# In-Plane Rotation and Scale Invariant Clustering Using Dictionaries

Yi-Chen Chen, *Student Member, IEEE*, Challa S. Sastry, *Member, IEEE*, Vishal M. Patel, *Member, IEEE*, P. Jonathon Phillips, *Fellow, IEEE*, and Rama Chellappa, *Fellow, IEEE*

***Abstract*— In this paper, we present an approach that simultaneously clusters images and learns dictionaries from the clusters. The method learns dictionaries and clusters images in the radon transform domain. The main feature of the proposed approach is that it provides both in-plane rotation and scale invariant clustering, which is useful in numerous applications, including content-based image retrieval (CBIR). We demonstrate the effectiveness of our rotation and scale invariant clustering method on a series of CBIR experiments. Experiments are performed on the Smithsonian isolated leaf, Kimia shape, and Brodatz texture datasets. Our method provides both good retrieval performance and greater robustness compared to standard Gabor-based and three state-of-the-art shape-based methods that have similar objectives.**

***Index Terms*— Clustering, content-based image retrieval (CBIR), dictionary learning, radon transform, rotation invariance, scale invariance.**

## I. Introduction

In RECENT years, sparse representation has emerged as a powerful tool for efficiently processing data in non-traditional ways. This is mainly due to the fact that signals and images of interest tend to enjoy the property of being sparse or compressible in an appropriate dictionary. In other words, they can be well approximated by a linear combination of a few elements (also known as atoms) of a dictionary. The dictionaries can be composed of wavelet or Fourier basis functions or can be learned from data. It has been observed that dictionaries learned directly from data provide better representation and hence improve the performance on many practical applications such as restoration and classification [1]–[4].

Dictionaries can be learned for both reconstruction and discrimination applications. In the late nineties, Etemad and Chellappa proposed a linear discriminant analysis (LDA) based basis selection and feature extraction algorithm for classification using wavelet packets [5], and Phillips proposed a dictionary method for face recognition [6]. Recently, algorithms for simultaneous sparse signal representation and discrimination have been proposed [2]–[14]. Additional techniques may be found within these references.

Dictionary learning techniques for unsupervised clustering have gained traction in recent years. In [15], a method for simultaneously learning a set of dictionaries that optimally represent each cluster was proposed. To improve the accuracy of sparse coding, this approach was later extended by adding a block incoherence term in their optimization problem [16]. Additional sparsity motivated subspace clustering methods include [17]–[19].

Invariance to rotation and scale are desirable in many practical applications. One important application is image classification and retrieval where one wants to classify or retrieve images having the same content but at different orientation and scale. For instance, in content based image retrieval (CBIR), images are retrieved from a database using features that best describe the orientation and scale of objects in the query image. Gabor filters have been used to extract features for retrieval and classification [20]. However, the chosen directions of Gabor filters may not correspond to the orientation of the content in the query image. Hence, a feature extraction method that is independent of orientation and scale in the image is desirable [21]. Wavelet-based methods have been proposed to achieve rotation invariance for image classification and retrieval [22], [23]. There have also been methods proposed to learn invariant dictionaries in the image domain [10], [24], [25]. Recently, a shift, scale and rotation invariant dictionary learning method for multivariate signals was proposed in [26]. Hierarchical dictionary learning methods for invariant classification have also been proposed in [13] and [27]. These methods learn a dictionary in a log-polar domain to be invariant to scale and rotation. A cellular neural network-based method for rotation invariant texture has also been proposed in [28].

Numerous descriptors have been proposed in the literature that are invariant to image transformations [29], [30]–[33]. A shape matching approach based on correspondences

Manuscript received July 6, 2012; revised January 15, 2013; accepted January 31, 2013. Date of publication February 8, 2013; date of current version March 29, 2013. The work of Y.-C. Chen and R. Chellappa was supported by a Cooperative Agreement from the National Institute of Standards and Technology under Grant 70NANB11H023. The work of V. M. Patel and R. Chellappa was supported by an ONR under Grant N00014-12-1-0124. The work of C. S. Sastry was supported by IIT Hyderabad and DST, India under Grant SR/FTP/ETA-054/2009. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Rebecca Willett.

Y.-C. Chen is with the Department of Electrical and Computer Engineering and the Center for Automation Research, UMIACS, University of Maryland, College Park, MD 20742 USA (e-mail: cheny08@umiacs.umd.edu).

C. S. Sastry is with the Department of Mathematics, Indian Institute of Technology Hyderabad, Hyderabad 500016, India (e-mail: csastry@iith.ac.in).

V. M. Patel is with the Center for Automation Research, UMIACS, University of Maryland, College Park, MD 20742 USA (e-mail: pvishalm@umd.edu).

P. J. Phillips is with the National Institute of Standards and Technology, Gaithersburg, MD 20899 USA (e-mail: jonathon@nist.gov).

R. Chellappa is with the Department of Electrical and Computer Engineering and the Center for Automation Research, UMIACS, University of Maryland, College Park, MD 20742 USA (e-mail: rama@umiacs.umd.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2013.2246178

between points on two shapes was proposed in [30]. This shape context descriptor essentially estimates the shape similarity and solves the correspondence problem. A shock graph-based feature extraction method that uses object silhouettes was proposed in [31]. The length of the shortest path within the shape boundary (called inner-distance) was used to build shape descriptors in [32]. These descriptors were shown to be robust to articulation in complicated shapes. In [33], a feature extraction method based on features that characterize the geometric relationships between each pair of images was proposed. This method was shown to be invariant to articulations and rigid transforms. Some of these methods are only shape-based and require the extraction of shape contour. These methods do not perform well on non-shape images such as textures.

In this paper, we present an in-plane rotation and scale invariant clustering approach (box 1∼3 in Fig. 1), which extends the dictionary learning and sparse representation framework (box 4, 5 in Fig. 1) for clustering and retrieval of images. Fig. 1 illustrates the overview of the proposed approach. Given a database of images $\{\mathbf{x}_j\}_{j=1}^{J}$ and the number of clusters $K$, our method uses the Radon transform to find scale and rotation-invariant features. It then uses sparse representation methods to simultaneously cluster the data and learn dictionaries for each cluster. One of the main features of our method is that it is effective for shape-based and certain texture-based images. We demonstrate the effectiveness of our approach in image retrieval experiments, where we report significant improvements in performance. A preliminary version of this work appeared in [34]. The following extensions have been made to [34]:

1. We propose a normalization method validated by a mathematical proof, to achieve scale invariance in the Radon domain.
2. We propose a method to obtain initial classes and class dictionaries in a deterministic way. We refer to our in-plane rotation and scale invariant clustering using dictionaries method as RSICD.
3. In the experiment section, we introduce both rotation and scale variations to shape-based databases (Smithsonian isolated leaf and Kimia) and compare RSICD with four other state-of-the-art methods (LBP [35], BAC [33], IDSC+DP [32] and SC [30]). We further demonstrate the effectiveness of RSICD on a texture-based database (Brodatz) by comparing it with three other texture-based methods (Modified Gabor [21], DC [15] and LBP [35]).

The organization of the paper is as follows. A method based on scale and rotation invariant features that are extracted using the Radon transform is detailed in Section II. Our simultaneous clustering and dictionary learning method is described in Section III. Experimental results are presented in Section IV. In Section V, we conclude the paper with a brief summary and future research directions.

## II. Radon-Based Rotation and Scale Invariance

In this section, we show how the Radon transform is used to achieve in-plane rotation and scale invariance (box 1 ∼ 3 in Fig. 1).

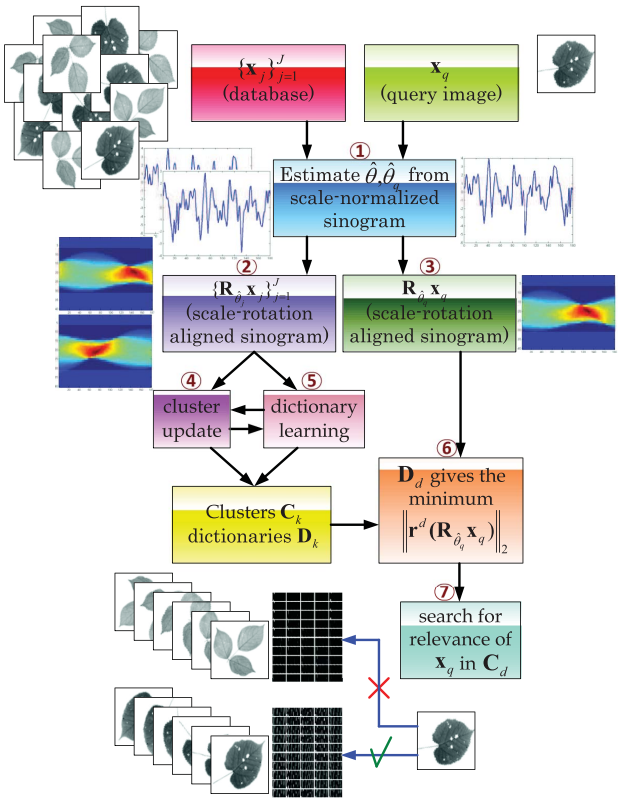



Fig. 1. Overview of the proposed simultaneous clustering and dictionary learning method.

### A. Estimating the Rotation Present in an Image

The Radon transform of a two variable function $x$ is defined as

$$R_\theta x(t) = \int_{-\infty}^{\infty} x(t\cos\theta - s\sin\theta, t\sin\theta + s\cos\theta)ds \quad (1)$$

where $(t, \theta) \in (-\infty, \infty) \times [0, \pi)$. Fig. 2(a) illustrates how the Radon transform is calculated. We use (1) to compute the value at any given point $(\theta, t)$ in the Radon domain by integrating along the line: $(u, v) = (t\cos\theta - s\sin\theta, t\sin\theta + s\cos\theta)$, $\forall s \in \mathbb{R}$. If $\tilde{x}$ is a rotated copy of $x$ by an angle $\hat{\theta}$, then a simple proof shows that their Radon transforms are related as

$$R_\theta \tilde{x}(t) = R_{\theta+\hat{\theta}} x(t) \quad \forall t, \theta. \quad (2)$$

For directional texture images, the principal orientation is roughly defined as the direction where there are more straight lines. The Radon transform can be used to detect linear trends in images. For general images, the principal orientation may be taken as the direction along which the Radon transform has the maximum variability. Let $\sigma_\theta \triangleq \mathbf{Var}_t[R_\theta x(t)]$ denote the variance with respect to $t$ of the Radon transform at $\theta$. In [36], $\sigma_\theta$ was found to be useful in estimating the principal orientation in an image. An important observation was that the Radon transform along $\hat{\theta}$ has larger variations with respect to $t$ and hence the variance $\sigma_{\hat{\theta}}$ is a local maximum along the $\theta$ axis. Based on the observation, one can estimate $\hat{\theta}$ of a given image

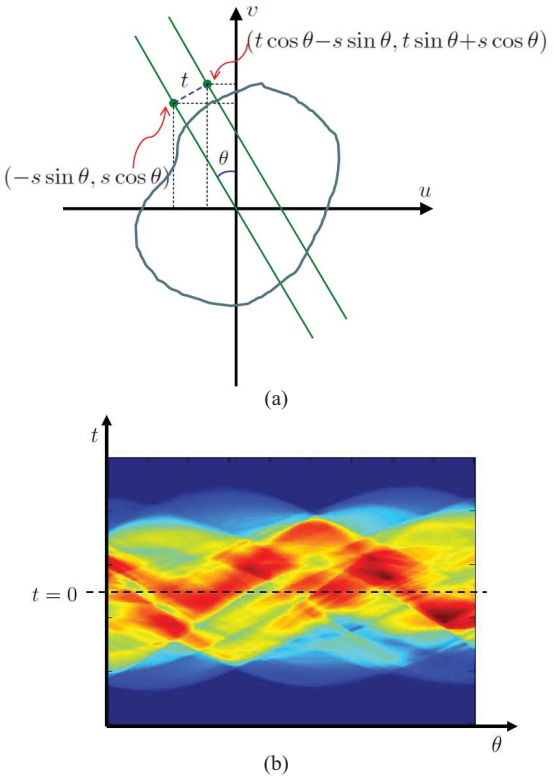



Fig. 2. (a) Illustration of how the Radon transform is calculated. Given any point $(u, v)$ in the image domain, we can express $u$ and $v$ as: $u = -s \sin\theta$, $v = s \cos\theta$ for some $s$ and $\theta$, where $s$ is the distance between $(u, v)$ and the origin; and $\theta$ is the angle between the positive vertical axis direction and the line passing through $(u, v)$ and the origin. As indicated, a $t$-translated point from $(u, v)$ is located at $(t \cos\theta - s \sin\theta, t \sin\theta + s \cos\theta)$. $t$ is the distance between the line that passes through $(u, v)$ and the origin, and the parallel line that passes through $(t \cos\theta - s \sin\theta, t \sin\theta + s \cos\theta)$. (b) In practice, the Radon transform of an image is represented as a matrix called sinogram, where the column indices correspond to discrete values of $\theta$ and row indices correspond to discrete values of $t$. $\theta$ and $t$ are the two continuous variables of $R_\theta x(t)$ given in (1).

$\tilde{x}$ from the following formula[1]

$$\hat{\theta} = \arg\min_{\theta} \left( \frac{d^2 \sigma_\theta}{d\theta^2} \right). \tag{3}$$

The global minimum of the second derivative of $\sigma_\theta$ is computed in order to locate the angle at which the change rate of the first derivative of $\sigma_\theta$ is the maximum, which represents the maximum number of line trends (i.e., along the principal orientation where the local maximum is the narrowest in shape, as indicated by Figs. 1(b) and (d) in [36]). Once the orientation is estimated, this estimate and (2) can be used to align the rotation in the Radon domain. Hence, we achieve rotation invariance.

In practice, the Radon transform of an image is represented as a matrix, called a *sinogram*. Fig. 2(b) gives the illustration of a sinogram. In a sinogram, column indices correspond to discrete values of $\theta$, while row indices correspond to discrete values of $t$, where $\theta$ and $t$ are the two continuous variables of

[1]In this paper, we apply this approach not only to directional texture images, but also to other isotropic textures (any direction is the principal orientation) and shape-based images.

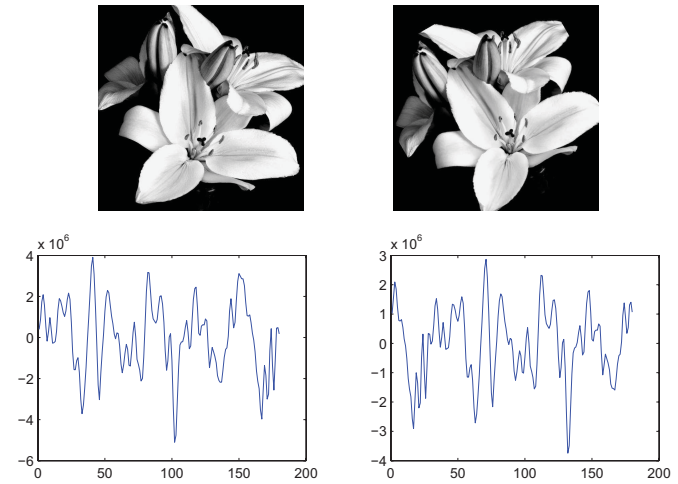



Fig. 3. For the rotated images present on the first row, the plots on the second row show their $\frac{d^2\sigma_\theta}{d\theta^2}$ (along the vertical axis) versus $\theta$ (along the horizontal axis). The second row plots indicate that the difference between the points of global minimum of both curves preserves the rotation present in the second image.

$R_\theta x(t)$ given in (1). $\sigma_\theta$ is the variance computed from values of the column that corresponds to angle $\theta$.

Fig. 3 illustrates how the sinogram is used to estimate the angle present in an image.[2] The second image shown on the first row of Fig. 3 is a rotated copy of the first image by 30°. The plots on the second row are the second derivatives of variances $\sigma_\theta$ (vertical axis) versus $\theta$ (horizontal axis) of their sinograms, where $\sigma_\theta$ is the variance over all entries in the column that corresponds to $\theta$. It may be noted that the difference between the points of global minima of both curves is 30°, coinciding with the rotation present in the second image. Consequently, the estimate presented in (3) is useful for estimating the presence of rotation in the images.

### B. Scale Invariance

Let $\bar{x}$ be a scaled copy of $x$ with the scaling factor $\xi$ such that $\bar{x}(u, v) = x(\xi u, \xi v)$. We can relate their Radon transforms as follows:

$$\begin{aligned} R_\theta \bar{x}(t) &= \int_{-\infty}^{\infty} \bar{x}(t \cos\theta - s \sin\theta, t \sin\theta + s \cos\theta) ds \\ &= \int_{-\infty}^{\infty} x(\xi t \cos\theta - \xi s \sin\theta, \xi t \sin\theta + \xi s \cos\theta) ds \\ &= \frac{1}{\xi} R_\theta x(\xi t). \end{aligned}$$

From the above equations, size scaling in the image domain results in scaling and normalization of the Radon transform. From this observation, scale invariance can be achieved through the following normalization in the Radon domain:

$$\frac{1}{M_{\bar{x}}} R_\theta \bar{x}\,(T_{\bar{x}} t) \tag{4}$$

[2]http://www.flowersofpictures.com/wp-content/uploads/2011/07/Lily-Flowers.jpg.

where

$$M_{\bar{x}} \triangleq \max_{t,\theta} |R_\theta \bar{x}(t)|$$

and

$$T_{\bar{x}} = \inf\{T | R_\theta \bar{x}(t) = 0 \ \forall |t| > T, \theta \in [0, 2\pi)\}.$$

Based on this formulation, one can derive the following result:

*1) Proposition 1:* Let $\mathcal{Y}$ be a set of functions related by different scales. For any pair $\bar{x}, x \in \mathcal{Y}$ related by $\bar{x}(u, v) = x(\xi u, \xi v)$ where $\xi > 0$, the following holds

$$\frac{1}{M_{\bar{x}}} R_\theta \bar{x}\left(T_{\bar{x}} t\right) = \frac{1}{M_x} R_\theta x\left(T_x t\right).$$

*Proof:* For any pair $\bar{x}, x \in \mathcal{Y}$ related by $\bar{x}(u, v) = x(\xi u, \xi v)$ where $\xi > 0$, we have

$$\begin{aligned} M_{\bar{x}} &= \max_{t,\theta} |R_\theta \bar{x}(t)| \\ &= \max_{t,\theta} \left|\tfrac{1}{\xi} R_\theta x(\xi t)\right| \\ &= \tfrac{1}{\xi} \max_{t,\theta} |R_\theta x(\xi t)| \\ &= \tfrac{1}{\xi} \max_{\tau,\theta} |R_\theta x(\tau)| (\text{let } \tau = \xi t) \\ &= \tfrac{1}{\xi} M_x \end{aligned} \tag{5}$$

and

$$\begin{aligned} T_{\bar{x}} &= \inf\{T | R_\theta \bar{x}(t) = 0 \ \forall |t| > T, \theta \in [0, 2\pi)\} \\ &= \inf\{T | \frac{1}{\xi} R_\theta x(\xi t) = 0 \ \forall |t| > T, \theta \in [0, 2\pi)\} \\ &= \inf\{T | R_\theta x(\xi t) = 0 \ \forall |t| > T, \theta \in [0, 2\pi)\} \\ &= \inf\{T | R_\theta x(\xi t) = 0 \ \forall |\xi t| > \xi T, \theta \in [0, 2\pi)\} \\ &= \frac{1}{\xi} \inf\{\xi T | R_\theta x(\xi t) = 0 \ \forall |\xi t| > \xi T, \theta \in [0, 2\pi)\} \\ &= \frac{1}{\xi} \inf\{T' | R_\theta x(\tau) = 0 \ \forall |\tau| > T', \theta \in [0, 2\pi)\} \\ &\qquad (\text{let } T' = \xi T, \tau = \xi t) \\ &= \frac{1}{\xi} T_x. \end{aligned} \tag{6}$$

Therefore,

$$\begin{aligned} \frac{1}{M_{\bar{x}}} R_\theta \bar{x}\left(T_{\bar{x}} t\right) &= \frac{\xi}{M_x} R_\theta \bar{x}\left(\frac{T_x}{\xi} t\right) \\ &= \frac{\xi}{M_x \xi} R_\theta x\left(\frac{\xi T_x}{\xi} t\right) \\ &= \frac{1}{M_x} R_\theta x\left(T_x t\right). \end{aligned} \tag{7}$$

■

As the above result holds for any pair of $\bar{x}, x \in \mathcal{Y}$, we have shown the invariance of $\frac{1}{M_{\bar{x}}} R_\theta \bar{x}\left(T_{\bar{x}} t\right)$ over all $\bar{x} \in \mathcal{Y}$.

Fig. 4 illustrates an example of scale alignment in the Radon domain. Figs. 4(a) and (d) show the flower images with the same orientation but in different scales. The corresponding sinograms are shown in Figs. 4(b) and (e), respectively. The corresponding normalized sinograms obtained according to (4) are shown in Figs. 4(c) and (f), respectively. As can be seen from the figure, after the adjustment, the resulting sinograms are scale-aligned to each other.

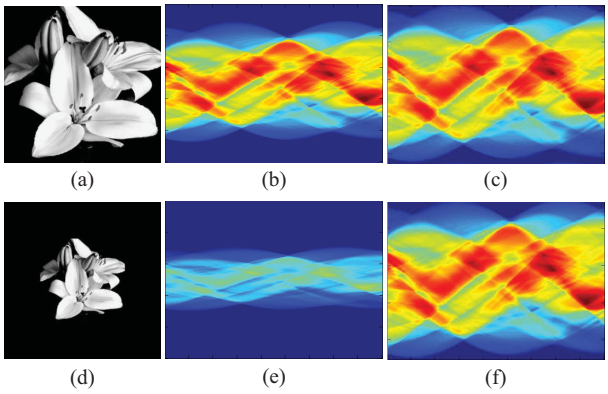



Fig. 4. Alignment of sinograms (a) and (d) show the flowers images with different scales. (b) and (e) Show their corresponding sinograms. Sinograms obtained after normalization are shown in (c) and (f). Note that after the adjustment, the resulting sinograms are closely scale-aligned to each other.

## III. Simultaneous Clustering and Dictionary Learning

In this section, we present the proposed rotation and scale invariant clustering and dictionary learning framework. Our method learns dictionaries and clusters images in the Radon domain. Let $\{\mathbf{x}_j\}_{j=1}^J$ be the database of images represented as vectors and $K$ be the number of clusters. $\mathbf{x}_j$ is a $Z_I \times 1$ column vector representing the $j$th image, where $Z_I$ is the image size (i.e., product of width and height in pixels). Let $\mathbb{C} \triangleq \{\mathbb{C}_k\}_{k=1}^K$ denote the collection of $K$ clusters such that $\mathbb{C}_k$ is a cluster that contains images belonging to the $k$th class. Given an image $\mathbf{x}_j$ and its estimated orientation $\hat{\theta}_j$ that is calculated from the discretized version of (3), we use a $Z_R \times 1$ column vector $\mathbf{R}_{\hat{\theta}_j} \mathbf{x}_j$ to denote the (column-vectorized) vector version of the column-shifted, scale normalized sinogram, where $Z_R$ is the sinogram size (i.e., product of width and height in pixels). This sinogram is obtained by left shifting columns of the scale normalized sinogram of $\mathbf{x}_j$ by $\hat{\theta}_j$. Let $\mathbf{C}_k$ be a $Z_R \times L_k$ matrix containing $\mathbf{R}_{\hat{\theta}_j} \mathbf{x}_j$s as columns, where $L_k$ is the population size of $\mathbb{C}_k$ (i.e., number of images in $\mathbb{C}_k$). Let $\mathbf{D}_k$ be the class dictionary learned from $\mathbf{C}_k$ such that $\mathbf{D}_k$ is a $Z_R \times d_k$ matrix where $d_k$ is the number of dictionary atoms. Define $\mathbf{D} = [\mathbf{D}_1 \ldots \mathbf{D}_K]$ as the concatenation of class dictionaries. Note that $\mathbf{C}_k$ is the matrix with columns as vector forms of $k$th-class images' sinograms. Table I gives a summary of the notations.

Our objective is to simultaneously cluster the data into $K$ groups and learn the best dictionaries for each cluster by solving the following optimization problem

$$\min_{\mathbb{C},\mathbf{D},\boldsymbol{\alpha}} \sum_{k=1}^{K} \sum_{\mathbf{x} \in \mathbb{C}_k} \min_{\theta} \left\{ \|R_\theta \mathbf{x} - \mathbf{D}\delta_k(\boldsymbol{\alpha})\|_2^2 + \mu_1 \|\boldsymbol{\alpha}\|_1 + \mu_2 \left| \frac{d^2 \tilde{\sigma}_\theta}{d\theta^2} \right| \right\} \tag{8}$$

where $\mu_1, \mu_2 > 0$, and $\|.\|_1$ denotes the $\ell_1$ norm, $\boldsymbol{\alpha}$ is the representation vector and $\delta_k(\boldsymbol{\alpha})$ is the masked version of $\boldsymbol{\alpha}$ such that its only nonzero entries are those of $\boldsymbol{\alpha}$ that correspond to the $k$th dictionary. Here, $\tilde{\sigma}_\theta$ is the variance of the column corresponding to $\theta$ of the scale-normalized sinogram of $\mathbf{x}$. In other words, $\tilde{\sigma}_\theta$ is the variance of the first column of

TABLE I
SUMMARY OF NOTATIONS

| Variable | Definition | Dimensions | Domain |
|---|---|---|---|
| $\mathbf{x}_j$ | a column representing the $j$th image | $Z_I \times 1$ | image |
| $\mathbb{C}_k$ | a cluster that contains $\mathbf{x}_j$s belonging to the $k$th class | $Z_I \times L_k$ | image |
| $\mathbb{C}$ | the collection all $\mathbb{C}_k$s | $Z_I \times (\sum_{k=1}^{K} L_k)$ | image |
| $\mathbf{R}_{\hat{\theta}_j}\mathbf{x}_j$ | column form of the column-shifted, scale normalized sinogram | $Z_R \times 1$ | Radon |
| $\mathbf{C}_k$ | a matrix with $\mathbf{R}_{\hat{\theta}_j}\mathbf{x}_j$s as columns | $Z_R \times L_k$ | Radon |
| $\mathbf{D}_k$ | the class dictionary learned from $\mathbf{C}_k$ | $Z_R \times d_k$ | Radon |
| $\mathbf{D}$ | the concatenation of class dictionaries $\mathbf{D}_k$s | $Z_R \times (\sum_{k=1}^{K} d_k)$ | Radon |

the sinogram that is column left-shifted (by $\theta$) version of the scale normalized sinogram of $\mathbf{x}$. According to (3), if $\theta$ is not the principal orientation, then the last term in (8) can never be the minimum. Therefore, this term introduces a penalty due to rotation misalignment. It uses $\tilde{\sigma}_\theta$ to estimate the presence of rotation in images. Our approach for solving the above optimization problem essentially consists of two steps: cluster assignment and dictionary learning. Detailed descriptions of these two steps are given below.

### A. Cluster Assignment (Box 4 in Fig. 1)

Given dictionary $\mathbf{D}^{(i)} = [\mathbf{D}_1^{(i)} \dots \mathbf{D}_K^{(i)}]$ at iteration $i$, we obtain the sparse representation of $\mathbf{R}_{\hat{\theta}_j}\mathbf{x}_j$ in $\mathbf{D}^{(i)}$ by solving the following optimization problem:

$$\boldsymbol{\alpha}^j = \arg\min_{\omega} \|\boldsymbol{\omega}\|_1 \ \text{ subject to } \ \mathbf{R}_{\hat{\theta}_j}\mathbf{x}_j = \mathbf{D}^{(i)}\boldsymbol{\omega}. \quad (9)$$

Several approaches have been suggested for solving (9) [37]. In our approach, we employ a highly efficient algorithm that is suitable for large-scale applications known as the spectral projected gradient (SPGL1) algorithm [38]. Once the sparse coefficients are found, $\mathbf{x}_j$ is set to belong to cluster $\hat{k}$ if the coefficients corresponding to the $\hat{k}$th dictionary give the best representation of the sinogram $\mathbf{R}_{\hat{\theta}_j}\mathbf{x}_j$ [39]. In other words, if

$$\hat{k} = \arg\min_{k} \|\mathbf{R}_{\hat{\theta}_j}\mathbf{x}_j - \mathbf{D}^{(i)}\delta_k(\boldsymbol{\alpha}^j)\|_2^2, \quad j = 1, \dots, J \quad (10)$$

then $\mathbf{x}_j$ is set to belong to $\mathbb{C}_{\hat{k}}^{(i)}$. The motivation for this consideration is that if $\mathbf{x}_j$ belongs to the $k$th cluster, then the dictionary corresponding to cluster $k$ will represent $\mathbf{R}_{\hat{\theta}_j}\mathbf{x}_j$ well.

### B. Dictionary Learning (Box 5 in Fig. 1)

The K-SVD algorithm is a common dictionary learning algorithm [1]. Given clusters $\{\mathbb{C}_k^{(i)}\}_{k=1}^K$, we use the K-SVD algorithm to learn the dictionary $\mathbf{D}^{(i+1)} = [\mathbf{D}_1^{(i+1)} \dots \mathbf{D}_K^{(i+1)}]$. In this section, we first detail the K-SVD principle. Then, we show how it is included in the dictionary learning step of our RSICD method.

*1) K-SVD Algorithm:* Given a set of input examples (in a column-vectorized form) $\{\mathbf{y}_l^k\}_{l=1}^{n_k}$ belonging to the $k$th class, the K-SVD algorithm finds a dictionary $\hat{\mathbf{B}}_k$ that provides the best representation for each example in this set by solving the following optimization problem:

$$(\hat{\mathbf{B}}_k, \hat{\Lambda}_k) = \arg\min_{\mathbf{B}_k, \Lambda_k} \|\mathbf{Y}_k - \mathbf{B}_k\Lambda_k\|_F^2, \ \text{s.t. } \|\boldsymbol{\lambda}_{k,l}\|_0 \le T_0 \quad \forall l \in \{1, ..., n_k\} \forall k \in \{1, ..., K\} \quad (11)$$

where $\boldsymbol{\lambda}_{k,l}$ represents the $l^{th}$ column of $\Lambda_k$, $\mathbf{Y}_k$ is the matrix whose columns are $\mathbf{y}_l^k$s and $T_0$ is the sparsity parameter. Here, the Frobenius norm is defined as $\|\mathbf{A}\|_F = \sqrt{\sum_{ij} A_{ij}^2}$ and the norm $\|\boldsymbol{\lambda}\|_0$ counts the number of non-zero elements in $\boldsymbol{\lambda}$. The K-SVD algorithm alternates between sparse-coding and dictionary update steps.

In the sparse-coding step, $\mathbf{B}_k$ is fixed and the representation vectors $\boldsymbol{\lambda}_{k,l}$s are found for each example $\mathbf{y}_l^k$ by solving the following equation:

$$\min_{\boldsymbol{\lambda}_{k,l}} \|\mathbf{y}_l^k - \mathbf{B}_k\boldsymbol{\lambda}_{k,l}\|_2^2 \quad \text{such that } \|\boldsymbol{\lambda}_{k,l}\|_0 \le T_0 \quad \forall l \in \{1, .., n_k\} \forall k \in \{1, ..., K\}. \quad (12)$$

As solving (12) is NP-hard, approximate solutions are usually sought [37], [40]. Greedy pursuit algorithms such as matching pursuit and orthogonal matching pursuit [41] are often used to find the approximate solutions to the above sparse coding problem [42]. In the dictionary update step, the dictionary is updated atom-by-atom in an efficient way. The K-SVD algorithm has been observed to converge in a few iterations.

*2) Learning $\mathbf{D}^{(i+1)}$:* Having obtained clusters $\{\mathbb{C}_k^{(i)}\}_{k=1}^K$, we update the dictionaries $\mathbf{D}_k^{(i+1)}$ with the K-SVD algorithm. In particular, we find the best representation of the members in $\mathbf{C}_k^{(i)}$ by solving the following optimization problem:

$$\left(\mathbf{D}_k^{(i+1)}, \Gamma_k^{(i+1)}\right) = \arg\min_{\mathbf{D}_k, \Gamma_\mathbf{k}} \|\mathbf{C}_k^{(i)} - \mathbf{D}_k\Gamma_k\|_F^2, \text{s.t.} \|\boldsymbol{\gamma}_l\|_0 \le T_0 \quad \forall l \in \{1, ..., L_k\} \quad \forall k \in \{1, ..., K\} \quad (13)$$

where $\Gamma_k^{(i+1)}$ is a $d_k \times L_k$ coefficient matrix that contains $\boldsymbol{\gamma}_l$s as its columns. Here $\mathbf{C}_k^{(i)}$, $L_k$, $\mathbf{D}_k$ and $\Gamma_k$ in (13) correspond to $\mathbf{Y}_k$, $n_k$, $\mathbf{B}_k$ and $\Lambda_k$ in (11), respectively. Note that each $\mathbf{R}_{\hat{\theta}_j}\mathbf{x}_j$ corresponds to a $\mathbf{y}_l^k$ for some $k$ and $l$ in (11).

### C. RSICD Algorithm

Our RSICD algorithm is an iterative approach, where there are global iterations and local iterations. Each global iteration consists of cluster assignment and dictionary learning. As the K-SVD algorithm is used for the dictionary learning step, this step further consists of local K-SVD iterations. In particular, given $\mathbf{C}_k^{(i)}$ in the beginning of the dictionary learning step in the $i$th global iteration, $\mathbf{D}_k$ in (13) is set by $\mathbf{D}_k^{(i)}$. After a few local K-SVD iterations, $\mathbf{D}^{(i+1)}$ is obtained as the updated dictionary for the next global iteration. We iteratively repeat the cluster assignment and dictionary learning steps till there is no significant change in $\{\mathbb{C}_k^{(i+1)}\}_{k=1}^K$.

Note that (9) is used in each global iteration under an error constraint to find the sparsest coefficients using the concatenation of all class dictionaries, while (12) is used in each local iteration under a sparsity constraint to find the coefficients that give the minimum representation error for each class dictionary.

### *D. Obtaining Initial Dictionaries*

As one can see from the previous discussion, the performance of our algorithm depends on the choice of initial dictionaries. In this section, we describe a method for obtaining the initial dictionary $\mathbf{D}^{(0)} = [\mathbf{D}_1^{(0)} \dots \mathbf{D}_K^{(0)}]$.

Let $L \triangleq \min_{k\in\{1,\dots K\}} L_k$ be the minimum cluster population size. To determine initial clusters $\{\mathbb{C}_1^{(0)}, \dots, \mathbb{C}_K^{(0)}\}$, we propose an approach that uses the Hamming distance between $R_i$ and $R_j$ for any pair $(i, j) \in \{1, 2, ..., J\}$, where $R_j$ is the set that consists of $\mathbf{R}_{\hat{\theta}_j}\mathbf{x}_j$ and its $L-1$ nearest neighbors. Algorithm 1 details our approach.

Let $S$ be the set that consists of $\mathbf{R}_{\hat{\theta}_j}\mathbf{x}_j$, $j = 1, 2, \dots, J$ (i.e., $S = \{\mathbf{R}_{\hat{\theta}_j}\mathbf{x}_j\}_{j=1}^{J}$). In step 1, for each $\mathbf{R}_{\hat{\theta}_j}\mathbf{x}_j$, we find its $L-1$ nearest neighbors and obtain the set $R_j$. $\mathbf{R}_{\hat{\theta}_j}\mathbf{x}_j$ in general should be closer to other within-class members than to other different-class members. Therefore, we expect $R_j$s that correspond to within-class members to be similar to each other, while different to those that correspond to other different-class members. In step 2, we calculate the Hamming distance between $R_i$ and $R_j$, defined by

$$d(R_i, R_j) \triangleq (L - \#\text{ of common elements in } R_i \text{ and } R_j).$$

Like the Euclidean distance, the Hamming distance $d(R_i, R_j)$ is an indication of how far the feature pair $\mathbf{R}_{\hat{\theta}_i}\mathbf{x}_i$ and $\mathbf{R}_{\hat{\theta}_j}\mathbf{x}_j$ are separated from each other. However, $d(R_i, R_j)$ for $\mathbf{R}_{\hat{\theta}_i}\mathbf{x}_i$ and $\mathbf{R}_{\hat{\theta}_j}\mathbf{x}_j$ that belong to two different classes, is always upper bounded by $L$. Hence, the Hamming distance $d(R_i, R_j)$ preserves the class separability between any different-class pair $\mathbf{R}_{\hat{\theta}_i}\mathbf{x}_i$ and $\mathbf{R}_{\hat{\theta}_j}\mathbf{x}_j$. Using this property, we are able to choose $K$ initial representatives such that they belong to $K$ different classes with high probability (as shown in steps 3 to 6). With these $K$ initial representatives, the corresponding initial $K$ partitions are determined by the nearest neighbor criterion (step 7). We assume each of the $K$ initial partitions contains one exemplar. For all subsequent iterations (steps 8 and 9), $K$ distinct representatives $\{\mathbf{s}_k\}_{k=1}^{K}$ are always chosen from these predetermined $K$ initial partitions, and are used to calculate the associated score $M(S(\{\mathbf{s}_k\}_{k=1}^{K}))$ defined by

$$M\left(S(\{\mathbf{s}_k\}_{k=1}^{K})\right) = \frac{div\left(S(\{\mathbf{s}_k\}_{k=1}^{K})\right)}{err\left(S(\{\mathbf{s}_k\}_{k=1}^{K})\right)}.$$

The terms $err\left(S(\{\mathbf{s}_k\}_{k=1}^{K})\right)$ and $div(S(\{\mathbf{s}_k\}_{k=1}^{K}))$ are *square error* and *diversity*, respectively [43]. They are defined as follows:

$$err\left(S\left(\{\mathbf{s}_k\}_{k=1}^{K}\right)\right) \triangleq tr\left[\sum_{k=1}^{K}\sum_{\mathbf{s}\in S_k}(\mathbf{s}-\mathbf{s}_k)(\mathbf{s}-\mathbf{s}_k)^T\right]$$

**Algorithm 1** Design of Initial Dictionary, $\mathbf{D}^{(0)}$

**Input**: Scale and rotation aligned sinograms, $\mathbf{R}_{\hat{\theta}_j}\mathbf{x}_j$, $j = 1, 2, ...J$.

**Initialization of sets:** $S \leftarrow \{\mathbf{R}_{\hat{\theta}_j}\mathbf{x}_j\}_{j=1}^{J}$, $I \leftarrow \{1, 2, ..., J\}$ $T \leftarrow \phi$. **Procedure:**

**1.** For each $\mathbf{R}_{\hat{\theta}_j}\mathbf{x}_j$ in $S$, find its $L-1$ nearest neighbors. $\mathbf{R}_{\hat{\theta}_j}\mathbf{x}_j$ and its $L-1$ nearest neighbors form a set denoted by $R_j$.

**2.** For all pairs $(i, j)$, calculate the Hamming distance between $R_i$ and $R_j$, $d(R_i, R_j)$.

**3.** Find $(i^*, j^*) = \underset{i,j\in I, i\neq j}{\text{argmax}}\, d(R_i, R_j)$.

**4.** Update of sets: $t_1 \leftarrow i^*, t_2 \leftarrow j^*, T \leftarrow T \bigcup \{t_1, t_2\}$, $I \leftarrow I \setminus \{i^*, j^*\}$. If $|T| = K$, goto 7.

**5.** Find $k^* = \underset{k\in I}{\text{argmax}} \prod_{l=1}^{|T|} d(R_{t_l}, R_k)$.

**6.** Update of sets: $t_{|T|+1} \leftarrow k^*, T \leftarrow T \bigcup \{t_{|T|+1}\}$, $I \leftarrow I \setminus \{k^*\}$, and goto step 3.

**7.** Given $\{\mathbf{R}_{\hat{\theta}_{t_k}}\mathbf{x}_{t_k}\}_{k=1}^{K}$, use the nearest neighbor criterion to partition $S$ into $K$ partitions, denoted by $S(\{\mathbf{R}_{\hat{\theta}_{t_k}}\mathbf{x}_{t_k}\}_{k=1}^{K}) = \bigcup_{k=1}^{K} S_k$.

**8.** Randomly select $\mathbf{s}_k$ from $S_k$, $k = 1, 2, ..., K$, as representatives. Find the corresponding nearest neighbor partitions $S(\{\mathbf{s}_k\}_{k=1}^{K})$, and calculate the corresponding score $M(S(\{\mathbf{s}_k\}_{k=1}^{K}))$.

**9.** Repeat step 8, and keep updating for $\{\mathbf{s}_k^*\}_{k=1}^{K}$ that gives the highest score $M$, until the number of repeating iterations for step 9 reaches $W_1$. In other words,

$$\{\mathbf{s}_1^*, ..., \mathbf{s}_K^*\} = \underset{\mathbf{s}_k\in S_k, k=1,2,...,K, \text{ in } W_1 \text{ iterations}}{\text{argmax}} M(S(\{\mathbf{s}_k\}_{k=1}^{K})).$$

**10.** Obtain $K$ initial clusters $\{\mathbb{C}_1^{(0)}, \dots, \mathbb{C}_K^{(0)}\}$ from $S(\{\mathbf{s}_k^*\}_{k=1}^{K})$.

**Output**: Initial dictionaries, $\mathbf{D}^{(0)} = [\mathbf{D}_1^{(0)} \dots \mathbf{D}_K^{(0)}]$, where $\mathbf{D}_k^{(0)} = \mathbf{C}_k^{(0)}$, $k = 1, \dots, K$.

and

$$div\left(S\left(\{\mathbf{s}_k\}_{k=1}^{K}\right)\right) \triangleq tr\left[\sum_{k=1}^{K}(\mathbf{s}_k-\bar{\mathbf{s}})(\mathbf{s}_k-\bar{\mathbf{s}})^T\right]$$

where $\bar{\mathbf{s}} = \frac{1}{K}\sum_{k=1}^{K}\mathbf{s}_k$ and $tr(\mathbf{A})$ denotes the trace of matrix $\mathbf{A}$. The *diversity* represents the scatter of representatives to their mean, while the *square error* represents the total summation of partition-specific scatters, over all $K$ partitions. The maximization of $M(S(\{s_k\}_{k=1}^{K}))$ is achieved through maximizing the *diversity* while minimizing the *square error*. Since these $K$ exemplars by assumption respectively fall within the $K$ initial partitions, they can be found after a sufficient number of iterations. The representatives that give the maximum score $M(S(\{\mathbf{s}_k\}_{k=1}^{K}))$ in $W_1$ iterations, are recorded as exemplars. The corresponding final partitions are obtained by finding nearest neighbors of the exemplars.

**Algorithm 2** RSICD-Based CBIR

**Input**: Database $\{\mathbf{x}_j\}_{j=1}^{J}$ and query image $\mathbf{x}_q$.
**1.** Use (3) and (4) to obtain scale and rotation aligned sinograms of $\{\mathbf{x}_j\}_{j=1}^{J}$ and $\mathbf{x}_q$.
**2.** Use Algorithm 1 to design initial dictionaries $\{\mathbf{D}_k^{(0)}\}_{k=1}^{K}$.
**3.** Given $\mathbf{D}^{(i)} = [\mathbf{D}_1^{(i)} \cdots \mathbf{D}_K^{(i)}]$, assign each $\mathbf{x}_j$ to $\mathbb{C}_{\hat{k}}^{(i)}$, where $i$ denotes the current iteration number, and $\hat{k}$ is obtained from (9) and (10).
**4.** Given $\{\mathbb{C}_k^{(i)}\}_{k=1}^{K}$, use the K-SVD algorithm to learn $\mathbf{D}_k^{(i+1)}$ from $\mathbb{C}_k^{(i)}$, $k = 1, 2, \ldots, K$. Then increment $i$ by 1 (i.e., $i \leftarrow (i+1)$).
**5.** Repeat **3** and **4** until the number of repeating iterations reaches $W_2$.
**6.** Determine the closest cluster to $\mathbf{x}_q$ from (17), from which the relevances are found by the nearest neighbor criterion.
**Output**: Clusters $\{\mathbb{C}_k^{(W_2)}\}_{k=1}^{K}$, dictionaries $\{\mathbf{D}_k^{(W_2)}\}_{k=1}^{K}$, and the relevance of $\mathbf{x}_q$ in its closest cluster.

### *E. Application to CBIR*

In this subsection, we show how the proposed simultaneous clustering and dictionary learning method is used in CBIR. Once the dictionaries have been learned for each class in the Radon domain, given a query image $\mathbf{x}_q$, we obtain its scale and rotation normalized sinogram $\mathbf{R}_{\hat{\theta}_q}\mathbf{x}_q$. Then, we project $\mathbf{R}_{\hat{\theta}_q}\mathbf{x}_q$ onto the span of the atoms in each $\mathbf{D}_k$ using the orthogonal projector

$$\mathbf{Proj}_{\mathbf{D}_k} = \mathbf{D}_k(\mathbf{D}_k^T\mathbf{D}_k)^{-1}\mathbf{D}_k^T. \tag{14}$$

The approximation and residual vectors can then be calculated as

$$\mathbf{R}_{\hat{\theta}_q}^k\mathbf{x}_q = \mathbf{Proj}_{\mathbf{D}_k}(\mathbf{R}_{\hat{\theta}_q}\mathbf{x}_q) \tag{15}$$

and

$$\mathbf{r}^k(\mathbf{R}_{\hat{\theta}_q}\mathbf{x}_q) = \mathbf{R}_{\hat{\theta}_q}\mathbf{x}_q - \mathbf{R}_{\hat{\theta}_q}^k\mathbf{x}_q = (\mathbf{I} - \mathbf{Proj}_{\mathbf{D}_k})\mathbf{R}_{\hat{\theta}_q}\mathbf{x}_q \tag{16}$$

respectively, where $\mathbf{I}$ is the identity matrix. Since the dictionary learning step in our algorithm finds the dictionary $\mathbf{D}_k$ that leads to the best representation for each member of $\mathbb{C}_k$ in the Radon domain, we assume $\|\mathbf{r}^k(\mathbf{R}_{\hat{\theta}_q}\mathbf{x}_q)\|_2$ is small if $\mathbf{x}_q$ belongs to the $k$th cluster and larger for the other clusters. Based on this, if

$$d = \arg\min_{1\le k\le K} \|\mathbf{r}^k(\mathbf{R}_{\hat{\theta}_q}\mathbf{x}_q)\|_2, \tag{17}$$

we search for the relevance of $\mathbf{x}_q$ in $\mathbb{C}_d$ by means of a nearest neighbor search (box 6, 7 in Fig. 1). We refer to our in-plane rotation and scale invariant clustering using dictionaries method as RSICD. Algorithm 2 summarizes the overall RSICD-based CBIR procedure.

The RSICD-based CBIR algorithm consists of two main steps: cluster assignment and dictionary learning. The overall algorithm is not convex on both of these steps. It is likely that the approach may get stuck in a local minima. However, experiments on various training sets have shown that it usually takes about 20 iterations for the algorithm to converge.

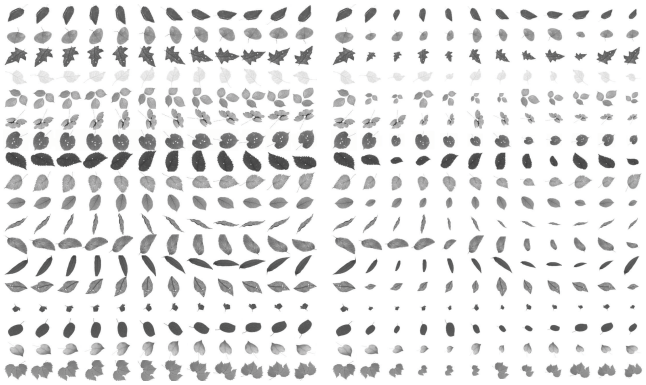


Fig. 5. (a) Sample images from the generated dataset containing the rotated images from the Smithsonian dataset. (b) Sample images from the Smithsonian dataset containing both scale and rotation variations.

## IV. Experimental Results

In this section, we show the effectiveness of the proposed simultaneous clustering and dictionary approach. We report the results of empirical evaluation of our method and compare it with six state-of-the-art matching algorithms on three standard datasets: the Smithsonian isolated leaf dataset [32], Kimia's object dataset [31] and Brodatz texture dataset [44]. We compare the performance of our method with a modified Gabor-based approach [21], a local binary pattern (LBP)-based approach [35], and three recently proposed feature-based approaches [30], [32], [33]. We refer to the indexing and retrieval method presented in [33] as the BAC[3] method. Note that methods presented in [30], [32], [33] are (in-plane) rotation and scale invariant as well. In addition, we compare our method with a recently proposed unsupervised discriminative dictionary learning method [15]. We refer to the method presented in [15] as dictionary-based clustering (DC).

For all the experiments implemented using LBP [35], we first resized each image to $40 \times 40$ pixels. Each resized image consists of 25 square patches, each with 64 pixels. On each patch we implemented uniform rotation-invariant LBP with $P = 8$ and $R = 1$, where $P$ is the member number in a circularly symmetric neighbor set, and $R$ is the corresponding radius [35]. Results of all 25 patches are then combined to form a feature vector of the image. For the experiments implemented using the dictionary-based methods (RSICD and DC), we set the sparsity parameter $T_0$ to be 20.

We evaluate the performance of various methods using precision-recall curves, average retrieval performance [21], [29] and recognition rates. Recall and precision are defined as

$$\text{Precision} = \frac{\text{Number of relevant images retrieved}}{\text{Total number of images retrieved}}, \tag{18}$$

$$\text{Recall} = \frac{\text{Number of relevant images retrieved}}{\text{Total number of relevant images}}. \tag{19}$$

[3]As no representative name was given for the method presented in [33], we chose the first letter of each author's last name (i.e., "B", "A", and "C" in, respectively) and connected these letters as 'BAC' to stand for their respective methods.

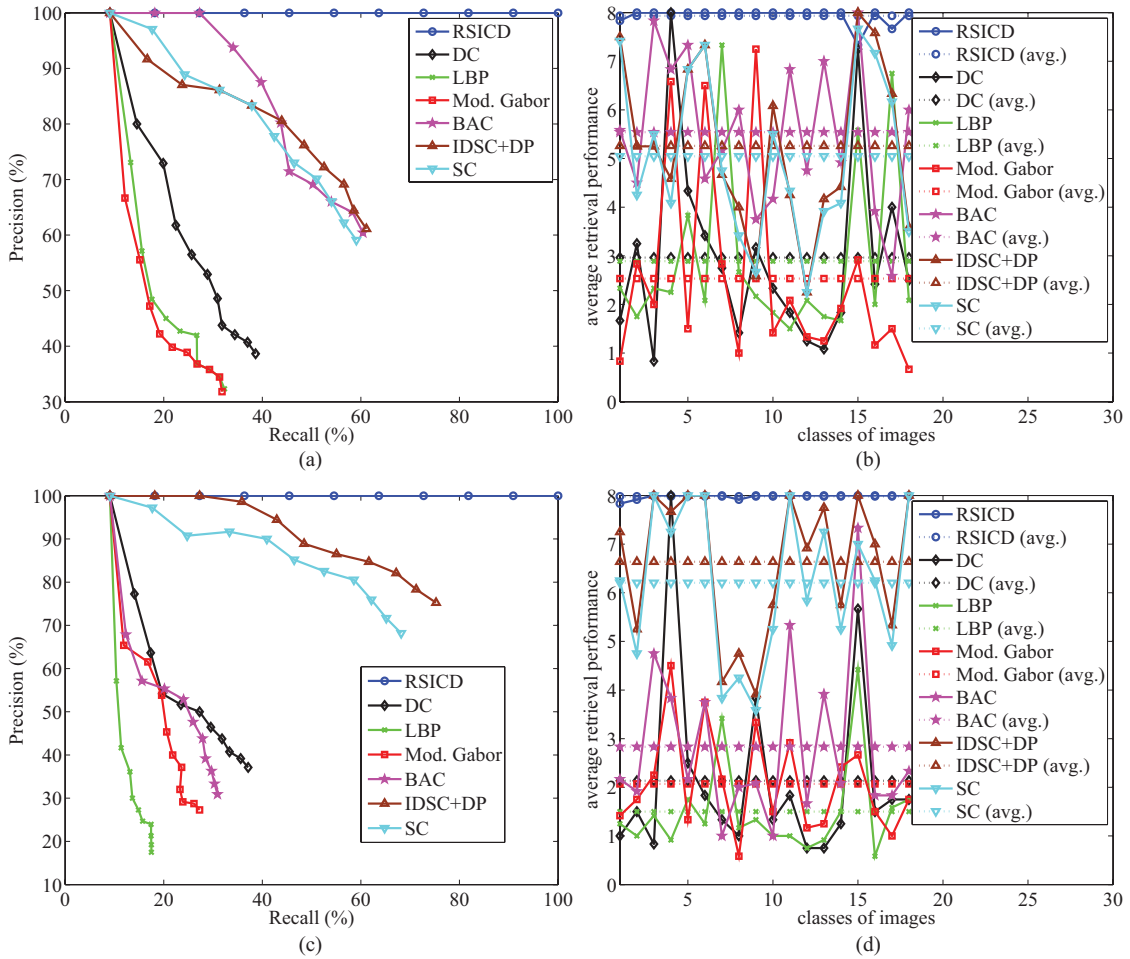



Fig. 6. Results on rotated 18-class Smithsonian datasets. (a) Precision-recall curves. (b) Average retrieval performance corresponding to the dataset containing the rotated images. (c) Precision-recall curves and (d) average retrieval performance corresponding to the dataset containing the rotated and scaled images. For both of these datasets, the proposed RSICD achieves the best precision rates for almost all recall rates and outperforms other methods.

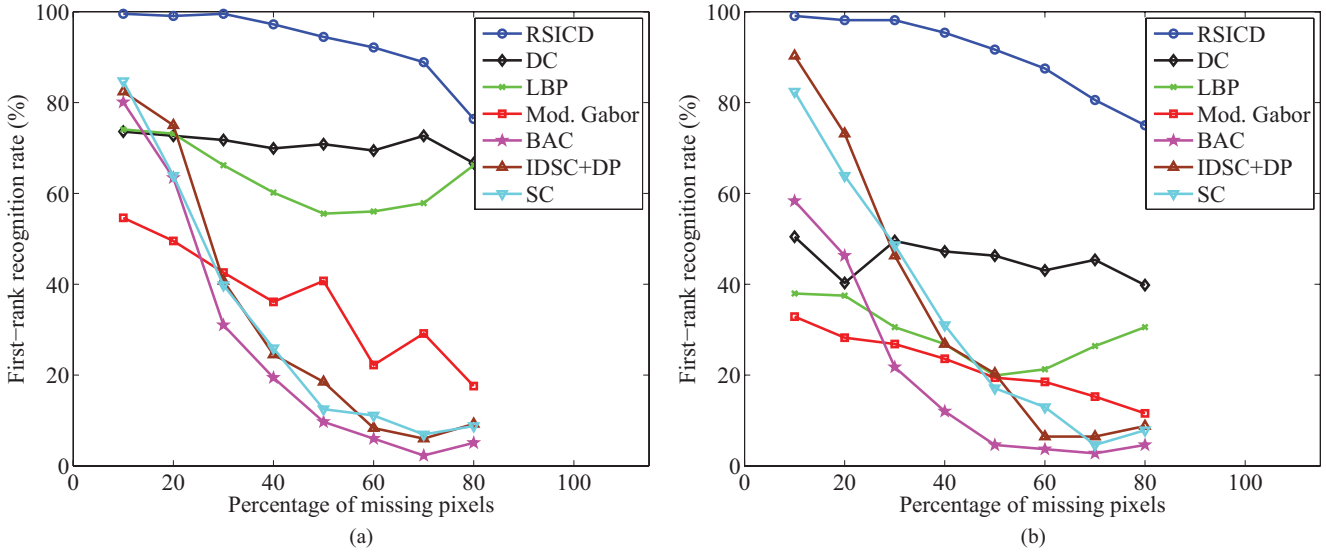



Fig. 7. First-rank recognition rates on 18-class Smithsonian datasets with missing pixels. (a) Experiment with the dataset with rotated images. (b) Experiment with the dataset containing both rotated and scaled images. These results show the proposed RSICD is robust to effects of missing pixels.

Recall is the portion of total relevant images retrieved whereas precision indicates the capability to retrieve only relevant images. An ideal retrieval should give precision rate that always equals 100% for any recall rate. Given a certain number of retrieved images, the average retrieval performance is defined as the average number of relevant retrieved images

TABLE II

RANK RECOGNITION RATES (%) CORRESPONDING THE DATASET CONTAINING 18-CLASSES WITH ROTATED IMAGES FROM THE SMITHSONIAN ISOLATED LEAF DATABASE

| Rank | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Modified Gabor [21] | 60.65 | 27.78 | 31.02 | 31.48 | 30.09 | 25.93 | 25.00 | 21.30 | 22.22 | 15.28 | 19.44 |
| LBP [35] | 77.78 | 49.54 | 35.19 | 34.26 | 25.46 | 22.69 | 23.15 | 20.83 | 18.06 | 15.74 | 14.35 |
| DC [15] | 74.54 | 55.56 | 48.61 | 40.74 | 32.41 | 27.78 | 25.00 | 20.83 | 23.61 | 17.59 | 17.13 |
| SC [30] | 91.20 | 80.09 | 74.54 | 68.98 | 58.33 | 51.39 | 40.28 | 39.81 | 36.57 | 26.39 | 17.13 |
| IDSC+DP [32] | 92.13 | 78.70 | 73.15 | 68.98 | 66.67 | 55.09 | 48.15 | 43.06 | 35.19 | 29.63 | 24.07 |
| BAC [33] | 91.20 | 92.13 | 85.19 | 75.93 | 66.20 | 54.17 | 49.07 | 40.28 | 33.33 | 33.33 | 25.46 |
| RSICD | **100** | **100** | **100** | **99.54** | **100** | **97.69** | **99.54** | **99.07** | **100** | **99.54** | **93.52** |

TABLE III

RANK RECOGNITION RATES (%) CORRESPONDING THE DATASET CONTAINING 18-CLASSES WITH ROTATED AND SCALED IMAGES FROM THE SMITHSONIAN ISOLATED LEAF DATABASE

| Rank | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Modified Gabor [21] | 48.61 | 32.87 | 31.48 | 24.54 | 20.83 | 17.59 | 18.06 | 12.96 | 9.72 | 12.96 | 10.65 |
| LBP [35] | 41.67 | 20.83 | 13.89 | 20.83 | 16.20 | 15.28 | 8.80 | 12.50 | 8.80 | 10.65 | 5.09 |
| DC [15] | 55.09 | 35.65 | 26.39 | 23.61 | 18.98 | 19.44 | 16.20 | 15.28 | 16.67 | 14.81 | 11.11 |
| SC [30] | 96.30 | 93.52 | 87.96 | 83.80 | 78.70 | 67.13 | 58.33 | 54.63 | 53.70 | 45.83 | 40.74 |
| IDSC+DP [32] | 99.07 | 99.07 | 93.52 | 86.57 | 79.17 | 73.61 | 69.91 | 62.96 | 63.43 | 56.02 | 49.07 |
| BAC [33] | 67.59 | 49.07 | 42.59 | 34.26 | 26.39 | 25.46 | 21.30 | 16.67 | 17.59 | 12.96 | 8.80 |
| RSICD | **100** | **100** | **100** | **99.07** | **99.54** | **100** | **99.07** | **99.54** | **99.07** | **98.15** | **95.37** |

over all query images of a particular class. On the other hand, the rank-$n$ recognition rates indicate how well the recognition performance of an approach can maintain from the best-match retrieval up to the $n$-th best-match retrieval. An ideal retrieval should also maintain 100% recognition rate for any rank-$n$ retrieval.

### A. Smithsonian Isolated Leaf Database

The original Smithsonian isolated leaf database consists of 93 different leaves [32]. From the original database, we created two challenging datasets, one containing rotated images and the other containing both rotated and scaled images. For the first set of experiments using this dataset, one representative image is selected from each of the last 18 leaves. We created an 18-class Smithsonian dataset by generating 11 additional in-plane rotated images with the following angles

$$18^\circ, 36^\circ, 54^\circ, 72^\circ, 90^\circ, 108^\circ, 126^\circ, 144^\circ, 162^\circ, 180^\circ, 198^\circ. \tag{20}$$

This sub-dataset contains rotated images of different leaves. We also created a dataset that contains both rotated and scaled images. This dataset is created by using the same rotated images as before. However, a random scaling that ranges from 0.25 to 1, is further applied to these rotated images. As a result, both of these sub-datasets have 18 classes and each class contains 12 different images. Fig. 5(a) and (b) show the resulting datasets containing rotated as well as rotated and scaled images, respectively. In both of these datasets, the final images were resized to $100 \times 80$ pixels.

*1) Results on the Smithsonian Dataset With 18 Classes:* In the first experiment using this dataset, we selected the last image (i.e., the 12-th image) of each class to form a query set (with 18 query images), and all the other images to form an unsupervised[4] training set (with $216 - 18 = 198$ training images). The dictionary $\mathbf{D}$ is of size $288 \times 90$. Five atoms per class dictionary are learned and concatenated to form the dictionary $\mathbf{D}$. Here, 288 is the dimension of the vectorized sinogram.

The results of this experiment are shown in Fig. 6(a) and (c). From the precision-recall curves we see that the proposed RSICD achieves ideal precision rates for all recall rates and outperforms other competitive methods. Fig. 6(b) and (d) show the total average retrieval performance over all shapes. For the sub-dataset containing rotated images only, on average RSICD obtained 7.9352 out of 8 retrieved images per shape. Whereas the BAC [33], IDSC+DP [32], SC [30], DC [15], LBP [35] and Gabor-based methods [21] obtained 5.5417, 5.2593, 5.0463, 2.9630, 2.8809 and 2.5324, respectively.

For the dataset containing rotated and scaled images, our RSICD obtained 7.9815 out of 8 retrieved images per shape. The BAC, IDSC+DP, SC, DC, LBP and Gabor-based methods obtained 2.8333, 6.6389, 6.2037, 2.1343, 1.5000 and 2.5324, respectively.

In our CBIR experiments, to determine the class label of a given test image, we find its $n$th nearest neighbor among the training images, and then assign that training image's estimated class label (given by our RSICD algorithm) to the test image. The $n$th rank recognition rate is therefore defined as the ratio of the number of test images' $n$th nearest neighbors (among the training images) that are assigned with the same class labels as the true labels of the test images, to the total number of test images. Tables II and III show the rank recognition rates for the above two 18-class datasets. Numbers in the abscissa of Table II, and III are the values of $n$ (same

[4]Here 'unsupervised' means samples' class labels are unknown to the algorithm initially.

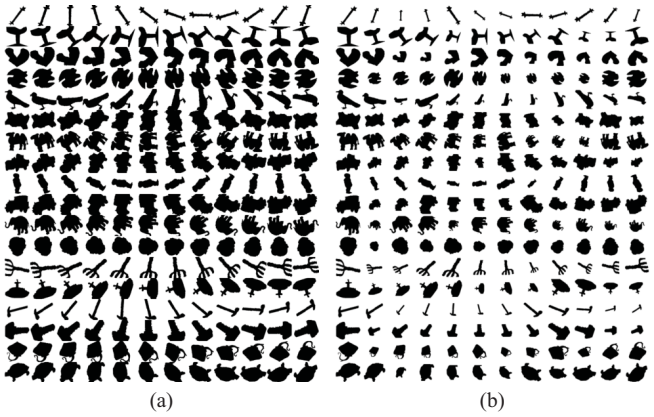


Fig. 8. Kimia datasets containing (a) rotated images and (b) rotated and scaled images.

for Tables IV, V, VI, VII, VIII, and IX). We observe that the proposed RSICD performs favorably in comparison to other methods.

*2) Results on All 93 Classes of the Smithsonian Dataset:* In the second set of experiments with the Smithsonian leaves dataset, we used all 93 classes to evaluate the rank recognition rates of different methods. Similar to the 18-class sub-datasets, we created rotated and scaled images for all 93 classes. Five atoms per class dictionary are learned and concatenated to form the dictionary $\mathbf{D}$ of size $288 \times 465$.

Tables IV and V show the rank recognition rates of the 1st up to the 15th rank retrieval. For both datasets, the recognition rates of the RSICD are the highest and at least 10% above those of the others for all rank retrievals. Comparing the RSICD results in Table II and Table III, the average recognition rate goes from 98.99% (18-classes) to 90.95% (93-classes). This decrease is only 8.04%, which shows that the RSICD is robust in maintaining recognition performances on rotated and scaled leaf databases across different class numbers.

*3) Robustness of RSICD to Missing Pixels:* We compare the results obtained by different methods when pixels are randomly removed from the query and probe images. The results are shown in Fig. 7 where we compare the rank-1 recognition rates of different methods as we vary the percentage of missing pixels. We can see that both dictionary methods (RSICD and DC) outperform the other methods. The RSICD is able to maintain its recognition rate at 75% even when 80% of the pixels are missing and it performs better than the DC.

### *B. Kimia Shape Database*

The Kimia database [31] consists of 216 images, where there are 18 shapes with small rotation, and each shape has 12 different images. Similar to how we generated new datasets for the Smithsonian Leaf database in the previous experiments, we created two sub-datasets from the original Kimia database: one containing the rotated images and the other containing the rotated and scaled images.

To obtain the rotated images, we selected one representative image from each of the 18 shapes in the original Kimia dataset. For each selected shape, 11 in-plane rotated images with the same angles as in (20) were created. We created a dataset that contains rotated and scaled images by scaling the rotated images as before. The resulting datasets are shown in Fig. 8(a) and (b), respectively. They possess more rotation and scale challenges than the original Kimia's dataset.

We selected the last image (i.e., the 12-th image) of each class to form a query set, and used all the other images to form an unsupervised training set. The precision-recall curves are shown in Fig. 9(a) and the average retrieval performance curves are shown in Fig. 9(b) for the datasets containing the rotated images. The dictionary size is set as 288.

As can be seen from both these figures, our method outperforms other competitive methods. Regarding the overall retrieval performance, our RSICD obtained 7.0324 out of 8 retrieved images per shape. Whereas the BAC [33], IDSC+DP [32], SC [30], DC [15], LBP [35] and Gabor-based methods [21] obtained 6.5185, 4.0231, 3.8935, 2.0926, 2.3657 and 1.2778, respectively. Table VI shows rank recognition rates for the above two 18-class datasets. The rank recognition rates of our RSICD remains the second while they still are close to the best results from BAC, for up to the 5th rank recognition.

Figs. 9(c) and (d) show the results obtained using the dataset containing both scaled and rotated images. For the overall retrieval performance, our RSICD obtained 7.0463 out of 8 retrieved images per shape. The BAC, IDSC+DP, SC, DC, LBP and Gabor-based methods obtained 6.5185, 7.8472, 7.2269, 1.2500, 1.3889 and 1.3426, respectively. Table VII shows the corresponding rank recognition rates.

### *C. Brodatz Texture Database*

In addition to shape-based datasets, we demonstrate the effectiveness of our RSICD on the Brodatz texture dataset [44]. We selected 25 textures and 60 textures from the Brodatz database, which are the dataset 1 and the dataset 3 defined in [36]. For each selected texture, we generated its in-plane rotated versions at the following angles: 10°, 20°, 30°, ..., 170°. Each original texture image and its 17 rotated images form a new in-plane rotated class. Fig. 10 shows the resulting sample images from the 25-class dataset. The dictionaries are of size $192 \times 300$ and $192 \times 720$ for 25 and 60 class datasets, respectively. Here, 192 is the size of vectorized sinogram.

Experimental results using 25 classes and 60 classes are compared in Table VIII and Table IX, respectively. The modified Gabor method gives 100% recognition rate on its first rank retrieval but degrades faster than the other methods within the first 4 rank retrievals. The average recognition rates of the RSICD are 60.17% (25-class) and 53.18% (60-class), which are higher than those of CD: 39.86% (25-class), 34.17% (60-class); LBP: 44.24% (25-class), 30.14% (60-class); and modified Gabor: 46.54% (25-class), 36.41% (60-class). This experiment shows that the RSICD is general enough that it can also perform well on a dataset that contains rotated textures.

### *D. Discussion*

In this section, we discuss our experimental results in the aspects of performance, complexity and limitation.

TABLE IV

RANK RECOGNITION RATES (%) CORRESPONDING THE SMITHSONIAN DATASET CONTAINING 93-CLASSES WITH ROTATED IMAGES

| Rank | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Modified Gabor [21] | 48.66 | 19.44 | 17.29 | 11.47 | 12.10 | 10.04 | 9.68 | 9.68 | 9.95 | 9.32 | 8.24 |
| LBP [35] | 56.09 | 31.09 | 24.55 | 16.58 | 17.65 | 11.29 | 11.38 | 10.39 | 9.68 | 7.44 | 8.06 |
| DC [15] | 54.75 | 34.59 | 25.27 | 22.13 | 17.74 | 14.78 | 13.98 | 12.99 | 11.83 | 11.38 | 9.95 |
| SC [30] | 85.13 | 72.31 | 64.78 | 59.95 | 51.16 | 43.10 | 37.90 | 31.81 | 28.76 | 22.85 | 18.91 |
| IDSC+DP [32] | 89.07 | 80.11 | 71.77 | 64.25 | 56.54 | 46.51 | 42.29 | 37.19 | 30.65 | 27.06 | 22.40 |
| BAC [33] | 83.87 | 75.72 | 68.46 | 55.73 | 44.09 | 35.93 | 28.32 | 27.33 | 22.94 | 18.91 | 16.58 |
| RSICD | **99.28** | **97.49** | **95.79** | **95.25** | **93.55** | **90.41** | **91.49** | **90.23** | **87.46** | **83.78** | **75.72** |

TABLE V

RANK RECOGNITION RATES (%) CORRESPONDING THE SMITHSONIAN DATASET CONTAINING 93-CLASSES WITH ROTATED AND SCALED IMAGES

| Rank | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Modified Gabor [21] | 26.43 | 16.40 | 12.10 | 10.39 | 10.84 | 10.30 | 9.23 | 9.41 | 6.72 | 6.99 | 6.63 |
| LBP [35] | 26.16 | 11.47 | 8.87 | 6.36 | 6.27 | 4.93 | 4.12 | 4.21 | 4.93 | 3.41 | 3.67 |
| DC [15] | 29.75 | 15.86 | 11.02 | 8.33 | 7.97 | 6.27 | 6.54 | 5.73 | 5.56 | 5.38 | 3.94 |
| SC [30] | 92.13 | 81.81 | 73.03 | 63.35 | 55.11 | 48.48 | 44.00 | 36.65 | 33.87 | 29.57 | 27.06 |
| IDSC+DP [32] | 97.58 | 92.83 | 83.96 | 72.85 | 62.01 | 54.84 | 49.73 | 43.28 | 43.37 | 36.83 | 32.08 |
| BAC [33] | 50.54 | 31.54 | 20.52 | 19.27 | 15.50 | 14.52 | 11.20 | 11.11 | 9.95 | 8.60 | 7.44 |
| RSICD | **98.21** | **95.52** | **92.83** | **91.85** | **90.95** | **89.16** | **86.29** | **86.65** | **81.36** | **78.23** | **73.03** |

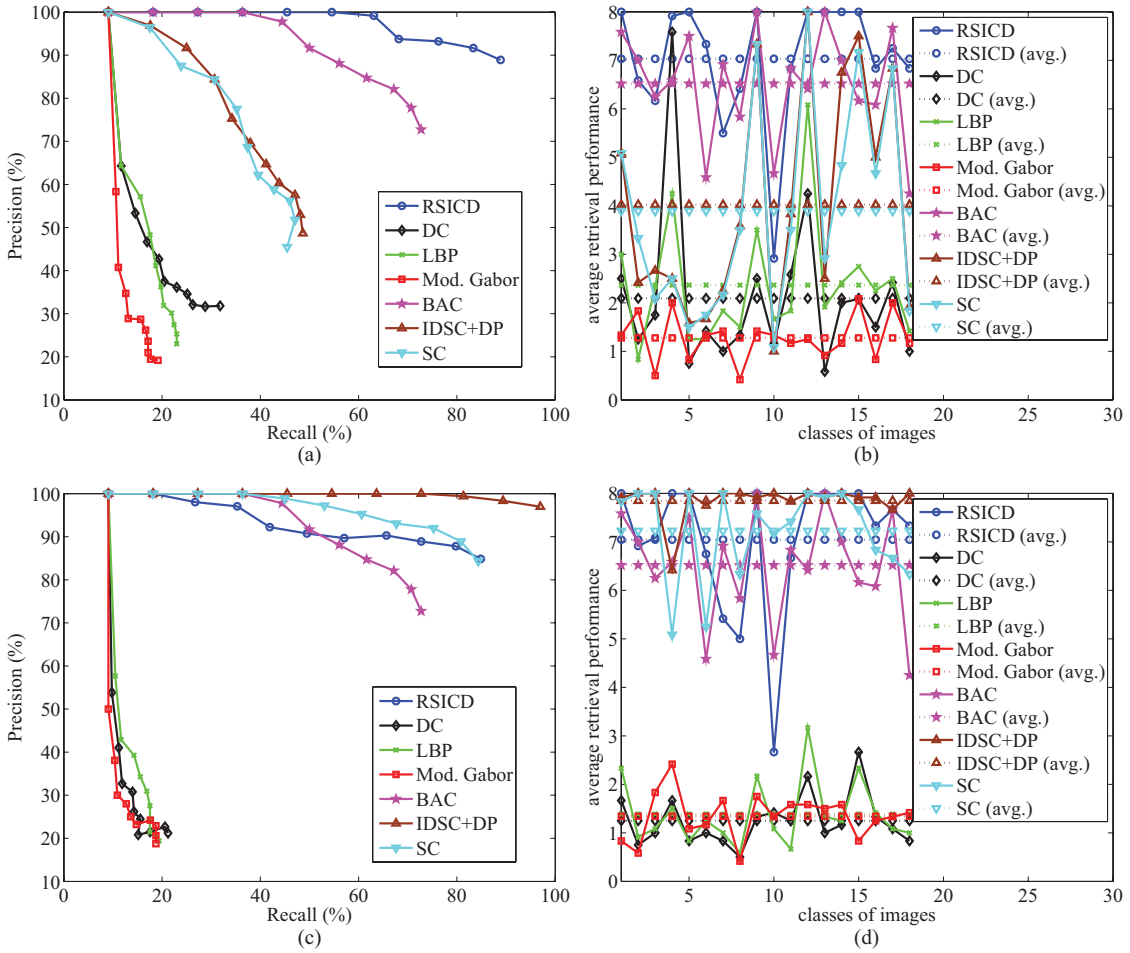



Fig. 9. Results on Kimia dataset. (a) Precision-recall curves. (b) Average retrieval performance of the dataset containing rotated images. (c) Precision-recall curves. (d) Average retrieval performance of the dataset containing rotated and scaled images.

*1) Performance:* For texture-based approaches, the Gabor method [21] extracts features using a modified Gabor filter to achieve independence of orientation and scale in the textures. The LBP [35] is a computationally simple method, but an efficient multiresolution approach based on uniform local binary patterns and nonparametric discrimination of sample and prototype distributions for rotation invariant texture classification. Both methods are designed for texture-based images.

TABLE VI

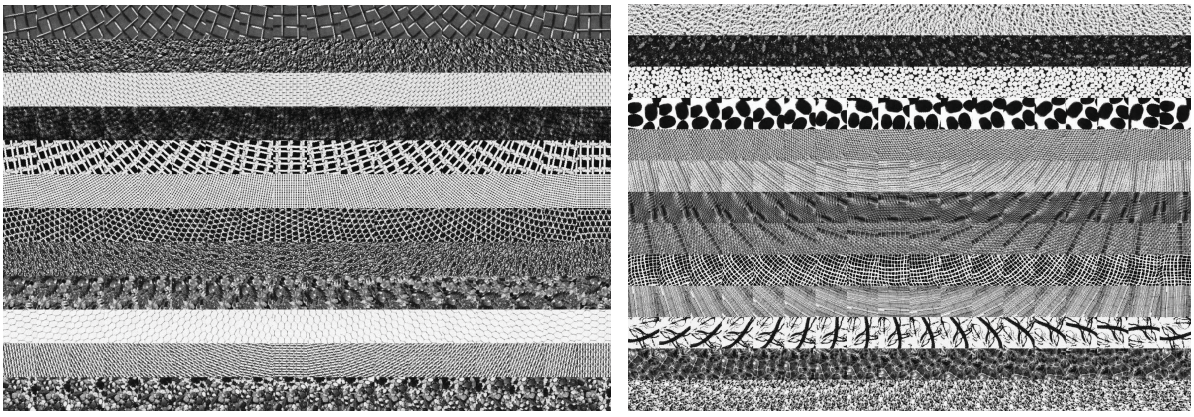
RETRIEVAL RESULTS (RANK RECOGNITION PERCENTAGE RATES) ON KIMIA DATASET CONTAINING ROTATED IMAGES

| Rank | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Modified Gabor [21] | 42.13 | 21.30 | 15.28 | 13.43 | 7.41 | 12.50 | 8.33 | 7.41 | 4.17 | 4.63 | 9.26 |
| LBP [35] | 62.96 | 45.37 | 26.85 | 29.17 | 24.07 | 15.28 | 16.20 | 16.67 | 11.11 | 10.65 | 12.50 |
| DC [15] | 60.65 | 38.43 | 23.61 | 21.76 | 17.59 | 17.59 | 16.20 | 13.43 | 17.59 | 12.04 | 10.19 |
| SC [30] | 86.11 | 66.67 | 59.26 | 48.61 | 40.28 | 32.41 | 28.24 | 27.78 | 25.00 | 15.28 | 12.04 |
| IDSC+DP [32] | 86.57 | 69.91 | 57.41 | 48.61 | 42.13 | 37.50 | 31.94 | 28.24 | 28.24 | 21.30 | 12.04 |
| BAC [33] | **100** | **97.69** | 88.89 | **90.28** | **84.72** | 70.37 | 63.89 | 56.02 | 48.61 | 32.41 | 41.67 |
| RSICD | 98.15 | 92.13 | **89.81** | 88.43 | 82.87 | **84.26** | **81.48** | **86.11** | **81.02** | **69.91** | **62.50** |

TABLE VII

RETRIEVAL RESULTS (RANK RECOGNITION PERCENTAGE RATES) ON KIMIA DATASET CONTAINING ROTATED AND SCALED IMAGES

| Rank | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Modified Gabor [21] | 36.11 | 15.28 | 16.67 | 13.43 | 15.28 | 11.11 | 13.43 | 12.96 | 11.57 | 8.33 | 11.57 |
| LBP [35] | 41.67 | 21.76 | 17.59 | 13.89 | 9.72 | 14.81 | 10.65 | 8.80 | 9.72 | 7.41 | 7.41 |
| DC [15] | 43.52 | 18.52 | 13.89 | 10.19 | 11.57 | 11.11 | 11.57 | 8.80 | 6.48 | 7.41 | 6.48 |
| SC [30] | 99.07 | 98.61 | 95.83 | 95.37 | 92.59 | 84.26 | 82.41 | 74.54 | 75.93 | 67.59 | 56.94 |
| IDSC+DP [32] | **100** | **99.54** | **98.15** | **99.07** | **98.61** | **97.69** | **96.30** | **95.37** | **90.74** | **84.72** | **76.85** |
| BAC [33] | **100** | 97.69 | 88.89 | 90.28 | 84.72 | 70.37 | 63.89 | 56.02 | 48.61 | 32.41 | 41.67 |
| RSICD | 97.22 | 94.44 | 89.81 | 89.81 | 85.19 | 84.72 | 81.94 | 84.26 | 79.17 | 76.39 | 56.94 |



(a) (b)

Fig. 10. Samples images from the 25-class in-plane rotated Brodatz texture database. (a) 1st $\sim$ 12th classes: D01, D04, D06, D19, D20, D21, D22, D24, D28, D34, D52, and D53. (b) 13th $\sim$ 25th classes: D56, D57, D66, D74, D76, D78, D82, D84, D102, D103, D105, D110, and D111.

TABLE VIII

RANK RECOGNITION RATES (%) ON 25-CLASS (DATA SET 1 IN [36]) IN-PLANE ROTATED BRODATZ DATABASE

| Rank | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Modified Gabor [21] | **100** | 80.44 | 81.33 | 57.11 | 57.56 | 43.33 | 44.44 | 38.44 | 40.89 | 35.78 | 34.89 | 33.56 | 34.22 | 27.78 | 26.44 |
| LBP [35] | 79.33 | 67.78 | 56.89 | 49.78 | 46.89 | 45.78 | 46.22 | 43.11 | 41.56 | 39.78 | 38.89 | 39.11 | 37.11 | 33.11 | 30.44 |
| DC [15] | 88.22 | 69.56 | 59.33 | 55.11 | 47.33 | 44.44 | 41.78 | 38.00 | 35.33 | 34.67 | 32.00 | 28.22 | 27.56 | 22.89 | 19.33 |
| RSICD | 92.00 | **86.89** | **85.56** | **82.00** | **76.00** | **75.56** | **71.11** | **63.56** | **57.56** | **53.11** | **52.67** | **46.67** | **45.11** | **42.22** | **38.00** |

As a result, they did not obtain good results on shape-based datasets, which can be seen from the experimental results in Sections IV-A and IV-B.

For shape-based approaches, the SC [30] is a shape matching approach based on correspondences between points on two shapes. The SC descriptor essentially estimates the shape similarity and solves the correspondence problems. The IDSC+DP [32] uses the length of the shortest path within the shape boundary (called inner-distance) to build shape descriptors, which were shown to be robust to articulation in complicated shapes. The BAC [33] extracts features that characterize the geometric relationships between each pair of images. This method was shown to be invariant to articulations and rigid transforms.

TABLE IX
RANK RECOGNITION RATES (%) ON 60-CLASS (DATA SET 3 IN [36]) IN-PLANE ROTATED BRODATZ DATABASE

| Rank | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Modified Gabor [21] | **100** | 71.11 | 72.41 | 40.28 | 42.50 | 34.44 | 34.44 | 27.22 | 30.37 | 27.41 | 24.26 | 21.20 | 22.41 | 19.63 | 18.70 |
| LBP [35] | 76.67 | 59.35 | 43.33 | 38.15 | 31.30 | 30.28 | 27.50 | 25.28 | 25.37 | 25.00 | 22.31 | 18.98 | 21.67 | 17.31 | 18.24 |
| DC [15] | 86.48 | 69.07 | 58.43 | 51.48 | 46.11 | 39.72 | 33.80 | 31.57 | 26.11 | 25.83 | 22.50 | 21.57 | 19.17 | 14.44 | 12.41 |
| RSICD | 93.06 | **86.11** | **82.13** | **75.74** | **70.09** | **66.39** | **62.69** | **54.17** | **46.57** | **44.44** | **42.41** | **40.19** | **35.65** | **32.87** | **31.48** |

TABLE X
COMPUTATION TIME OF DIFFERENT METHODS TO OBTAIN PRECISION-RECALL CURVES SHOWN IN FIG. 6(a)

| | Modified Gabor [21] | LBP [35] | DC [15] | SC [30] | IDSC+DP [32] | BAC [33] | RSICD |
|---|---|---|---|---|---|---|---|
| Execution Time (s) | 10.42 | 13.34 | 383.40 | 1226.34 | 1383.48 | 92.31 | 217.32 |
| Unsupervised Clustering | no | no | yes | no | no | no | yes |

The SC descriptor [30] relies on the correspondences between points on two shapes, while the IDSC+DP descriptor [32] is built based on the normalized inner distance. In practice, SC and IDSC+DP descriptors remain the same for similar shapes with different scales and change significantly for different shapes with different scales. Therefore, introducing scale variations (shown in Fig. 5(b) and Fig. 8(b)) in fact boosts the discriminative power of SC and IDSC+DP features, in that between-class distances are increased due to scale variations while within-class distances remain the same. Hence, SC and IDSC+DP obtained better results on the datasets containing both rotation and scale variations (see Fig. 6, Fig. 9, Table II, Table III, Table IV, Table V, Table VI and Table VII). Furthermore, pixels in the Smithsonian leaf images appear in different grayscales. Without additional preprocessing, the shape contours as well as inner distances-based and point correspondence-based descriptors are sensitive to changes in grayscale or missing pixels (e.g., Fig. 7(a) and (b)). These two reasons explain why SC and IDSC+DP in our experiments perform much better in the Kimia dataset with both rotation and scale variations. The BAC, too, obtained better results on the Kimia dataset. However, it is sensitive to both scale and rotation changes in Smithsonian leaf images. On the other hand, the proposed RSICD does not require any knowledge of the shape contour, and is not sensitive to grayscale changes and missing pixels in an image. In addition, the Smithsonian leaf datasets used in our experiments consist of more directional leaves than isotropic leaves, which are in favor of our assumption on directional images described in Section II-A. Hence, the proposed RSICD obtained good results on the Smithsonian leaf datasets. Finally, from the experimental results, we observe that DC [15] does not give satisfactory performances in both shape-based and texture-based datasets because it uses pixel intensities as features.

*2) Complexity:* We present the relative complexity of all the methods by comparing the computation time required to obtain precision-recall curves for the rotated 18-class Smithsonian datasets in Table X. This table shows both the computation time[5] and whether the unsupervised clustering is provided by each method. Note that the RSCID provides both unsupervised clustering and dictionary learning. As a result, its computation time is higher than some of the other methods. Also, both SC [30] and IDSC+DP [32] require a large amount of time for image retrieval.

*3) Limitation:* We have examined the performance of our method on various shape-based and texture-based datasets. In practice, there may be objects with background clutter. For our method to be effective, the background needs to be removed before applying our algorithm to obtain good retrieval performances. Hence, it may not provide good results on datasets where images contain objects with background clutter. The second limitation of our method is that it does not work so well for texture-based images where there are more within-class variations such as illumination changes, noise, occlusion, variant distances with 3D rotations and spatial shifts. Moreover, for textures where there are no linear trends (e.g., isotropic textures) or inapparent linear trends, the Radon-domain sinogram may no longer be used to accurately capture the direction to give rotation-aligned features.

## V. CONCLUSION

In this paper, we presented a rotation and scale invariant clustering algorithm suitable for applications such as CBIR. We extracted in-plane rotation and scale invariant features of images in the Radon domain. The initial dictionaries are learned through initial clusters that are determined using the Hamming distance between nearest-neighbor sets of each feature pair. With a view to achieve rotation and scale invariance in clustering, the proposed method learns dictionaries and clusters images in the Radon transform domain. We demonstrated the effectiveness of our approach by a series of CBIR experiments on shape-based and texture-based datasets, its robustness to missing pixels, and performance improvements compared to other Gabor-based and shape-based methods.

[5]We conducted our experiments using MATLAB installed in the 64-bit Windows OS on a machine with Intel(R) Core(TM) i5 CPU (2.8 GHz) and 8 GB RAM.

Robustness to within-class variations of texture images is one of the important research directions. We will continue with the feature extraction from the Radon-domain sinogram, via filtering or transformation techniques. We will also consider using local features. One possible way is to divide the texture (or its sinogram) into several patches, from each of which the local feature is extracted. Then we combine all extracted features using an efficient fusion technique, to improve the recognition performance.

In addition, we will modify the Radon-domain sinogram based approach for our algorithm to adapt for textures without apparent line trends, and for isotropic textures. Furthermore, as our method being based on heuristic K-SVD lacks theoretical guarantee on convergence, we will also provide analytical evidences on convergence.

## ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their many valuable comments and suggestions, which significantly improved this paper. The identification of any commercial product or trade name does not imply endorsement or recommendation by NIST.

## REFERENCES

[1] M. Aharon, M. Elad, and A. M. Bruckstein, "The K-SVD: An algorithm for designing of overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[2] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. Huang, and S. Yan, "Sparse representation for computer vision and pattern recognition," *IEEE Proc.*, vol. 98, no. 6, pp. 1031–1044, Jun. 2010.

[3] R. Rubinstein, A. Bruckstein, and M. Elad, "Dictionaries for sparse representation modeling," *IEEE Proc.*, vol. 98, no. 6, pp. 1045–1057, Jun. 2010.

[4] V. M. Patel, T. Wu, S. Biswas, P. J. Phillips, and R. Chellappa, "Dictionary-based face recognition under variable lighting and pose," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 3, pp. 954–965, Jun. 2012.

[5] K. Etemad and R. Chellappa, "Separability-based multiscale basis selection and feature extraction for signal and image classification," *IEEE Trans. Image Process.*, vol. 7, no. 10, pp. 1453–1465, Oct. 1998.

[6] P. J. Phillips, "Matching pursuit filters applied to face identification," *IEEE Trans. Image Process.*, vol. 7, no. 8, pp. 150–164, Aug. 1998.

[7] F. Rodriguez and G. Sapiro, "Sparse representations for image classification: Learning discriminative and reconstructive non-parametric dictionaries," Univ. Minnesota, Twin Cities, MN, USA, Tech. Rep. IMA-2213, Dec. 2007.

[8] K. Huang and S. Aviyente, "Sparse representation for signal classification," in *Proc. Neural Inf. Process. Syst. Conf.*, vol. 19. 2007, pp. 609–616.

[9] Q. Zhang and B. Li, "Discriminative K-SVD for dictionary learning in face recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jan. 2010, pp. 2691–2698.

[10] M. Ranzato, F. Haung, Y. Boureau, and Y. LeCun, "Unsupervised learning of invariant feature hierarchies with applications to object recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.

[11] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Discriminative learned dictionaries for local image analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[12] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Supervised dictionary learning," in *Advance Neural Information Processing Systems*, Ithaca, NY, USA: Cornell Univ. Press, 2008.

[13] L. Bar and G. Sapiro, "Hierarchical invariant sparse modeling for image analysis," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2011, pp. 2397–2400.

[14] V. M. Patel and R. Chellappa, "Sparse representations, compressive sensing and dictionaries for pattern recognition," in *Proc. Asian Conf. Pattern Recognit.*, Nov. 2011, 325–329.

[15] P. Sprechmann and G. Sapiro, "Dictionary learning and sparse coding for unsupervised clustering," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Mar. 2010, pp. 2042–2045.

[16] I. Ramirez, P. Sprechmann, and G. Sapiro, "Classification and clustering via dictionary learning with structured incoherence and shared features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 3501–3508.

[17] E. Elhamifar and R. Vidal, "Sparse subspace clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 2790–2797.

[18] S. Rao, R. Tron, R. Vidal, and Y. Ma, "Motion segmentation via robust subspace separation in the presence of outlying, incomplete, or corrupted trajectories," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[19] M. Soltanolkotabi and E. J. Candes, *A geometric analysis of subspace clustering with outliers*. Ithaca, NY, USA: Cornell Univ. Press, 2011.

[20] G. Haley and B. Manjunath, "Rotation-invariant texture classification using a complete space-frequency model," *IEEE Trans. Image Process.*, vol. 8, no. 2, pp. 255–269, Feb. 1999.

[21] C. S. Sastry, M. Ravindranath, A. K. Pujari, and B. Deekshatulu, "A modified Gabor function for content based image retrieval," *Pattern Recognit. Lett.*, pp. 293–300, 2007.

[22] M. Do and M. Vetterli, "Rotation invariant texture characterization and retrieval using steerable wavelet-domain hidden Markov models," *IEEE Trans. Multimedia*, vol. 4, no. 4, pp. 517–527, Dec. 2002.

[23] C.-M. Pun and M.-C. Lee, "Log-polar wavelet energy signatures for rotation and scale invariant texture classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 5, pp. 590–603, May 2003.

[24] H. Wersing, J. Eggert, and E. Körner, "Sparse coding with invariance constraints," in *Proc. Int. Conf. Artif. Neural Netw.*, 2003, pp. 385–392.

[25] Y. Tomokusa, M. Nakashizuka, and Y. Iiguni, "Sparse image representations with shift and rotation invariance constraints," in *Proc. Int. Symp. Intell. Signal Process. Commun. Syst.*, Jan. 2009, pp. 256–259.

[26] Q. Barthelemy, A. Larue, A. Mayoue, D. Mercier, and J. I. Mars, "Shift and 2-D rotation invariant sparse coding for multivariate signals," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 1597–1611, Apr. 2012.

[27] L. Bar and G. Sapiro, "Hierarchical dictionary learning for invariant classification," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Mar. 2010, pp. 3578–3581.

[28] P. Ungureanu, E. David, and L. Goras, "On rotation invariant texture classification using two-grid coupled CNNs," in *Proc. Seminar Neural Netw. Appl. Electr. Eng.*, Sep. 2006, pp. 33–36.

[29] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1615–1630, Oct. 2005.

[30] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 509–522, Apr. 2002.

[31] T. B. Sebastian, P. N. Klein, and B. B. Kimia, "Recognition of shapes by editing their shock graphs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 5, pp. 550–571, May 2004.

[32] H. Ling and D. W. Jacobs, "Shape classification using the inner-distance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 2, pp. 286–299, Feb. 2007.

[33] S. Biswas, G. Aggarwal, and R. Chellappa, "Robust estimation of albedo for illumination-invariant matching and shape recovery," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 2, pp. 884–899, Mar. 2009.

[34] Y.-C. Chen, C. S. Sastry, V. M. Patel, P. J. Phillips, and R. Chellappa, "Rotation invariant simultaneous clustering and dictionary learning," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Mar. 2012, pp. 1053–1056.

[35] T. Ojala, P. Matti, and M. Topi, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[36] K. Jafari-Khouzani and H. Soltanian-Zadeh, "Radon transform orientation estimation for rotation invariant texture analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 6, pp. 1004–1008, Jun. 2005.

[37] J. A. Tropp and S. J. Wright, "Computational methods for sparse solution of linear inverse problems," *Proc. IEEE*, vol. 98, no. 6, pp. 948–958, Jun. 2010.

[38] E. van den Berg and M. P. Friedlander, "Probing the pareto frontier for basis pursuit solutions," *Soc. Ind. Appl. Math. J. Sci. Comput.*, vol. 31, no. 2, pp. 890–912, 2008.

[39] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[40] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *Soc. Ind. Appl. Math. J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1998.
[41] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. Conf. Rec. 27th Asilomar Conf. Signals, Syst. Comput.*, pp. 40–44, 1993.
[42] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.
[43] N. Shroff, P. Turaga, and R. Chellappa, "Video précis: Highlighting diverse aspects of videos," *IEEE Trans. Multimedia*, vol. 12, no. 8, pp. 853–868, Dec. 2010.
[44] P. Brodatz, *Texture: A Photographic Album for Artists and Designers*. New York, USA: Dover, 1966.

**Yi-Chen Chen** (S'08) received the B.S. degree in electrical engineering from National Tsing Hua University, Hsinchu, Taiwan, and the M.S. degree in communication engineering from National Taiwan University, Taipei, Taiwan. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD, USA.

His current research interests include computer vision and pattern recognition, particularly face recognition, salient view selection, clustering, and content-based image retrieval.

**Challa S. Sastry** (M'10) received the Ph.D. degree in applied mathematics from the Indian Institute of Technology Kanpur, Kanpur, India, in 2003.

He is currently with the Indian Institute of Technology Hyderabad, Hyderabad, India. His current research interests include wavelets, computed tomography, and compressive sampling theory.

**Vishal M. Patel** (M'01) received the B.S. degree in electrical engineering and applied mathematics (Hons.) and the M.S. degree in applied mathematics from North Carolina State University, Raleigh, USA, in 2004 and 2005, respectively, and the Ph.D. degree in electrical engineering from the University of Maryland, College Park, USA, in 2010.

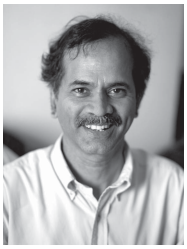
He is currently a Member of the Research Faculty with the University of Maryland Institute for Advanced Computer Studies, College Park. His current research interests include signal processing, computer vision, and pattern recognition with applications in biometrics and imaging.

Dr. Patel was a recipient of the ORAU postdoctoral fellowship in 2010. He is a member of Eta Kappa Nu, Pi Mu Epsilon, and Phi Beta Kappa.

**P. Jonathon Phillips** (M'96–SM'06–F'10) received the Ph.D. degree in operations research from Rutgers University, New Brunswick, NJ, USA.

He is currently a Leading Researcher involved in research on computer vision, face recognition, biometrics, and human identification. He is with the National Institute of Standards and Technology, Gaithersburg, MD, USA, where he is involved in directing challenge problems and evaluations in face recognition and biometrics. He has contributed his efforts in the Multiple Biometrics Evaluation 2010, the Iris Challenge Evaluations, the Face Recognition Vendor Test (FRVT) 2006, and the Face Recognition Grand Challenge and FERET. From 2000 to 2004, he was involved in research with the Defense Advanced Research Projects Agency as the Program Manager for the Human Identification at a Distance program. His contribution is reported in print media of record including The New York Times and The Economist. He has appeared on NPR's ScienceFriday. In an Essential Science Indicators analysis of face recognition publication over the past decade, his published paper ranks at 2 by total citations and 1 by citations per paper.

Dr. Phillips was a recipient of the Gold Medal from the Department of Commerce for his contribution on the FRVT 2002. He was an Associate Editor of the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE and a Guest Editor of an issue of the *Proceedings of the IEEE* on biometrics from 2004 to 2008. He is a Fellow of the IAPR. He was the Test Director of the FRVT 2002.



**Rama Chellappa** (S'78–M'79–SM'83–F'92) received the B.E. (Hons.) degree in electronics and communication engineering from the University of Madras, Chennai, India, and the M.E. (with Distinction) degree from the Indian Institute of Science, Bangalore, India, in 1975 and 1977, respectively, and the M.S.E.E. and Ph.D. degrees in electrical engineering from Purdue University, West Lafayette, IN, USA, in 1978 and 1981, respectively.

He was a Faculty Member with the Department of of Electrical Engineering Systems, University of Southern California (USC) from 1981 to 1991. Since 1991, he has been a Professor of Electrical and Computer Engineering and an Affiliate Professor of Computer Science with the University of Maryland (UMD), College Park, MD. He is also affiliated with the Center for Automation Research, and is a Permanent Member with the Institute for Advanced Computer Studies. In 2005, he was named a Minta Martin Professor of Engineering. He holds three patents. His current research interests include face recognition, clustering and video summarization, 3-D modeling from video, image and video-based recognition of objects, events and activities, dictionary-based inference, compressive sensing, domain adaptation, and hyper spectral processing.

Prof. Chellappa was a recipient of the NSF Presidential Young Investigator Award, four IBM Faculty Development Awards, the Excellence in Teaching Award from the School of Engineering at USC, two Paper Awards from the International Association of Pattern Recognition (IAPR), the K.S. Fu Prize from IAPR, the Society, Technical Achievement, and Meritorious Service Awards from the IEEE Signal Processing Society, the Technical Achievement and Meritorious Service Awards from the IEEE Computer Society, the Outstanding Innovator Award from the Office of Technology Commercialization, the Outstanding GEMSTONE Mentor Award from the Honors College, and the Outstanding Faculty Research Award and the Poole and Kent Teaching Award for Senior Faculty from the College of Engineering. At UMD, he was elected as a Distinguished Faculty Research Fellow, and as a Distinguished Scholar-Teacher. In 2010, he was recognized as an Outstanding ECE by Purdue University. He is a Fellow of the IAPR, the Optical Society of America, and the American Association for the Advancement of Science. He was the Editor-in-Chief of the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE. He was the General or the Technical Program Chair of several IEEE international and national conferences and workshops. He is a Golden Core Member of the IEEE Computer Society and was a Distinguished Lecturer of the IEEE Signal Processing Society. He has completed a two-year term as the President of the IEEE Biometrics Council.