Radon Representation-Based Feature Descriptor for Texture Classification

Guangcan Liu, Zhouchen Lin, *Senior Member, IEEE*, and Yong Yu

***Abstract*—In this paper, we aim to handle the intraclass variation resulting from the geometric transformation and the illumination change for more robust texture classification. To this end, we propose a novel feature descriptor called Radon representation-based feature descriptor (RRFD). RRFD converts the original pixel represented images into Radon-pixel images by using the Radon transform. The new Radon-pixel representation is more informative in geometry and has a much lower dimension. Subsequently, RRFD efficiently achieves affine invariance by projecting an image (or an image patch) from the space of Radon-pixel pairs onto an invariant feature space by using a *ratiogram*, i.e., the histogram of ratios between the areas of triangle pairs. The illumination invariance is also achieved by defining an illumination invariant distance metric on the invariant feature space. Comparing to the existing Radon transform-based texture features, which only achieve rotation and/or scaling invariance, RRFD achieves affine invariance. The experimental results on CUReT show that RRFD is a powerful feature descriptor that is suitable for texture classification.**

***Index Terms*—Image classification, image texture analysis.**

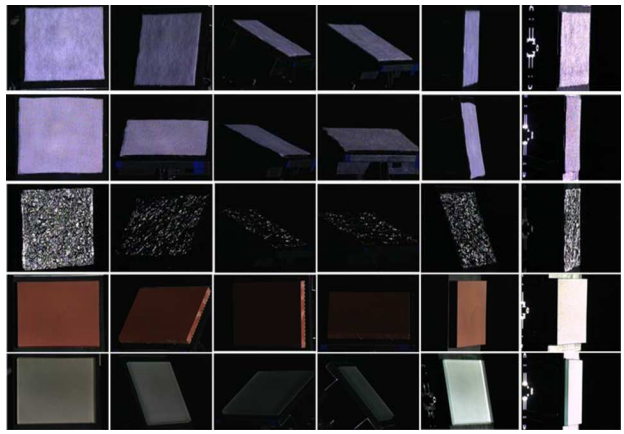


Fig. 1. Changing viewpoint and illumination can have a dramatic impact on the appearance of a material. These images are taken from the CUReT database. Each row is of the same class.

I. Introduction

TEXTURE analysis is important for the interpretation and understanding of real-world visual patterns. It has been applied to many practical vision systems, such as biomedical imaging, ground classification, segmentation of satellite imagery, and pattern recognition. The automated analysis of image textures has been a topic of extensive research in the past decades. Existing features and techniques for modeling textures include graylevel co-occurrence matrices [1], Gabor transforms [2], [3], bidirectional texture function [4], [5], local binary patterns [6]–[8], random fields [9], [10], autoregressive models [11], wavelet-based features [12], [13], textons [14]–[17], affine adaption [18]–[20], fractal dimension [21]–[24], SIFT [20], [25], and some invariant feature descriptors such as Zernike moments [26], circular harmonic functions [27], and others [28]–[31]. Recently, some researchers have considered using the Radon transform [32] for texture analysis. In [33], the Radon transform is used to achieve rotation invariance, and [34] considers rotation and scale invariance. In this work, our approach is also based on the Radon transform. However, our approach achieves affine invariance in both geometry transform and illumination transform, which is much more challenging.

Despite lots of work on this topic, some problems remain unsolved. First, as shown in Fig. 1, the illumination variation can have a dramatic impact on the appearance of a material and the existing methods cannot handle under-illuminated images, as in the third and the fourth columns of Fig. 1. One evidence is that previous researchers always discard such unfavorable images in their experiments [17], [19], [20], [23], [24], [35]. It is, therefore, desirable to overcome the difficulty caused by illumination variance. Second, eliminating the interclass confusion and the intraclass variation *simultaneously* is usually difficult: reducing the interclass confusion may produce more false-positives, which goes against reducing the intraclass variation, and vice versa. It is, therefore, crucial to develop texture features that are not only discriminative across many classes, but also invariant to key transformations, mainly including the geometric affine transformation and the illumination change. Third, many recently developed applications demand more robust and effective texture features, e.g., the construction of an appearance model in object recognition requires the clustering of local image patches into visual words [36], which essentially is an *unsupervised* texture clustering problem that needs the texture descriptors to be simple (have few parameters to tune) and robust (perform well and stably).

In this paper, we propose a novel feature descriptor, called radon representation-based feature descriptor (RRFD), that can handle unfavorable change in illumination conditions, such as underexposure, and also the variance caused by the geometric

Manuscript received June 15, 2008; revised December 10, 2008. First published March 16, 2009; current version published April 10, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Ying Wu.

G. Liu and Y. Yu are with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, 200240, China (e-mail: roth@sjtu.edu.cn; yyu@apex.sjtu.edu.cn).

Z. Lin is with the Visual Computing Group of Microsoft Research Asia, Beijing, 100190, China (e-mail: zhoulin@microsoft.com).

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.2009.2013072

affine transformation. RRFD achieves both the geometric affine transformation and the illumination change invariance in three steps. First, RRFD converts original pixel represented images into Radon-pixel images by using the Radon transform [32]. The new Radon representation is more informative in geometry and has a much lower dimension. Second, RRFD projects an image from the space of Radon-pixel pairs onto a feature space by collecting the histogram of ratios between the areas of triangle pairs,[1] which we call *ratiogram*. As the ratio of areas is invariant up to the affine group, for a given image, RRFD produces a feature vector that is affine invariant in geometry, which means that the variations resulting from geometric affine transformations can be reduced. Unlike traditional feature vectors which are points in R^n (Euclidean space), there is an l-variate[2] Gaussian distribution for each dimension of the RRFD vector. Since all the information, including the color if available, of an image is kept, the feature vector has high discriminability. Finally, we define an illumination invariant distance metric on the feature space such that the affine invariance in illumination is also achieved. With these pairwise distances given, we compute a kernel matrix and use kernel consistent learning algorithms (e.g., LDA [37] and SVM [38]) to perform texture classification. Compared to previous approaches, including the Radon transform-based ones [33], [34], [39], RRFD has several advantages.

- It is able to handle the affine transformation in the geometry of texture.
- It is able to handle unfavorable change in illumination conditions, e.g., underexposure.
- It is easy to use. There are only two parameters in RRFD and neither of them need careful adjustment.
- It is simple and robust. It yields satisfactory results in our experiments.

One may have noticed that RRFD handles an image in a *global* way. Hence, it cannot detect local features of interest from the image. RRFD is a general descriptor whose goal is to produce effective feature vectors for any image patterns given. When applied to complex images that consist of several different textures, instead of being applied to the whole image, RRFD should be used to describe a local image region or window.

The rest of this paper is organized as follows. Section II introduces the details of RRFD, Section III demonstrates the experimental results and Section IV concludes this paper.

II. Radon Representation-Based Feature Descriptor

Our RRFD is outlined in Fig. 2. In the following, we describe details of each step.

A. Radon Transform

The 2-D Radon transform (Fig. 3) [32] is the integral transform that computes the integral of a function along straight lines. Every straight line can be represented as

$$(X(t), Y(t)) = t(\sin\alpha, -\cos\alpha) + s(\cos\alpha, \sin\alpha)$$

where s is the signed distance from the origin to the line and α is the angle between the normal of the line and the x axis. The

[1]A pair of Radon-pixels correspond to a pair of triangles.

[2]$l = 6$ for three-channel color images and $l = 2$ for grayscale images.

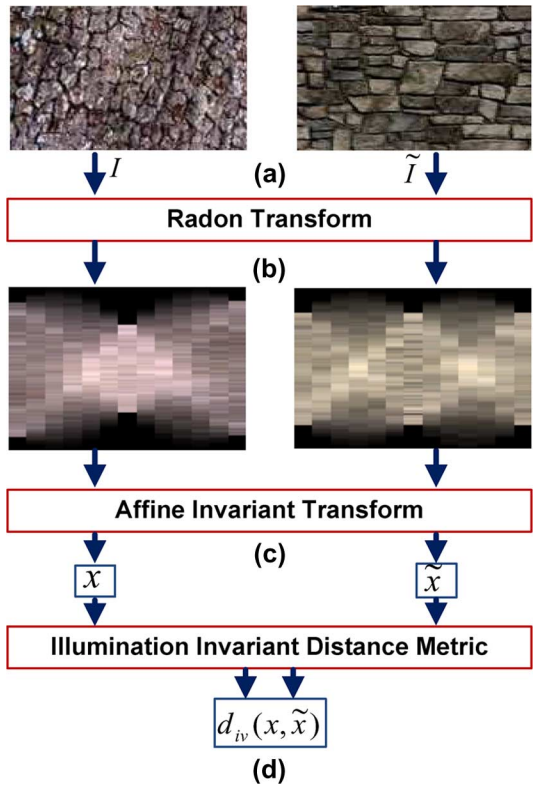


Fig. 2. Illustration of the RRFD. Given two texture images, (a) we want to compute a distance and (d) measure the similarity between them. To this end, we first convert each image into Radon-pixel images (b) by the Radon transform [32]. As one Radon-pixel actually indicates a line segment in the original image and a pair of Radon-pixels correspond to four triangles, there are two affine invariants associated to a pair of Radon-pixels. We, therefore, define a fast affine invariant transform on the Radon-pixel image and each image is then transformed into a vector of an m-dimensional vector space (c). The attributes of the vector are modeled by multivariate Gaussian distributions, i.e., $x = (N_1(\mu_1, \Sigma_1), \ldots, N_m(\mu_m, \Sigma_m))^T$. Finally, we define an illumination invariant distance metric on the vector space to measure the similarity between texture images.

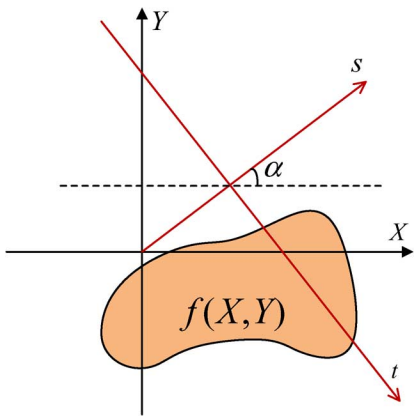


Fig. 3. Illustration of the Radon transform.

Radon transform of a function $f(X, Y)$ on the plane is defined by

$$R(f)(\alpha, s) = \int_{-\infty}^{+\infty} f(X(t), Y(t))\, dt.$$

The Radon transform is a special case of image projection operations [32]. It has found wide applications in many areas such as tomographic reconstruction. It has also been applied to many computer vision areas, such as image segmentation, structural extraction by projections, determining the orientation of an object, recognition of Arabic characters [40] and 1-D processing, filtering and restoration of images [41], [42]. Al-Shaykh and Doherty [39] used the Radon transform for invariant image analysis. By combing the Radon transform and singular value decomposition (SVD), they developed an efficient invariant analysis method. In this work, we have four major reasons to adopt the Radon transform. First, as in [39], the Radon-pixel image [Fig. 2(b)] brings a large advantage to achieving global geometric affine invariance. This is because the Radon-pixel image has more geometric information than the original pixel image. It can be seen that *one* Radon-pixel corresponds to a line segment which needs *two* pixels in the original image to describe. Second, a single Radon-pixel contains the information of a line segment in the original image. This property makes Radon-pixels more robust to image noise and it also fits our motivation of using global features to describe textures. Third, the dimension of Radon-pixel representation is much lower than that of the original image. For an n-pixel image, the number of Radon-pixels is about $O(\sqrt{n})$. Finally, the Radon transform is invertible. In principle, the original image can be roughly recovered from its Radon-pixel image. The invertibility is the chief characteristic that distinguishes the Radon transform from other transformations such as SIFT. In summary, the Radon transform converts a pixel represented image into an equivalent, lower dimensional and more geometrically informative Radon-pixel image, which is a good basis of defining invariant features.

B. Affine Invariant Feature Transform

To achieve the affine invariance, we want to find a projection from the image space to a feature space so that the projection is invariant up to the affine group. We achieve this in three steps:

1) Selecting the Observation Space of an Image: This step plays the role of feature selection. It is important because an inappropriate observation space will result in ineffective features. For example, if we view an image as a set of *single* pixels, then we can only obtain a 1-D affine invariant feature space and, therefore, can only obtain a single scalar to describe an image. Under the affine transformation, to ensure the discriminability of features, one needs to consider at least pixel *quadruples* (four-pixel groups), which causes enormous computation. However, we just need to consider Radon-pixel *pairs* in the Radon-pixel representation, as every Radon pixel corresponds to all the pixels on the corresponding line segment in the original image. Let an image I be represented by a Radon-pixel image $\{r_1, \ldots, r_k\}$. The observation space is then a set of Radon-pixel pairs $I_{rp} = \{(r_i, r_j)\}$. Since for an n-pixel image, the number of Radon-pixels is $O(\sqrt{n})$, the dimension of I_{rp} thus is $O(n)$.

2) Generating the Affine Invariant Feature Vector: Given an image represented by a set of Radon-pixel pairs $I_{rp} = \{(r_i, r_j)\}$, RRFD generates an m-dimensional feature vector

$$x = \left([(r_i, r_j)]_1, \ldots, [(r_i, r_j)]_m\right)^T \tag{1}$$

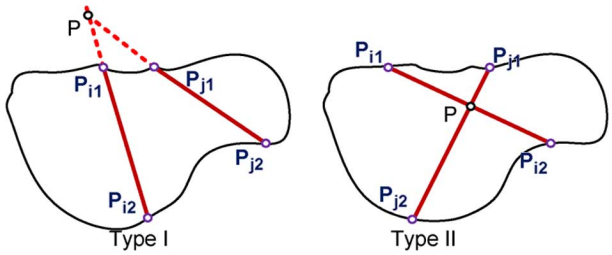


Fig. 4. Two types of Radon-pixel pairs. For type I pairs, the corresponding line segments in the original pixel image have intersection points outside them. For type II pairs, the intersection points are inside them. As the area is a relative invariant under the affine transformation group, the quotient of the areas of any two triangles is invariant. A pair of Radon-pixels, therefore, result in two invariants.

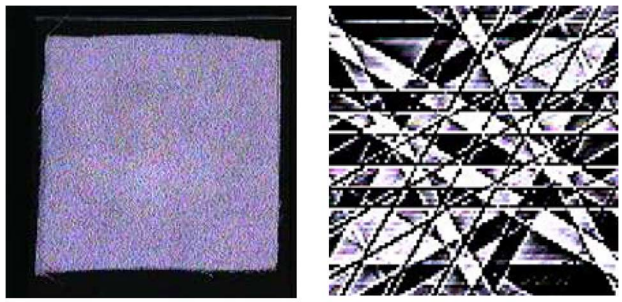


Fig. 5. Example of the information contained in *one* RRFD key. (a) The original texture image. (b) The image recovered from one RRFD key which is a collection of Radon-pixels that belong to a cluster. In this experiment, we set $\Delta\alpha = 30^\circ$ and $\Delta iv = 0.1$.

where $[(r_i, r_j)]_k$ $(k = 1 \cdots m)$ denotes a *cluster* of Radon-pixel pairs. To make x affine invariant, we may find the invariants under the affine transformation and break them into m bins, each of which determines a cluster. Namely, the feature vector is formed by collecting a histogram of the affine invariants. In general, it is hard and unnecessary to find *all* invariants in practice. We just need to find sufficient invariants to determine part of the feature space. For a Radon-pixel pair (r_i, r_j) whose ends in the original pixel image are P_{i1}, P_{i2}, P_{j1} and P_{j2} (Fig. 4), respectively, there are two invariants under the affine transformation

$$iv_1 = \frac{|PP_{i1}P_{j1}|}{|PP_{i2}P_{j2}|} \quad \text{and} \quad iv_2 = \frac{|PP_{i1}P_{j2}|}{|PP_{i2}P_{j1}|} \tag{2}$$

where $|.|$ denotes the area of a triangle. As the order of two triangles is unimportant, we assume that $0 < iv_1 \leq iv_2 \leq 1$. Moreover, as shown in Fig. 4, the intersection type is also preserved by affine transformation. This can be embodied by the above two invariants by using oriented area instead, i.e., $-1 \leq iv_1 \leq iv_2 \leq 1$. By breaking the interval $[-1, 1]$ into bins $[-1, -1+\Delta iv], [-1+\Delta iv, -1+2\Delta iv], \ldots, [1-\Delta iv, 1]$, where Δiv is the bin size, we can have a finite-dimensional vector by putting the Radon-pixel pairs whose invariants locate in the same bin into the same cluster. As the bins are constructed by using the affine invariants, this vector is naturally invariant up to the affine transformation. The basis of the feature vector is only dependent on the bin size Δiv. So, different images are projected onto the same feature space. In practice, because it

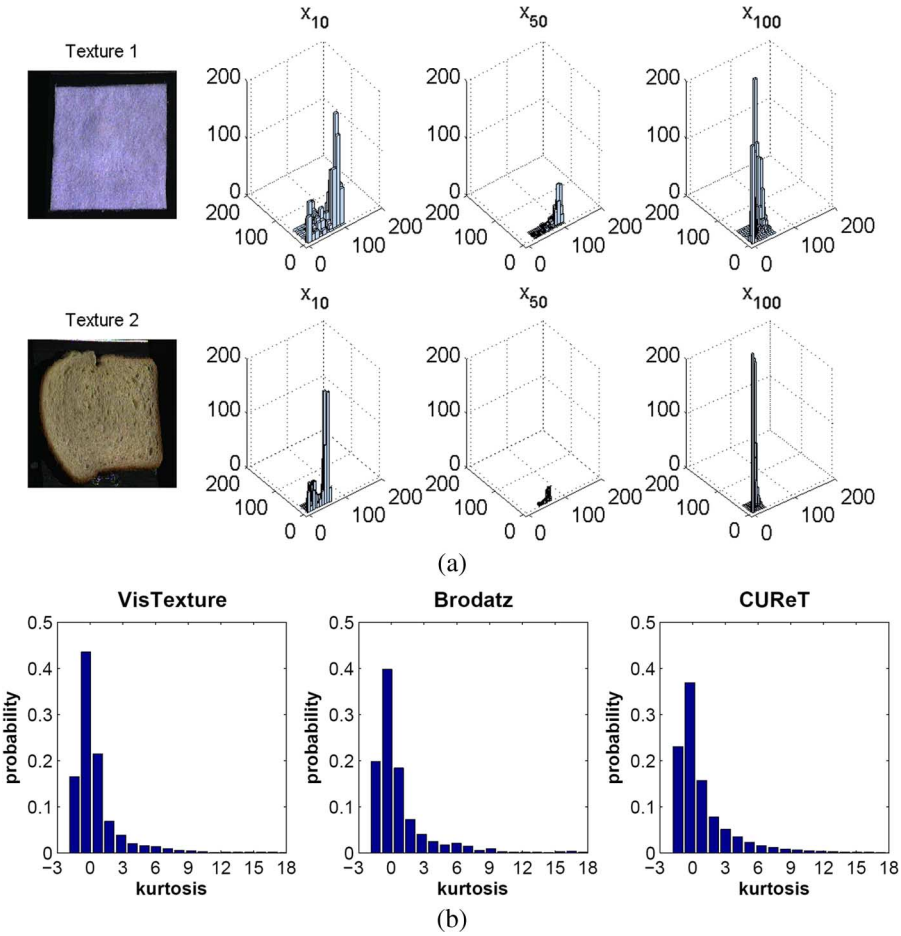


Fig. 6. (a) Examples of 10 × 10 histograms of RRFD keys. Each row is from the same texture. Each column is at the same dimension. For visualization, the original 6-D vectors have been projected to be 2-D by PCA. (b) The distribution of the kurtosis of the vectors sampled from the three databases we used in this paper. There are 62,568 samples for VisTexture [43] (also known as MIT Texture), 24,024 samples for Brodatz [44] and 162,360 samples for CUReT. It can be seen that it is suitable to use a Gaussian distribution to model the RRFD keys because the majority of the kurtosis is 0. In these experiments, we set $\Delta\alpha = 30^\circ$ and $\Delta iv = 0.1$.

is hard to acquire the boundary of a texture, we simply regard the image rectangle as the boundaries. So, like previous affine invariant approaches, RRFD only achieves approximate affine invariance.

3) Describing the Feature Vector: Each dimension of the feature vector x is a cluster of Radon-pixel pairs $[(r_i, r_j)]_k$ $(k = 1, \cdots, m)$, which we call an *RRFD key* (Fig. 5). It is inconvenient to use such feature vectors for texture classification. So we need to reformat them such that it is easy to measure the distance or similarity between feature vectors. For three-channel images (such as RGB-color images), as a Radon-pixel contains three scalars, for each Radon-pixel pair we compute a 6-D vector $(k_1, \ldots, k_6)$ as follows:[3]

$$k_1 = \frac{1}{2}\left|R(r_i) - R(r_j)\right|, \quad k_2 = \frac{1}{2}\left|G(r_i) - G(r_j)\right|$$
$$k_3 = \frac{1}{2}\left|B(r_i) - B(r_j)\right|, \quad k_4 = \frac{1}{2}\left|R(r_i) + R(r_j)\right|$$
$$k_5 = \frac{1}{2}\left|G(r_i) + G(r_j)\right|, \quad k_6 = \frac{1}{2}\left|B(r_i) + B(r_j)\right|$$

[3]For grayscale images, an RRFD key is a set of 2-D vectors that can be computed in a similar way.

where $R(\cdot)$, $G(\cdot)$, and $B(\cdot)$ are the red, the green and the blue intensity values of a Radon-pixel, respectively. We define the above six quantities because the feature values should be independent of the permutation of r_i and r_j and these six quantities are the simplest invariants under permutation. As shown in Fig. 6, it is suitable to use a multivariate Gaussian to fit the *distribution* of the vector $(k_1, \ldots, k_6)$ for every RRFD key. Therefore, the RRFD feature vector of a texture image is represented by an m-dimensional Gaussian distribution vector, i.e.,

$$x = (N_1(\mu_1, \Sigma_1), \ldots, N_m(\mu_m, \Sigma_m))^T \tag{3}$$

where μ_i and Σ_i are the mean and the covariance matrix of a 6-variate Gaussian distribution, respectively.

C. Illumination Invariant Distance Metric

Modeling the illumination change is difficult because it is related to both the lighting condition and the material reflection property. However, from a global view of a texture, it is acceptable to only consider an affine model $I \rightarrow sI + t$ with two parameters s (scale) and t (translation) [19], [23]. Fig. 7 shows an example of applying the affine model to a color texture. It can

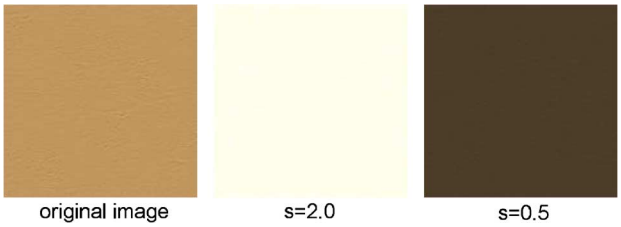


Fig. 7. Color texture images are sensitive to the illumination change. The second and the third images are obtained by multiplying a scale s to the first image.

be seen that considering the illumination change is very important. As mentioned in [16], previous approaches often handle this problem using some kinds of normalization. Obviously, the impact of the scale s can be eliminated by normalizing the intensities of an image to be summed to one. However, such normalization will change the image information and lose many useful features. In our approach, the illumination invariance is achieved by designing a special distance metric.

For simplicity, we start with a distance metric without considering the change of illumination. Given two RRFD vectors x and $\tilde{x}$ computed as (3), we compute their distance by

$$d(x,\tilde{x}) = \sum_{i=1}^{m} J(N_i, \tilde{N}_i)$$

where $J(\cdot,\cdot)$ is the Jeffrey divergence, i.e., the symmetric version of the KL divergence: $J(N_i,\tilde{N}_i) = KL(N_i|\tilde{N}_i) + KL(\tilde{N}_i|N_i)$. With the Gaussian model in (3), the distance can be computed as

$$d(x,\tilde{x}) = \frac{1}{2}\sum_{i=1}^{m}(\mu_i-\tilde{\mu}_i)^T\left(\Sigma_i^{-1}+\tilde{\Sigma}_i^{-1}\right)(\mu_i-\tilde{\mu}_i) + \frac{1}{2}\sum_{i=1}^{m} Tr\left(\Sigma_i\tilde{\Sigma}_i^{-1}+\Sigma_i^{-1}\tilde{\Sigma}_i\right) - ml \quad (4)$$

where l is the number of variables in the Gaussian distribution. This distance is a standard metric as it satisfies positive definiteness, symmetry and the triangle inequality.

Now consider that an image I is taken under a different illumination and becomes $I_{\{s,t\}} = sI + t$. The Gaussian distribution $N_i(\mu_i,\Sigma_i)$, therefore, becomes $N_i(s\mu_i + te, s^2\Sigma_i)$, where e is an l-dimensional vector with all ones. For two observed images $I_{\{s,t\}}$ and $\tilde{I}_{\{\tilde{s},\tilde{t}\}}$, their distance should be $d_{s,\tilde{s},t,\tilde{t}}(x,\tilde{x})$. Replacing μ_i, $\tilde{\mu}_i$, Σ_i and $\tilde{\Sigma}_i$ by $s\mu_i + te$, $\tilde{s}\tilde{\mu}_i + \tilde{t}e$, $s^2\Sigma_i$ and $\tilde{s}^2\tilde{\Sigma}_i$ in (4), respectively, one can see that $d_{s,\tilde{s},t,\tilde{t}}(x,\tilde{x})$ only depends on two variables: $Ds = s/\tilde{s}$ and $\Delta t = t - \tilde{t}$, i.e.,

$$d_{s,\tilde{s},t,\tilde{t}}(x,\tilde{x}) = d_{Ds,\Delta t}(x,\tilde{x}).$$

Although the illumination conditions are unknown and it is difficult or impossible to estimate the parameters for each image, we can achieve the illumination invariance by minimizing $d_{Ds,\Delta t}(x,\tilde{x})$, i.e., we compute an illumination invariant distance by

$$d_{iv}(x,\tilde{x}) = \min_{Ds,\Delta t} d_{Ds,\Delta t}(x,\tilde{x}) \quad (5)$$

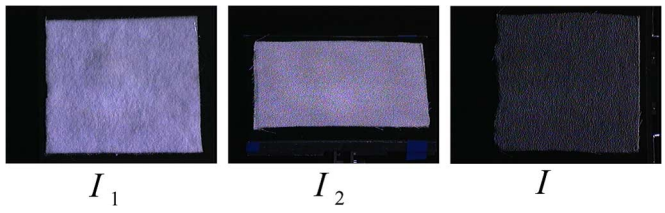


Fig. 8. Effectiveness of using the illumination invariant distance. Images I_1 and I_2 are from two different texture classes, and image I is from the same class as I_2. Without illumination considered, we get $d(I,I_1) = 9613.62$ and $d(I,I_2) = 11119.52$ and may obtain the wrong classification result. However, using the illumination invariant distance metric, we get $d_{iv}(I,I_1) = 7692.05$ and $d_{iv}(I,I_2) = 3926.15$, which correctly helps us to judge that I is closer to I_2 than I_1. The images are from the CUReT database. In this experiment, we set $\Delta\alpha = 15^\circ$ and $\Delta iv = 0.1$.

which means that we compute the distance between two textures I and $\tilde{I}$ after finding the best match between their illuminations. To minimize (5) one just needs to minimize a one-variable function of Ds, i.e.,

$$d_{iv}(x,\tilde{x}) = \min_{Ds} f(Ds)$$

where

$$\begin{aligned} f(Ds) &= \frac{1}{2}(Ds)^2\sum_{i=1}^{m}\left(Tr\left(\Sigma_i\tilde{\Sigma}_i^{-1}\right)+\mu_i^T\tilde{\Sigma}_i^{-1}\mu_i\right) \\ &+ \frac{1}{2(Ds)^2}\sum_{i=1}^{m}\left(Tr\left(\Sigma_i^{-1}\tilde{\Sigma}_i\right)+\tilde{\mu}_i^T\Sigma_i^{-1}\tilde{\mu}_i\right) \\ &- (Ds)\sum_{i=1}^{m}\mu_i^T\tilde{\Sigma}_i^{-1}\tilde{\mu}_i - \frac{1}{(Ds)}\sum_{i=1}^{n}\tilde{\mu}_i^T\Sigma_i^{-1}\mu_i \\ &- \frac{1}{2}\sum_{i=1}^{m}\frac{\left(e^T\left(\frac{1}{(Ds)}\Sigma_i^{-1}+(Ds)\tilde{\Sigma}_i^{-1}\right)((Ds)\mu_i-\tilde{\mu}_i)\right)^2}{e^T\left(\Sigma_i^{-1}+(Ds)^2\tilde{\Sigma}_i^{-1}\right)e} \\ &+ \frac{1}{2}\sum_{i=1}^{m}\left(\mu_i^T\Sigma_i^{-1}\mu_i+\tilde{\mu}_i^T\tilde{\Sigma}_i^{-1}\tilde{\mu}_i\right) - ml \end{aligned}$$

as Δt can be easily found as a function of Ds by letting $\partial(d_{Ds,\Delta t})/\partial(\Delta t) = 0$.[4] The final solution can be found numerically.

Fig. 8 shows the effectiveness of using the illumination invariant distance. It can be seen that the invariant distance is effective in handling large illumination change. One should notice that the distance computed by (5) satisfies positive definiteness and symmetry but does *not* satisfy the triangle inequality. This is natural because the illumination parameters are unknown and they are determined dynamically.

III. Experiments

CUReT is the most challenging database for texture classification. It is a large database that contains 61 texture classes. Nine out of the 61 classes contain 410 images per class and the remaining 52 classes contain 205 images per class. As shown

[4]Substituting the expression of Δt in Ds to $d_{Ds,\Delta t}(x,\bar{x})$ yields $f(Ds)$.

TABLE I
RESULTS ON THE CUReT DATABASE. MEANS AND STANDARD DEVIATIONS ARE COMPUTED OVER 1000 RANDOM SPLITS OF THE TRAINING AND THE TEST SETS

92-images-per-class dataset				all images
T	RRFD	[24]	[35]	RRFD
6	84.60±0.83	81.67±0.96	–	64.64±1.36
12	92.23±0.46	89.74±0.66	91.60	78.23±1.64
23	96.82±0.32	94.69±0.45	94.59	87.82±0.71
46	99.30±0.18	97.50±0.30	97.58	94.20±0.38

Fig. 9. "Invisibility" of images is relative. (a) An "invisible" image caused by underexposure. (b) The image obtained by multiplying a scale $s = 2.0$ to the "invisible" image (a). (c) The image obtained by multiplying a scale $s = 4.0$ to (a).

in Fig. 1, there exists both large interclass confusion and intraclass variation. The images of a class are obtained under unknown viewpoint and illumination (see each row in Fig. 1), and some different classes look similar in appearance (see the first two rows in Fig. 1). Although lots of previous algorithms have been demonstrated on this database and have shown high classification accuracy, as mentioned in Introduction, they have just been tested on a 92-images-per-class subset which does *not* contain the unfavorably illuminated images [17], [19], [20], [23], [24], [35], making the rest relatively easy to be classified. In this section, we demonstrate our RRFD on the *whole* database. To compare this with previous approaches, we also test RRFD on the 92-images-per-class database, which is built according to [17]. We compare RRFD with the approaches proposed in [24], [35] because Hayman's approach [35] is the most competitive according to the survey in [20] and the approach in [24] is a recently developed one.

Classification Algorithm: Although RRFD does not provide any explicit feature vector in the R^n space, kernel-based classifiers can still be designed. We first choose a Gaussian kernel and compute a kernel matrix by

$$K(x, \tilde{x}) = \exp\left(-\frac{d_{iv}(x, \tilde{x})}{2\sigma^2}\right)$$

where σ is set to be 55 in our experiments. With this kernel, we can use a kernel-based LDA algorithm proposed in [45] to project the original data onto a linearly separable space and adopt a multiclass SVM [38] (one-against-one) with a linear kernel to train class models and classify test images.

Classification Results: To assess the classification performance, T training images are randomly chosen from each class while the remaining images are put into the test set. The configuration of T follows that in [17], [24], and [35]. Table I compares the performance of RRFD with the methods of [24] and [35]. On the 92-images-per-class dataset, RRFD outperforms both baselines. The advantage of RRFD comes from two aspects. On the one hand, RRFD adopts a per-pixel dense descriptor to represent an image, which makes RRFD more discriminative (for high precision). On the other hand, RRFD has appropriately handled the change caused by the geometric affine transformation and the illumination change, which makes RRFD capable of manipulating large intraclass variation (for high recall).

As expected, RRFD also achieves promising results when tested with the entire CUReT database. At $T = 46$, it achieves a classification rate of 94.20 ± 0.38. As mentioned by previous researchers [17], [19], [20], [23], [24], [35], it is hard to apply their approaches to the entire CUReT database because there are a portion of "invisible" images: "... *report results for 92 images per class. The remaining images do not have a sufficiently large portion of the texture visible to be cropped from the background.*" [24]. Since previous methods depend on filter banks and/or interest points detectors, which require textures being clearly imaged, it is hard for them to handle such "invisible" images as in Fig. 9(a). According to our statistics, about 30% of the images of CUReT are "invisible." It is, therefore, hard for them to achieve a 70% classification rate on the entire CUReT database. In comparison, the results in the last column of Table I show that RRFD still performs satisfactorily on the entire CUReT database. The main reason is that those "invisible" images are actually *visible* to RRFD. First, unlike previous approaches, which always try to bypass the "useless" information and target on the features of interest only, RRFD does not discard any information in the texture images. Second, as shown in Fig. 9, an "invisible" image can be made visible by choosing appropriate parameters s and t in the affine illumination model $I_{s,t} = sI + t$. However, RRFD does not try to adjust the illumination explicitly. Rather, it utilizes the illumination invariant distance metric to compute the distance between two textures. In this way, Fig. 9(a)–(c) actually does *not* produce any difference to RRFD. We have demonstrated the power of the illumination invariant distance metric in Fig. 8.

Impact of Parameters: RRFD only has two parameters. The first one is $\Delta\alpha$ required by the discrete Radon transform, which projects an image in $180^\circ/\Delta\alpha$ directions.[5] Another one is the bin size Δiv used for collecting the invariants in (2). There is no special consideration for Δiv and we set $\Delta iv = 0.1$ in all experiments. Fig. 10 shows the impact of $\Delta\alpha$ on the classification rates. We see that the classification accuracy decreases very slowly with the increase of $\Delta\alpha$. As RRFD works much faster with larger $\Delta\alpha$'s (i.e., smaller Radon-pixel image size), in practice one may set this parameter by balancing the effectiveness and the efficiency.

Examples of Incorrectly Classified Images: There are about 5.8% images in CUReT are not correctly classified. Part of the error is due to the high similarity among different classes (e.g., the two classes shown in the first two rows of Fig. 1). Another reason is that RRFD is not good at classifying images that have

[5]Another parameter Δs in Radon transform is always set to be 1 pixel and it need not be specified.

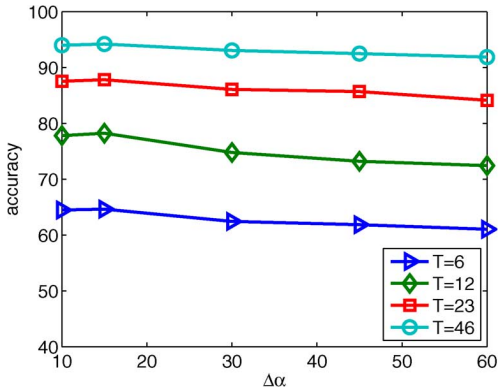


Fig. 10. Impact of $\Delta\alpha$. Another parameter is set to be $\Delta iv = 0.1$. The classification rates are the mean of 1000 random runs. These experiments are done on the entire CUReT database. T is the number of training images.

Fig. 11. Some examples of the images that are incorrectly classified. These images all have nontexture backgrounds.

nontexture backgrounds, as shown in Fig. 11. We have pointed out this at the end of Introduction.

Performance on Some Simple Databases: We test RRFD on VisTexture [43] and Brodatz [44] (Fig. 12) that have much less variations than CUReT. The part of VisTexture we use has 15 texture classes, each of which contains $4 \sim 20$ images. Brodatz has 13 classes with each class containing seven images. As these two databases are small, we do not perform random training/testing split, but instead apply RRFD on all images and observe its accuracy, defined as the percentage of the images that are correctly classified using their nearest neighbors. RRFD achieves 85.87% and 93.21% on VisTexture and Brodatz, respectively. These results illustrate that RRFD does not degenerate if the database has little or no variation caused by the geometry and/or the illumination transformations.

IV. Conclusion

In this paper, we propose an effective feature descriptor, RRFD, for texture classification. In contrast to previous filter-bank-based approaches, RRFD extracts Radon features that are invariant up to the geometric affine transformation. The illumination invariance is also achieved by computing the illumination invariant distance between affine invariant

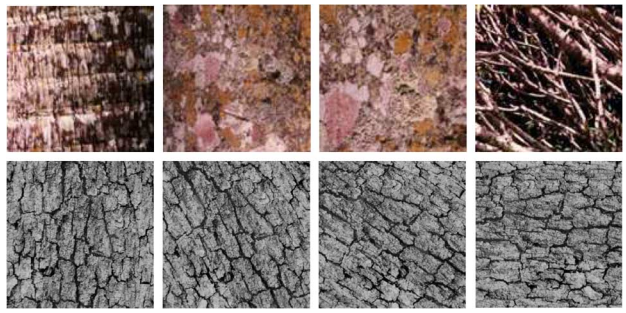


Fig. 12. Some examples of the images in VisTexture (the top row) and Brodatz (the bottom row). Each row is of the same class.

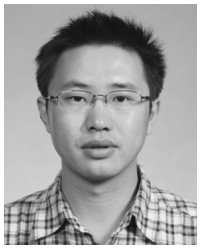
RRFD vectors. Another feature of RRFD is that it effectively makes use of the color information so as to produce more accurate texture descriptors. Note here that RRFD can also be applied to grayscale images. In this case, each dimension of the feature vector is a two-variate Gaussian distribution. Moreover, RRFD has only two parameters that do not need careful tuning. Finally, RRFD is computationally efficient. Our experiments demonstrated on a large scale database, CUReT, show that RRFD performed well and stably.

Actually, RRFD provides a general feature descriptor, so its application should not be limited to texture analysis only. The function of RRFD is to output a feature vector for a given image pattern, which can be an entire image or a local image window. We are seeking wider applications of RRFD in parallel to further improve the current approach.

References

[1] R. Haralick, "Statistical and structural approaches to texture," *Proc. IEEE*, vol. 67, no. 5, pp. 786–804, May 1979.
[2] A. K. Jain and F. Farrokhnia, "Unsupervised texture segmentation using gabor filters," *Pattern Recognit*, vol. 24, no. 12, pp. 1167–1186, 1991.
[3] T. Randen and J. H. Husoy, "Filtering for texture classification: A comparative study," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 4, pp. 291–310, Apr. 1999.
[4] O. Cula and K. Dana, "Compact representation of bidirectional texture functions," in *Proc. Computer Vision and Pattern Recognition Conf.*, 2001, vol. 1, pp. I-1041–I-1047, 1.
[5] O. G. Cula and K. J. Dana, "3d texture recognition using bidirectional feature histograms," *Int. J. Comput. Vis.*, vol. 59, no. 1, pp. 33–60, 2004.
[6] D. Harwood, T. Ojala, M. Pietikäinen, S. Kelman, and L. Davis, "Texture classification by center-symmetric auto-correlation, using kullback discrimination of distributions," *Pattern Recognit. Lett.*, vol. 16, no. 1, pp. 1–10, 1995.
[7] T. Ojala, M. Pietikäinen, and D. Harwood, "A Comparative Study of Texture Measures With Classification Based on Featured Distributions," *Patteron Recognit.* 1996.
[8] G. Zhao and M. Pietikainen, "Local binary pattern descriptors for dynamic texture recognition," in *Proc. 18th Int. Conf. Pattern Recognition*, Washington, DC, 2006, pp. 211–214.
[9] G. R. Cross, "Markov Random Field Texture Models," Ph.D. dissertation, East Lansing, MI, 1980.
[10] B. Caputo, E. Hayman, and P. Mallikarjuna, "Class-specific material categorisation," in *Proc. 10th IEEE Int. Conf. Computer Vision*, Washington, DC, 2005, pp. 1597–1604.
[11] J. Mao and A. K. Jain, "Texture classification and segmentation using multiresolution simultaneous autoregressive models," *Pattern Recognit.*, vol. 25, no. 2, pp. 173–188, 1992.
[12] M. Unser, "Texture classification and segmentation using wavelet frames," *IEEE Trans. Image Process.*, vol. 4, no. 11, pp. 1549–1560, Nov. 1995.
[13] D. Charalampidis and T. Kasparis, "Wavelet-based rotational invariant roughness features for texture classification and segmentation," *IEEE Trans. Image Process.*, vol. 11, no. 8, pp. 825–837, Aug. 2002.

[14] T. Leung and J. Malik, "Representing and recognizing the visual appearance of materials using three-dimensional textons," *Int. J. Comput. Vis.*, vol. 43, no. 1, pp. 29–44, 2001.
[15] C. Schmid, "Weakly supervised learning of visual models and its application to content-based retrieval," *Int. J. Comput. Vis.*, vol. 56, no. 1–2, pp. 7–16, 2004.
[16] F. Schaffalitzky and A. Zisserman, "Viewpoint invariant texture matching and wide baseline stereo," in *Proc. 8th IEEE Int. Conf. Computer Vision*, 2001, vol. 2, pp. 636–643, 2.
[17] M. Varma and A. Zisserman, "A statistical approach to texture classification from single images," *Int. J. Comput. Vis.*, vol. 62, no. 1–2, pp. 61–81, 2005.
[18] T. Tuytelaars and L. J. V. Gool, "Wide baseline stereo matching based on local, affinely invariant regions," presented at the BMVC, 2000.
[19] S. Lazebnik, S. M.-C. Schmid, and F.-J. Ponce, "A sparse texture representation using local affine regions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1265–1278, Aug. 2005.
[20] J. Zhang, M. Marsza, S. Lazebnik, and C. Schmid, "Local features and kernels for classification of texture and object categories: A comprehensive study," *Int. J. Comput. Vis.*, vol. 73, no. 2, pp. 213–238, 2007.
[21] J. Levy Vehel, P. Mignot, and J.-P. Berroir, "Multifractals, texture, and image analysis," presented at the IEEE Computer Society Conf. Computer Vision and Pattern Recognition, 1992.
[22] L. Kaplan, "Extended fractal analysis for texture classification and segmentation," *IEEE Trans. Image Process.*, vol. 8, no. 11, pp. 1572–1585, Nov. 1999.
[23] Y. Xu, H. Ji, and C. Fermuller, "A projective invariant for textures," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Washington, DC, 2006, pp. 1932–1939.
[24] M. Varma and R. Garg, "Locally invariant fractal features for statistical texture classification," presented at the IEEE Int. Conf. Computer Vision, Rio de Janeiro, Brazil, Oct. 2007.
[25] D. Lowe, "Object recognition from local scale-invariant features," in *Proc. 7th IEEE Int. Conf. Computer Vision*, 1999, vol. 2, pp. 1150–1157, 2.
[26] A. Khotanzad and Y. H. Hong, "Invariant image recognition by zernike moments," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-12, no. 5, pp. 489–497, May 1990.
[27] J. Yao and L. Chin, "Power-adjusted fractional power radial harmonic filters for shift- and scale-invariant pattern recognition with improved noise robustness and discrimination," *Opt. Commun.*, vol. 162, no. 5, pp. 26–30, 1999.
[28] D. Chetverikov and Z. Földvári, "Affine-invariant texture classification," in *Proc. Int. Conf. Pattern Recognition*, 2000, pp. 3901–3904.
[29] J. Zhang and T. Tan, "Affine invariant classification and retrieval of texture images," *Pattern Recognit.*, vol. 36, no. 3, pp. 657–664, 2003.
[30] C. Ballester and M. González, "Affine invariant texture segmentation and shape from texture by variational methods," *J. Math. Imag. Vis.*, vol. 9, no. 2, pp. 141–171, 1998.
[31] T. N. Tan, "Rotation invariant texture features and their use in automatic script identification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 7, pp. 751–756, Jul. 1998.
[32] J. Radon, The Radon Transform and Some of Its Applications ("translation of radon's 1917 paper"), 1983.
[33] M. R. Hejazi and Y.-S. Ho, "Texture analysis using modified discrete radon transform," *IEICE Trans. Inf. Syst.*, vol. E90-D, no. 2, pp. 517–525, 2007.
[34] P. Cui, J. Li, Q. Pan, and H. Zhang, "Rotation and scaling invariant texture classification based on radon transform and multiscale analysis," *Pattern Recognit. Lett.*, vol. 27, no. 5, pp. 408–413, 2006.
[35] E. Hayman, B. Caputo, M. Fritz, and J.-O. Eklundh, "On the significance of real-world conditions for material classification," *ECCV*, 2004.
[36] S. Agarwal and D. Roth, "Learning a sparse representation for object detection," *ECCV*, pp. 113–130, 2002.
[37] G. Balakrishnama, Linear Discriminant Analysis—A Brief Tutorial 1998.
[38] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Min. Knowl. Discov.*, vol. 2, no. 2, pp. 121–167, 1998.
[39] O. Al-Shaykh and J. Doherty, "Invariant image analysis based on radon transform and svd," *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 43, no. 2, pp. 123–133, 1996.
[40] H. Al-Yousefi and S. S. Udpa, "Recognition of arabic characters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 8, pp. 853–857, Aug. 1992.
[41] A. K. Jain, *Fundamentals of Digital Image Processing*. Upper Saddle River, NJ: Prentice-Hall, 1989.
[42] J. L. C. Sanz, E. B. Hinkle, and A. K. Jain, *Radon and Projection Transform-Based Computer Vision: Algorithms, a Pipeline Architecture, and Industrial Applications*. New York: Springer-Verlag, 1988.
[43] C. Rosalind Picard, C. Graczyk, and L. Campbell, Vision Texture [Online]. Available: http://vismod.media.mit.edu/~vismod/~imagery/~VisionTexture/~vistex.html
[44] [Online]. Available: http://sipi.usc.edu/database/database.cgi?volume=textures
[45] M. Sugiyama, "Local fisher discriminant analysis for supervised dimensionality reduction," in *Proc. 23rd Int. Conf. Machine Learning*, New York, 2006, pp. 905–912, ACM.



Guangcan Liu received the B.S. degree in applied mathematics from Shanghai Jiao Tong University, China, in 2004. He is pursuing the Ph.D. degree in the Department of Computer Science and Engineering, Shanghai Jiao Tong University.

He is a visiting student at Microsoft Research Asia. His research interests include machine learning and computer vision.

Zhouchen Lin (M'00–SM'08) received the Ph.D. degree in applied mathematics from Peking University, China, in 2000.

He is currently a researcher with the Visual Computing Group, Microsoft Research, Asia. His research interests include computer vision, computer graphics, pattern recognition, statistical learning, document processing, and human computer interaction.

Yong Yu received the M.S. degree from the Computer Science Department, East China Normal University.

He joined Shanghai Jiao Tong University (SJTU), China, in 1986. Now he is the Vice President of the Department of Computer Science and Engineering, Ph.D. candidate tutor, and the Chairman of E-generation technology research center (SJTU-IBM-HKU). He taught the courses "Computer Graphics and Human-machine Interface" and "Next Generation Web Infrastructure". As the head coach of the SJTU ACM-ICPC team, he and his team won the 2002 and 2005 ACM ICPC Championships. His research interests include semantic Web, Web mining, information retrieval, and computer vision.