left(\frac{1}{t_c} \sum_{i=1}^{t_c} \rho(x_i^c; A) \right). \quad (6)$$

The decisions made by the nearest neighbor classifier in the reduced feature space are insensitive to the scaling of matrix A . This is indicated by the fact that the performance function F is almost scale invariant; that is, $F(A) \approx F(\gamma A)$, for $\gamma > 0$. Although the equality does not hold exactly because $\epsilon > 0$, in practice ϵ is negligible. Therefore we can treat F as a scale-invariant function defined on $\mathbb{R}^{k \times m}$ which can be optimized over linear map A of unit Frobenius norm. Let

$$\mathbb{S} = \{A \in \mathbb{R}^{k \times m} : \|A\|^2 = \text{tr}(AA^T) = 1\} \quad (7)$$

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

$$\hat{A} = \underset{A \in \mathbb{S}}{\text{argmax}} F(A). \quad (8)$$

IV. NUMERICAL OPTIMIZATION

In order to estimate \hat{A} , a stochastic gradient approach with simulated annealing is adopted. The optimization strategy is similar to that used in OCA by Liu et al. [2]. The difference between the two is that the search of OFA is implemented on a sphere while that of OCA is implemented on a Grassmann manifold.

Let A be on \mathbb{S} which is the unit sphere, 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. The calculation of $\nabla_{\mathbb{R}^{k \times m}}F(A)$ is based on estimations of the partial derivatives. For $1 \leq i \leq k, 1 \leq j \leq m$, 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} , $\nabla_{\mathbb{S}}F$ can be estimated by the summation of partial derivatives along all the directions.

The next goal is to add a stochastic component to the deterministic gradient field $\nabla_{\mathbb{S}}F$ on \mathbb{S} . In order to simplify the calculation, we employ a stochastic component on $\mathbb{R}^{k \times m}$ projected onto the tangent space to \mathbb{S} at A . We define the notation $\Pi_A: \mathbb{R}^{m \times k} \rightarrow T_A\mathbb{S}$ for the orthogonal projection onto the tangent space of sphere at A . The stochastic gradient search procedure is listed as following:

- 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 $w_{ij}(t) \in \mathbb{R}, 1 \leq i \leq m, 1 \leq j \leq k$, from the standard normal distribution and construct the tangent vector

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

where Π_{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 $t \leq N$, 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

We evaluated our method for image retrieval applications, using the Corel dataset used in [9], [7]. The dataset contains 10 object categories with 100 images in each category. This dataset is referred as Corel-1000 in this paper. The categories are listed in Table I. Example images from some categories are shown in Figure 3. It can be observed that there are significant within-class variations in the images.

TABLE I
IMAGE CATEGORIES IN COREL-1000.

1	African People & Villages
2	Beach
3	Buildings
4	Buses
5	Dinosaurs
6	Elephants
7	Flowers
8	Horses
9	Mountains & Glaciers
10	Food

In our experiments, the images from each class are evenly split into two disjoint sets. One of the sets is used as the training set, and the other set is used as query image set. Each query image is indexed by the nearest neighbor classifier (based on the cross-correlation metric (3)) applied to the reduced feature learned from the dimension reduction algorithm presented in Section III. For a fair comparison with [7], an image has 5 filters (including intensity, Gabor, Laplacian, Gradxx and Grady) of 11 bins each applied to its R, G, and B channels to form a representative SH-feature vector $h(I, \mathcal{F})$ of dimension 165. The dimension is then reduced from $m = 165$ to $k = 12$. Figure 4 shows the plot of the categorization performance against the size of the training set T . The performance is evaluated by the proportion of query images that are correctly indexed by the classifier learned on the T training images.

VI. IMAGE RETRIEVAL

We use the low-dimensional representation learned for optimal image categorization to retrieve images from the dataset. It should be noted that the classifier is optimized for categorizing images, but not necessarily optimized for ranking the images matched to a query image. Thus, instead of ranking the retrieval results using the distances in the reduced feature space, we evaluate the image categorization strategy in a more theoretical way.

First, we index all the image in the dataset using the optimal low-dimensional representations learned by the dimension reduction algorithm and the NN classifier based on the cross-validation metric. The goal of retrieval task is to retrieve a ranked list of ℓ ($\ell > 0$) images from the database with a given query image I . Let $T_c, 1 \leq c \leq K$, denote the

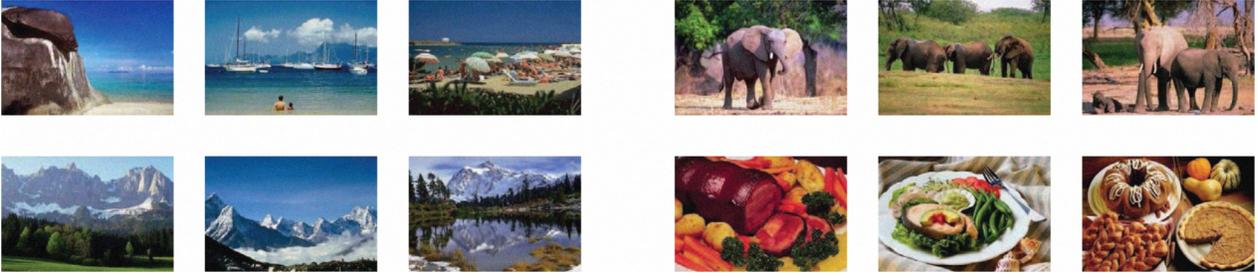


Fig. 3. Samples from 4 of ten classes of the Corel-1000 data set.

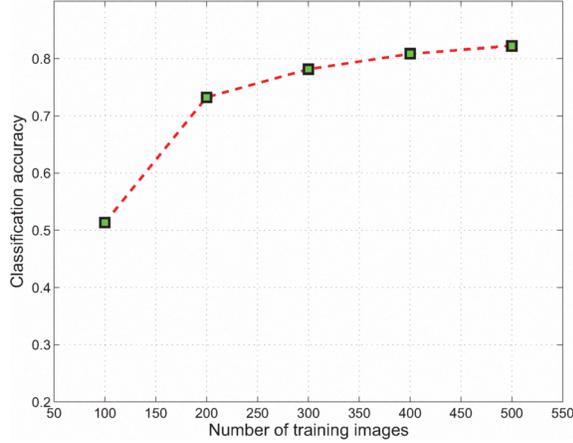


Fig. 4. Classification performance \times size of training set.

set of training images in class c . Then the distances $d(I; T_c)$ from I to T_c in reduced feature space are calculated and classes are ranked 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. The images within each class are ranked using by the cross correlation distance d , defined in (3), applied to the reduced SH-features of the images. We are to first retrieve more images from the correct class if the image classification step is accurate.

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

$$p(I) = \frac{1}{100} \sum_{\ell=1}^{100} \frac{n_\ell}{\ell} \quad (9)$$

as there are 100 elements in each category. We rank the order of all 1,000 images in the database for each image I 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

$$\bar{p}_i = \frac{1}{100} \sum_{I \in C_i} p(I) \quad \text{and} \quad \bar{r}_i = \frac{1}{100} \sum_{I \in C_i} r(I) \quad (10)$$

of the weighted precision and average rank within each class C_i , $1 \leq i \leq 10$ will be used to quantify retrieval accuracy. The higher mean precision and lower mean rank is, the higher performance retrieval system yields.

A. Experimental Results

We extract the intrinsic SH-histogram from the Corel-1000 dataset which is scaled based on representative scale calculated by the criterion described in Section II. Since each category contains 100 images, the maximum possible number of matches to a query image is 100. We compare our results to those obtained by the SIMPLiCity retrieval system in [9], the color histograms with earth movers distance (EMD) investigated in [8], and OFA-400 in [7]. The results of SIMPLiCity system and the color histograms have been reported in [9]. Our method is denoted as CC-400 intrinsic spectral histogram (method based on cross-correlation metrics and intrinsic spectral histogram with 400 training images). For a fair 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 (10). The plots shown in Figure 5 show a significant improvement in retrieval performance over many existing work except OFA via Euclidean metrics. This inferiority of performance can be explained by the possible effect between cross correlation metrics and the scaled intrinsic spectral histogram.

The retrieval performance is further quantified with a precision-recall curve. Let m_ℓ be the number of matching images among the top ℓ returns for an image I and a positive integer ℓ ,

$$p_\ell(I) = \frac{m_\ell(I)}{\ell} \quad \text{and} \quad r_\ell(I) = \frac{m_\ell(I)}{100}, \quad (11)$$

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

$$p_\ell = \frac{\sum_I p_\ell(I)}{1000} \quad \text{and} \quad r_\ell = \frac{\sum_I r_\ell(I)}{1000}, \quad (12)$$

respectively. In our experiment the sum is taken over all 1,000 images of the database. For an absolutely ideal situation, $p_\ell = 1$, for $1 \leq \ell \leq 100$, and gradually decays to $p_{1000} = 0.1$. Similarly $r_\ell = 1$, for $\ell \geq 100$, decaying to $r_1 = 0.01$.

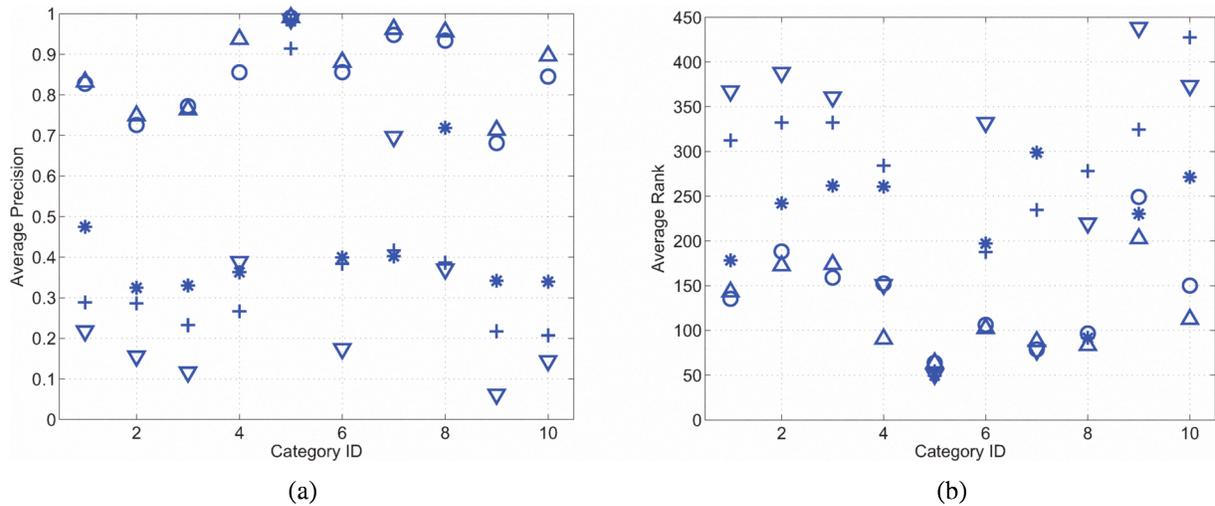


Fig. 5. Mean values over 10 experiments of the (a) average precision and (b) average rank within each class. The methods are labeled as follows: (o) CC-400 intrinsic spectral histogram; (∇) Spectral histogram via Euclidean metrics; (*) SIMPLiCity; (△) OFA-400 via Euclidean metrics; (+) color histogram.

The average-precision-recall plots for a 12-dimensional representation obtained with dimension reduction are shown in Figure 6 for 100, 200, 300, 400 and 500 training images.

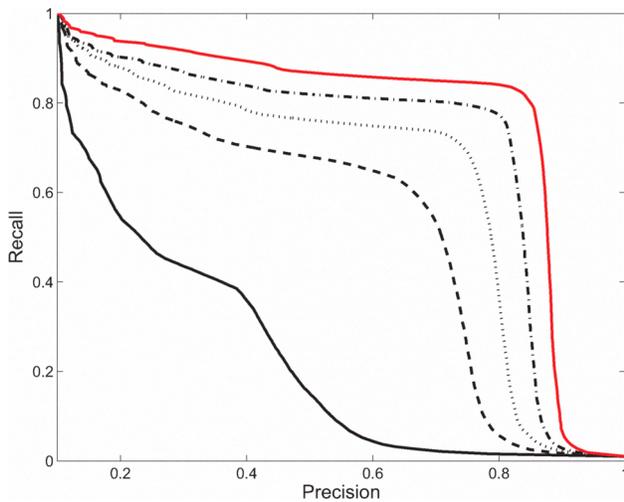


Fig. 6. Corel-1000: plots of average-precision \times average-recall for 100(black solid line), 200(- -), 300(· · ·), 400(- · -) and 500(red solid line) training images.

VII. CONCLUSIONS

In this paper we propose an approach using characteristic scale of image content to get rid of effect of image scaling. Before the extraction of images' histogram of spectral components for content-based image categorization and retrieval, the images are rescaled according to the ratio between average characteristic scale of the whole dataset to their own scale. OFA learning technique is employed to reduce the dimension of the representation and optimize the discriminativeness of the nearest-neighbor classifier based on metrics

derived from cross-correlation of histograms. Experiments are performed and the benefits of intrinsic spectral histogram are discussed. In the future we will do experiments on intrinsic spectral histogram with OFA via Euclidean metrics to have more direct comparison.

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

REFERENCES

- [1] T. Lindeberg. "Feature Detection with Automatic Scale Selection", *International Journal of Computer Vision*, vol. 30, no. 2, pp. 79-116, 1998.
- [2] X. Liu, A. Srivastava, and K. Gallivan. "Optimal Linear Representations of Images for Object Recognition", *IEEE transactions on pattern analysis and machine intelligence*, vol. 1, pp. 662-666, 2004.
- [3] D. Lowe. "Object recognition from local scale-invariant features", *International Conference on Computer Vision*, vol. 2, pp. 1150-1157, 1999.
- [4] D. Lowe. "Distinctive Image Features from Scale-Invariant Keypoints", *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110, 2004.
- [5] K. Mikolajczyk and C. Schmid. "Scale & Affine Invariant Interest Point Detectors", *International Journal of Computer Vision*, vol. 60, no. 1, pp. 63-86, 2004.
- [6] W. Mio, Y. Zhu, and X. Liu. "A learning approach to content-based image categorization and retrieval", *2nd International Conference on Computer Vision Theory and Applications*, Vol. 22, pp. 36-43, 2007.
- [7] 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.
- [8] Y. Rubner, L. Guibas, and C. Tomasi. "The earth movers distance, multi-dimensional scaling, and color-based image retrieval". *Proc. ARPA Image Understanding Workshop*, pp. 661-668, 1997.
- [9] J. Wang, J. Li, and G. Wiederhold. "SIMPLiCity: Semantics-Sensitive Integrated Matching for Picture Libraries", *IEEE transactions on pattern analysis and machine intelligence*, vol. 23, pp. 947-963, 2001.