Medical Image Retrieval via Histogram of Compressed Scattering Coefficients

Rushi Lan and Yicong Zhou, *Senior Member, IEEE*

***Abstract*—The features used in many current medical image retrieval systems are usually low-level hand-crafted features. This limitation may adversely affect the retrieval performance. To address this problem, this paper proposes a simple yet discriminative feature, called histogram of compressed scattering coefficients (HCSC), for medical image retrieval. In the proposed work, the scattering transform, a particular variation of deep convolutional networks, is first performed to yield more abstract representations of a medical image. A projection operation is then conducted to compress the obtained scattering coefficients for efficient processing. Finally, a bag-of-words (BoW) histogram is derived from the compressed scattering coefficients as the features of the medical image. The proposed HCSC takes the advantages of both scattering transform and BoW model. Experiments on three benchmark medical computer tomography image databases demonstrate that HCSC outperforms several state-of-the-art features.**

***Index Terms*—Computer tomography (CT) image, feature extraction, medical diagnosis, medical image retrieval (MIR), scattering transform.**

I. Introduction

MEDICAL imaging plays a significantly important role in the fields of computer-assisted diagnosis and medical analysis. Among different types of medical data, the images are regarded as a source for the diagnosis aid because they are able to directly capture the patient pathology [1]. Consequently, a huge number of medical images are collected every day in hospitals and medical institutions. To fully apply these medical images, many techniques have been developed to handle the obtained images from different aspects, such as data representation, storage, segmentation, and reconstruction.

Recently, medical image retrieval (MIR) has aroused considerable research interest. For an input query medical image, the MIR system will return several images whose contents are most relevant to the query one. The retrieval results may not only help to manage the massive medical images, but also provide useful information to support the doctors to make decisions. Many MIR systems have been proposed in the literature, and more insights can be found in [2] and [3].

The core component of an MIR system is to extract features from the medical images as many classical content-based image retrieval approaches. The derived features are required to reflect the essential and discriminative characteristics of the medical images. Numerous research efforts have been devoted to extracting discriminative features of the medical images from different perspectives so far.

A. Related Work

Shape, a basic attribute of the image, is commonly used to derive features for MIR. The authors in [4] evaluated the performance of several representative shape characteristics using the X-ray image retrieval. In their experiments, the following features were considered: global shape properties (such as center of gravity, area, perimeter, and major axis length and angle), invariant moments, scale space filtering-based features, polygon approximation-based features, and Fourier descriptors. Similarly, Xu *et al.* proposed shape representation methods using the length, absolute orientation, and relative orientation of the line segments as well as the angles associated with multiple open triangles to design a spine X-ray image retrieval system [5]. Though these shape representation features achieve satisfactory results for some medical retrieval applications, they are heavily dependent upon a boundary detection or the region-of-interest segmentation operations. These preprocessing steps may limit the feasibilities of shape-based features.

The color information of medical images has also been applied to feature extraction. The authors in [6] developed an MIR system using a quantization of the Hue, Saturation, Value (HSV) space as features. Akakin and Guran took more color spaces into account in their system [7], namely the mean and standard values of each channel of red–green–blue, CIE*Lab*, and HSV spaces. Additionally, eight statistics features, i.e., mean, standard deviation, skewness, kurtosis, maximum and minimum values, energy, and entropy, were also used. Promising results have been achieved by these features.

Compared with shape and color information, the texture characteristics of medical images are more important because they contain the core in identifying tissues [8]. The features used in mainstream MIR systems are mainly extracted using the texture information. The strategies of traditional texture representation approaches have been analogously applied in many MIR systems.

Manuscript received June 16, 2016; revised September 21, 2016; accepted October 28, 2016. Date of publication November 1, 2016; date of current version September 1, 2017. This work was supported in part by the Macau Science and Technology Development Fund under Grant FDCT/016/2015/A1 and by the Research Committee at University of Macau under Grant MYRG2014-00003-FST and Grant MYRG2016-00123-FST.

The authors are with the Department of Computer and Information Science, University of Macau, Macau 999078, China (e-mail: lrs0106@yahoo.com; yicongzhou@umac.mo).

Digital Object Identifier 10.1109/JBHI.2016.2623840

The well-known Gabor and wavelet transforms have been proven to be effective tools to extract features from natural texture images. Based on these findings, Traina *et al.* proposed a MultiWaveMed system based on both Daubechies and Gabor wavelets [9]. The mean and standard deviation values of the magnitudes of different transformed coefficients were utilized as features. Similarly, the authors of [6] developed the texture features using Gabor responses in different directions and scales, and Quddus and Basir extracted texture features of magnetic resonance (MR) images based on the gray-scale values of magnitudes of multiscale wavelet coefficients [10]. Apart from Gabor and wavelet transformed domains, Akakin and Gurcan extracted characteristics from the cooccurrence histograms [7]. Their features consist of different statistics (such as mean, standard deviation, and correlation) of the normalized cooccurrence histograms. Besides, Larsen *et al.* proposed shape index histograms that capture the second-order image structure at multiple scales as features of indirect immunoflourescence images of HEp-2 cells [11]. The texture features are usually integrated with the shape and color information to achieve a comprehensive representation of the medical image.

Owning to the great success of local binary pattern (LBP) [12] in texture image representation, many local patterns have also been proposed for medical images. LBP has been directly performed to the brain MR images [13] and lung computer tomography (CT) images [14] for feature extraction. A directional binary wavelet pattern was proposed by integrating LBP with bitplane decomposition and binary wavelet transform [15]. Similarly, Li and Meng applied LBP to the subimages of wavelet transform results of wireless capsule endoscopy images [16], and the authors in [17] extended the traditional local derivative pattern (LDP) to the cooccurrence of similar ternary edges to derive local features for biomedical images. It can be found that these methods mainly are either to perform LBP in a specific transformed domain or to improve the existing encoding ways. More work of local patterns for medical retrieval can be found in [18]–[21].

B. Our Main Contributions

Although having achieved success in certain tasks, the aforementioned features are usually low-level hand-crafted features, and they have difficulties in discovering more abstract representations of medical images in higher levels. To address this limitation, several researchers analyze the medical images using deep learning techniques. The notable convolutional neural network (CNN) has been successfully applied to derive features for lung pattern classification [22], bodypart recognition [23], and cell nuclei detection and classification [24]. Recently, the scattering transform [25], [26], a variation of deep convolutional networks, has been proposed to represent the image contents by cascading wavelet transform and modulus pooling operations. Compared with CNN, the properties and optimal configurations of scattering transform have been well understood. It offers abstract representations of images and shows impressive performance to classify the texture images [27], [28]. It is interesting to extract high-level features of medical images using the scattering transform.

In this paper, we propose a novel feature, named histogram of compressed scattering coefficients (HCSC), for MIR. Given a medical image, the scattering coefficients, obtained by the scattering transform, are stable to deformations and preserve high-frequency information. To efficiently handle these coefficients, a compression operation is carried out to reduce the dimension of the scattering coefficients. Finally, a bag-of-words (BoW) HCSC is used as the texture feature of the medical image. HCSC features are easy to implement yet discriminative because it takes advantages of both the scattering transform and the BoW model. Experiments are carried out to evaluate the proposed HCSC features, and state-of-the-art performance is obtained for medical CT images retrieval.

Compared with previous works, our main contributions are threefold.

1) To the best of our knowledge, this is the first time to derive texture characteristics of medical CT images using scattering transform.
2) We propose a novel feature, named HCSC, using the scattering representations of CT images.
3) We conduct experiments on three benchmark CT image databases to evaluate the performance of the proposed features, and state-of-the-art results have been achieved.

The rest of this paper is organized as follows: Section II details the proposed HCSC features, and Section III evaluates the proposed features from different aspects. Section IV finally draws the conclusions.

II. Methodology

In this section, we present the proposed HCSC features in detail. An overview of HCSC will be first given, and then, each step of the HCSC extraction will be introduced successively.

A. Overview of HCSC

The derivation of the proposed HCSC consists of the following modules. With a given training image set, we first perform the scattering transform to each image to obtain the translation invariant scattering representations of all images. These scattering coefficients cannot be directly applied for feature extraction because they are of high dimension. Then, a compression operation is carried out to the obtained coefficients for a dimension reduction purpose. Based on the compressed scattering coefficients (CSC), the BoW model is finally used to derive the HCSC.

B. Compressed Scattering Coefficients

A wavelet transform is obtained by a convolution of the signal and a specific wavelet. Let φ and $\boldsymbol{R}$ be a signal bandpass filter and a group of rotations r. $R_r x$ means the rotation of $x \in \mathbb{R}^2$ by an angle r. The directional wavelet is achieved by rotating φ along angle r and scaling it by 2^j as follows:

$$\varphi_{j,r}(x) = 2^{-2j}\varphi(2^{-j}R_r x). \tag{1}$$

The directional wavelet transform of a signal f at position x for scale 2^j is $f \star \varphi_{j,r}(x)$, where $\star$ is the convolution operation.

To improve the computational efficiency of the well-known scale-invariant feature transform (SIFT) descriptor, the authors in [29] proposed a DAISY descriptor that can be represented as follows:

$$|f \star \varphi_{j,r}| \star \phi_J(x) \tag{2}$$

where $\phi_J(x) = 2^{-2J}\phi(2^{-J}x)$ is a low-pass filter and $j \leqslant J$. Equation (2) indicates that the local features of f can be extracted by averaging the wavelet coefficient amplitude $|f \star \varphi_{j,r}|$.

The researchers in [27] found that the low-pass convolution operation in (2) may eliminate the high-frequency characteristics of f. To restore part of the information lost, the wavelet coefficient amplitude $|f \star \varphi_{j_1,r_1}|$ will be convoluted with another wavelet, which is $|f \star \varphi_{j_1,r_1}| \star \varphi_{j_2,r_2}$. Then, the scatter coefficients of f are represented as [27]

$$||f \star \varphi_{j_1,r_1}| \star \varphi_{j_2,r_2}| \star \phi_J(x). \tag{3}$$

Note that j_2 should be smaller than j_1 because it has been proven that $|f \star \varphi_{j_1,r_1}| \star \varphi_{j_2,r_2}$ is negligible if $j_2 \geqslant j_1$.

Iterating this operation p times, a vector of scattering coefficients at x is obtained as

$$S_{p,J}f(x) = |||f \star \varphi_{j_1,r_1}| \star \cdots \star |\varphi_{j_p,r_p}| \star \phi_J(x) \tag{4}$$

where $j_1 < \cdots < j_p < J$ and $\{r_1, \ldots, r_p\} \in \boldsymbol{R}$. The outputs of the first iteration ($p = 1$) are SIFT-type descriptors, and the results of the second iteration ($p = 2$) are the complementary invariant information [28]. Compared with traditional wavelet transforms, the modulus and averaging pooling operations in (4) ensure that the scattering transform is highly nonlinear and translation invariant from coefficients. More details about scattering transform can be found in [26]–[28].

For a given image I, we can obtain a series of coefficients after performing scattering transform. These coefficients contain different characteristics of the original image. Let $sc_l(m,n)$ denote the lth scattering coefficient matrix of I, where $l = 1, 2, \ldots, L$, $m = 1, 2, \ldots, M$, and $n = 1, 2, \ldots, N$. L is the total number of the matrix, and $\{M, N\}$ is the size of the matrix. Take the first CT image shown in Fig. 3 as an example. Its size is 512 × 512, and the $\{L, M, N\}$ for its first and second layer output after scattering transform are {32, 64, 64} and {384, 64, 64}, respectively. Fig. 1 visually illustrates several examples of these scattering coefficient matrices. It can be seen that, unlike the commonly used texture features, the scattering transform decomposes the original CT image into different semantic components, such as the edge and shape information. These coefficient matrices are high-level representations because they not only provide a comprehensive description, but also contain more structure characteristics of the image, which will benefit the feature extraction procedure.

Though $sc_l(m,n)$ carries plenty of image characteristics, it is difficult to directly handle these coefficients because of the high dimension. As mentioned before, for a 512 × 512 image, all $sc_l(m,n)$ of the second layer output contain 384 × 64 × 64 real numbers, which is six times of the image number itself. It will take a large amount of storage and computation cost to handle these data, and the time complexity is also unacceptable. The well-known principle component analysis is a commonly used dimension reduction algorithm, and it has been successfully used to many applications. However, it does not work well in this situation because the derivation of the eigenvector of the covariance matrix is also with high computation cost.

To address this problem, a compression process is conducted to the coefficients before feature extraction. Projection operations are commonly used compression methods because they may not only reduce the dimension, but also preserve some properties of original data. Due to these advantages, projection operations have been successfully applied in many feature extraction algorithms [30]–[34]. For the efficiency purpose, the following simple projection algorithms are considered.

1) Vertical Projection (VP) and Horizontal Projection (HP): These two projections compress the data by accumulating $sc_l(m,n)$ along the vertical or horizontal directions as follows:

$$sc_l^v(n) = \sum_{\mathrm{m}=1}^{M} sc_l(m,n) \tag{5}$$

$$sc_l^h(m) = \sum_{\mathrm{n}=1}^{N} sc_l(m,n). \tag{6}$$

After this projection operation, the size of $sc_l(m,n)$ reduces to 1 × N or M × 1.

2) Ring Projection (RP) [30] and Central Projection (CP) [31]: To perform these two projections to $sc_l(m,n)$, the Cartesian coordinate system is first transformed to polar coordinate system. Denote the polar coordinate version of $sc_l(m,n)$ by $\bar{sc}_l(\rho,\theta)$, where ρ and θ respond to the polar radius and angle, respectively. The RP and CP can be achieved as

$$sc_l^r(\rho) = \int \bar{sc}_l(\rho,\theta)d\theta \tag{7}$$

$$sc_l^c(\theta) = \int \bar{sc}_l(\rho,\theta)d\rho. \tag{8}$$

$sc_l^r(\rho)$ and $sc_l^c(\theta)$ perform integral along the polar radius and angle directions.

We call these projection results as CSC. Compared with $sc_l(m,n)$, CSC inevitably remove some image characteristics during the projection operation, but they have following advantages.

1) They require less storage space. The data number of the compression results is smaller than those of original image and scattering coefficients. Take the CT image with 512 × 512 in size as an example. The total data number of $sc_l^v(n)$ of the second layer output is 9.38% and 1.56% compared with the numbers of the original image and the corresponding scattering coefficients.
2) Extracting features from the compressed result is more effective. Because the data are reduced to much smaller dimension compared with the original size, it will take less computation complexity to handle the compressed data.
3) The compressed data reveal the distribution of image characteristics. Assuming $s_l(m,n)$ to be a joint probability function, $sc_l^v(n)$ and $sc_l^h(m)$ can be regarded as margin distributions of $sc_l(m,n)$ and are suitable for further processing. On the other hand, $sc_l^r(\rho)$ and $sc_l^c(\theta)$ are

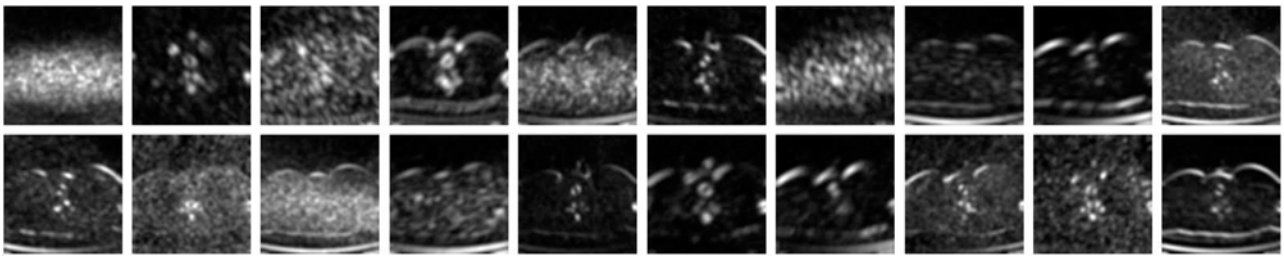


Fig. 1. Visual illustration of some selected scattering transform results of the first CT image in Fig. 3.

Fig. 2. Flowchart of codebook generation for HCSC.

the margin distributions of $\bar{sc}_l(\rho, \theta)$ on the polar coordinate system.

Due to the above superiorities, we derive features using CSC.

C. HCSC Extraction

There are many ways to derive features based on the CSC. We elected to use the well-known BoW model because of its good performance in several domains, as reported in the technical literature [35]–[37].

The key step of BoW is to develop a codebook from the obtained CSC. Let $\mathbb{S} = \{I_c^k | k = 1, \ldots, K, c = 1, \ldots, C\}$ be a set of training images, where C is the class number in the database, and K is the image number for each class for training. To each I_c^k, we extract the CSC as described in Section II-B. Here, we use the VP of the scattering coefficients in (5) as an example. Let $\boldsymbol{sc} = \{sc_h^v | h = 1, \ldots, H\}$ be all the vertical CSC extracted from $\mathbb{S}$, where H is the total number of the derived CSC. The codebook can be generated by supervised or unsupervised ways in the literate. For computational efficiency, we develop the codebook by clustering over $\boldsymbol{sc}$. The flowchart of the codebook generation is summarized in Fig. 2. The typical algorithm, K-means, is applied here because of its simpleness and good performance. Each cluster, derived by performing K-means to $\boldsymbol{sc}$, is regarded as a codeword, and all codewords form the codebook. It requires sufficient samples to generate an accuracy codebook. Hence, we make use of the second layer output of the scattering transform results, which contains more coefficients than those of the first layer output, to develop the codebook. Denote the obtained codebook by $\boldsymbol{B} = \{\boldsymbol{b}_g | g = 1, \ldots, G\}$, where $\boldsymbol{b}_g$ is the gth cluster of $\boldsymbol{sc}$, and G is the number of clusters. The detailed procedures to derive $\boldsymbol{B}$ are summarized as Algorithm 1.

After obtaining the codebook $\boldsymbol{B}$, a histogram is used to describe the distribution of the CSC of each image. The hard voting is utilized to generate the histogram, which assigns each CSC of the image to its closest codeword. Suppose there are a total of Z CSC for a CT image. The coding representation of the zth CSC, denoted by csc_z, can be achieved as

$$v_g^z = \begin{cases} 1, & \text{if } g = \arg\min_j \|csc_z - \boldsymbol{b}_j\|, \\ 0, & \text{otherwise.} \end{cases} \quad g = 1, \ldots, G. \tag{9}$$

Algorithm 1: Codebook generation of CSC.

Input: A set of training images, and the codeword number G.

Output: The codebook $\boldsymbol{B} = \{\boldsymbol{b}_g | g = 1, \ldots, G\}$.

(a) Perform scattering transform to all the image in the training set to obtain the scattering coefficients;

(b) Project each scattering coefficient matrix into a vector along the vertical or horizontal direction using (5) or (6) to obtain all the CSC.

(c) Find G clusters $\{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_G\}$ by conducting the K-means algorithm to all the CSC as codewords.

Then, the occurrence frequency of all CSC on the codeword $\boldsymbol{B}$ will be calculated as

$$h(g) = \sum_{z=1}^{Z} v_g^z, \quad g = 1, \ldots, G. \tag{10}$$

A normalization operation is finally performed as

$$\bar{h}(g) = \frac{h(g)}{\sum_{g=1}^{G} h(g)}, \quad g = 1, \ldots, G. \tag{11}$$

$\bar{h}$ is regarded as the HCSC of the original image.

From the above descriptions, we can conclude that the proposed HCSC features of medical images are flexible to implement. However, HCSC features are quite discriminative characteristics of medical images because 1) the scattering transform provides extensive texture characteristics with different scales and directions, and 2) the BoW model is able to reveal the essential distribution of the obtained texture characteristics. The projection operations perfectly bridge the gap between the scattering transform and the BoW model.

III. Experiments

This section provides detailed experimental results to evaluate the performance of the proposed HCSC features from different aspects. The experimental settings are first presented, including the databases, similarity measurement of two feature vectors, evaluation criteria of retrieval performance, and comparative retrieval methods. After that, experiments are carried out to test the effects of different parameters of HCSC. Subsequently, the proposed HCSC features are compared with several

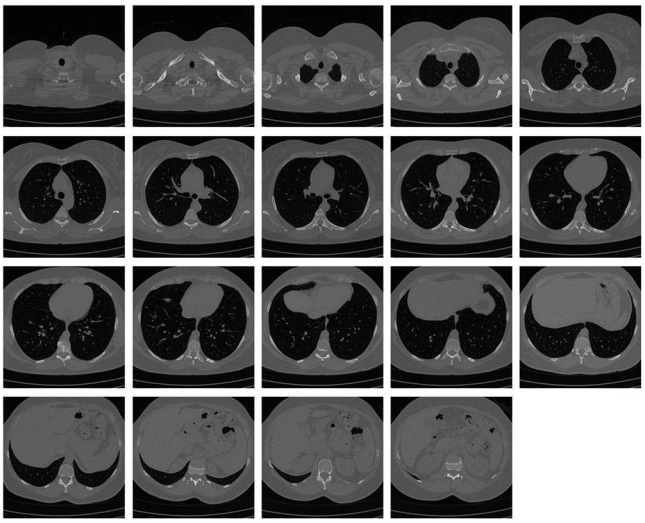


Fig. 3. Medical CT image examples of from EXACT09-CT database. Each image is randomly selected from the corresponding category.

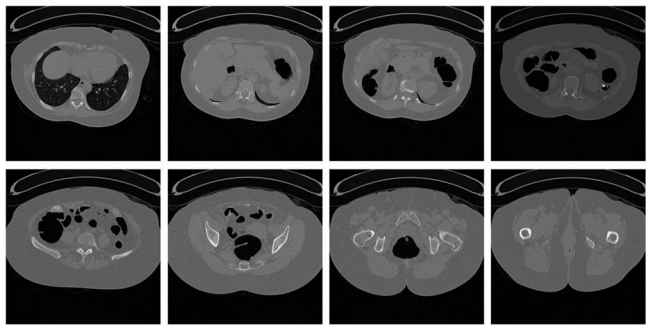


Fig. 4. Medical CT image examples of from TCIA-CT database. Each image is randomly selected from the corresponding category.

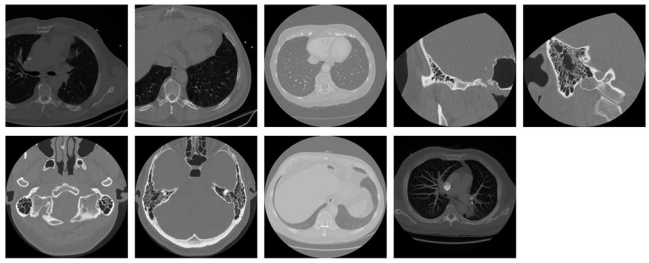


Fig. 5. Medical CT image examples of from NEMA-CT database. Each image is randomly selected from the corresponding category.

state-of-the-art features using three CT image databases. Finally, a discussion on these results is given.

A. Experimental Settings

1) Databases: Three publicly available CT image databases, namely EXACT09-CT [20], [38], TCIA-CT [20], [39], and NEMA-CT [40], are used in our experiments. The detailed information about these databases is as follows:

1) Extraction of Airways from CT 2009 (EXACT09) consists of CT scans at the chest [38]. Each image is stored using the Digital Imaging and Communications in Medicine (DICOM) format and is a size of 512 × 512. Following the setting in [20], we select a subset of EXACT09, the CASE23 of testing set, to form the EXACT09-CT database to evaluate our proposed features. There are totally 675 CT images in this database, and these images are grouped into 19 categories, as in [20]. The image numbers for each category are 36, 23, 30, 30, 50, 42, 20, 45, 50, 24, 28, 24, 35, 40, 50, 35, 30, 28, and 55, respectively. Fig. 3 illustrates the example images of this database with one image of each category.
2) There are a large number of cancer images in the cancer image archive (TCIA) [39]. The TCIA-CT database, collected by the authors in [20], is used here, which is generated by using the 604 CT images of the DICOM series number 1.3.6.1.4.1.9328.50.4.2 of study instance UID 1.3.6.1.4.1.9328.50.4.1 for subject 1.3.6.1.4.1.9328.50.4.0001. These images are grouped into eight categories, and the image numbers of each category are 75, 50, 58, 140, 70, 92, 78, and 41. Fig. 4 depicts the example images of this database with one image of each category.
3) The National Electrical Manufacturers Association (NEMA) [40] database provides CT images of different parts of the human body that can be applied for research and diagnosis purposes. Following the experimental setting in [20], we create the NEMA-CT database by selecting 315 CT images with dimension 512 × 512 from the CT0001, CT0003, CT0057, CT0060, CT0080, CT0082, and CT0083 cases. These images are further categorized into nine groups that have 36, 18, 36, 37, 41, 30, 23, 70, and 34 images, respectively, representing different parts of the body. The images of this database with one image of each category are shown in Fig. 5 as examples.

2) Similarity Measurement: Let $\bar{h}_1$ and $\bar{h}_2$ be two HCSC feature vectors extracted from two CT images. Many histogram distances can be used to measure the similarity between them. The well-known histogram intersection is applied in this paper, which can be achieved by

$$HI(\bar{h}_1, \bar{h}_2) = \sum_{g=1}^{G} \min(\bar{h}_1(g), \bar{h}_2(g)). \tag{12}$$

The larger $HI(\bar{h}_1, \bar{h}_2)$ indicates more similarity between $\bar{h}_1$ and $\bar{h}_2$.

3) Evaluation Criteria: In the following experiments, each CT image in the database is chosen as a query image and matched with the rest images. The images with top S largest histogram intersection values are treated as the retrieval results. We follow the setting of [20], namely the images with ten largest similarity measurements are returned as a set of retrieval results. The precision and recall for the query image are commonly used to evaluate the retrieval results, which can be achieved as

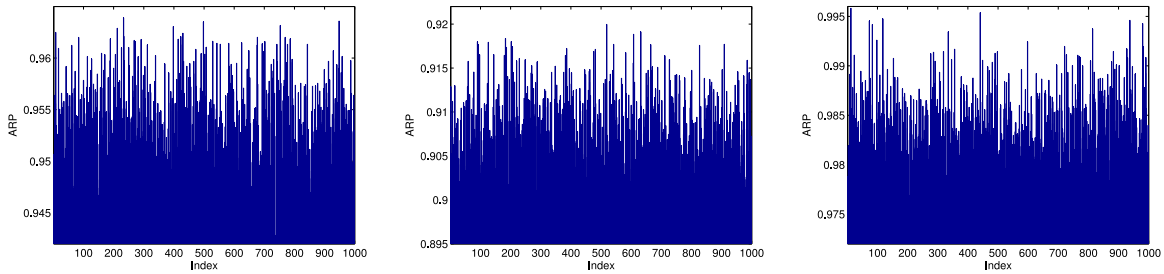


Fig. 6. Retrieval performance (ARP) of HCSC features obtained by different codebooks for 1000 repeats via (a) TCIA-CT, (b) EXACT09-CT, and (c) NAMA-CT databases, respectively.

follows:

$$\mathfrak{P} = \frac{\text{Number of relevant images retrieved}}{\text{Number of images retrieved}} \tag{13}$$

$$\mathfrak{R} = \frac{\text{Number of relevant images retrieved}}{\text{Number of relevant images in the database}}. \tag{14}$$

Note that a retrieved image is treated as relevant if it is in the same category as the query image.

In this work, the average retrieval precision (ARP) and average retrieval rate (ARR), the means of average precision and average recall per category, are used as evaluation criteria as in [20]. Besides, F_{score} is also considered, which can be obtained by [19]

$$F_{\text{score}} = \frac{2 \times \text{ARP} \times \text{ARR}}{\text{ARP} + \text{ARR}}. \tag{15}$$

4) Comparative Retrieval Methods: For each database, we compare our method for MIR with retrieval performance using features commonly used in the technical literature. These features include LBP [41], local ternary pattern (LTP) [42], LDP [43], local tetra pattern (LTrP) [44], local ternary cooccurrence pattern (LTCoP) [17], local mesh pattern (LMeP) [45], spherical symmetric 3-D local ternary pattern (SS-3D-LTP) [18], and local wavelet pattern (LWP) [20], respectively.

B. Effects of Parameters

This section investigates the effects of different parameters in the derivation of HCSC from following perspectives.

Because selection of different training samples from the database may yield different codebooks, we first study the robustness of HCSC obtained by different codebooks. All three databases are used here. More specifically, two training samples per category are randomly chosen from the database to generate the codebook, and the number of codewords in the codebook is set to 200. The VP is employed here. This procedure is repeated 1000 times to obtain 1000 different codebooks, and the HCSC features are extracted using these codebooks. Then, the image retrieval test is carried out as described in Section III-A3. The retrieval performance in terms of ARP using all three databases is plotted in Fig. 6. A statistics analysis of these results, including the minimum, maximum, mean, and variance, is also presented in Table I. The differences between the maximum and minimum values of the TCIA-CT, EXACT09-CT, and NEMA-CT databases are 0.0312, 0.0326, and 0.0316, respectively. These values are small compared with the corresponding mean values of each criterion. On the other hand, the variance values given in Table I also indicate that all criteria of 1000 repeats are quite stable. Therefore, we can conclude that HCSC features are robust to different codebooks.

TABLE I
STATISTICS ANALYSIS OF THE RETRIEVAL PERFORMANCE (ARP) OF HCSC FEATURES OBTAINED FROM DIFFERENT TRAINING SAMPLES

	Min	Max	Mean	Variance
TCIA-CT	0.9327	0.9639	0.9511	3.0973e-5
EXACT09-CT	0.8874	0.9200	0.9056	2.9694e-5
NEMA-CT	0.9642	0.9958	0.9817	2.1596e-5

TABLE II
RETRIEVAL PERFORMANCE (ARP) OF HCSC FEATURES OBTAINED BY DIFFERENT PROJECTION ALGORITHMS

	TCIA-CT	EXACT09-CT	NEMA-CT	Aver.
HCSC_v	**0.9510**	0.9057	**0.9818**	0.9462
HCSC_h	0.9461	0.8963	0.9813	0.9412
HCSC_c	0.9473	**0.9186**	0.9752	**0.9470**
HCSC_r	0.9273	0.9095	0.9503	0.9290

Next, the effects of different projection directions are investigated. As aforementioned in Section II-B, four directions are considered here, namely VP, HP, RP, and CP. To keep brevity, denote the obtained features by HCSC_v, HCSC_h, HCSC_r, and HCSC_c. For each database, we randomly select two samples per category to form the training set and derive the HCSC features by different projection algorithms. This procedure is run 500 times. The average ARP values of all databases are reported here. As seen from Table II, HCSC_v obtains the best performance for the TCIA-CT and NEMA-CT databases, while HCSC_c wins other features for the EXACT09-CT database. Considering the average performance, HCSC_c achieves the most satisfactory results, slightly surpassing HCSC_v. These results indicate that CP and VP are more suitable to explore the information distributions of CT images. Owing to VP is easier to implement than CP, it is used to derive HCSC in the following experiments.

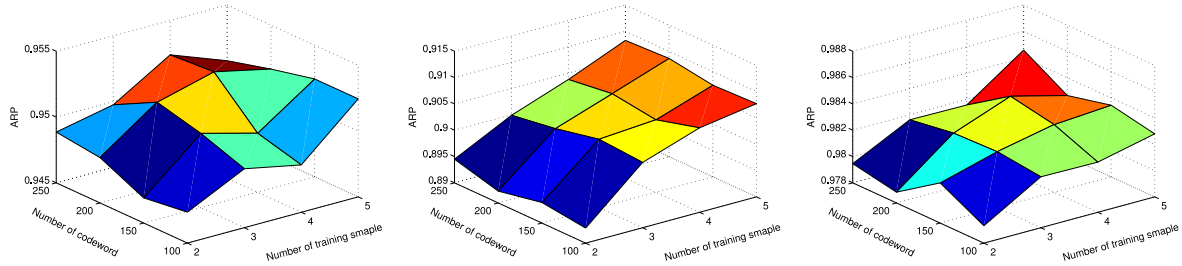


Fig. 7. Retrieval performance (ARP) of HCSC features obtained by different numbers of training sample and codeword via (a) TCIA-CT, (b) EXACT09-CT, and (c) NAMA-CT databases, respectively.

TABLE III
COMPARISON RESULTS OF DIFFERENT METHODS ON THE EXACT09-CT DATABASE

Methods	Evaluation Criteria		
	ARP	ARR	F_{score}
LBP	0.6503	0.1951	0.3002
LTP	0.6209	0.1854	0.2855
LDP	0.5440	0.1619	0.2495
LTrP	0.5782	0.1729	0.2662
LTCoP	0.7348	0.2216	0.3405
LMeP	0.6323	0.1891	0.2911
SS-3D-LTP	0.6700	0.2009	0.3091
LWP	0.8300	0.2487	0.3827
HCSC	**0.9150**	**0.2883**	**0.4384**

Best results are in bold.

TABLE IV
COMPARISON RESULTS OF DIFFERENT METHODS ON THE TCIA-CT DATABASE

Methods	Evaluation Criteria		
	ARP	ARR	F_{score}
LBP	0.6691	0.0974	0.1700
LTP	0.7183	0.1033	0.1806
LDP	0.6906	0.1005	0.1755
LTrP	0.7469	0.1095	0.1910
LTCoP	0.7440	0.1092	0.1904
LMeP	0.7371	0.1077	0.1879
SS-3D-LTP	0.8054	0.1171	0.2045
LWP	0.8840	0.1309	0.2280
HCSC	**0.9512**	**0.1452**	**0.2520**

Best results are in bold.

We finally test the performance of HCSC features obtained by different numbers of training samples and the codewords. The numbers of training samples are set to 2, 3, 4, and 5, respectively. The codebook sizes, namely the number of codewords, are set to 100, 150, 200, and 250, respectively. For each setting of training samples and codebook sizes, the retrieval experiment is run 25 times to obtain the average ARP value for evaluation. The retrieval results of all three databases with different parameter settings are illustrated in Fig. 7. It can be seen that the distributions of ARP values are different for all databases. In these cases, the differences between the maximum and minimum of the ARP measurement for TCIA-CT, EXACT09-CT, and NEMA-CT databases are 0.0057, 0.0164, and 0.0058, respectively. Generally speaking, the retrieval results of HCSC slightly improve as the increments of training sample and codeword numbers. However, it will take more computation cost to obtain the codebook with large training and codeword numbers. It is necessary to trade off the computation complexity and retrieval accuracy. In the following experiments, we empirically set the training sample and codeword numbers to 3 and 200 to obtain a balance.

C. Experiments on the EXACT09-CT Database

This experiment evaluates the performance of HCSC using the EXACT09-CT database. The performance of all methods in terms of ARP, ARR, and F_{score} are illustrated in Table III. For this CT image database, LDP performs worst, while the proposed HCSC features achieve the most satisfactory performance. The retrieval results of proposed method are improved by $\{68.20\%, 78.07\%, 75.71\%\}$ and $\{10.24\%, 15.92\%, 14.55\%\}$ as compared with LDP and the second best method LWP in terms of $\{\text{ARP, ARR, } F_{\text{score}}\}$.

D. Experiments on the TCIA-CT Database

The TCIA-CT database is used in this experiment. The comparative methods used in Section III-C are considered here too, and they are evaluated by the identical experimental settings used for the EXACT09-CT database. The retrieval performance of all algorithms is presented in Table IV. We can observe that the extended local patterns obtain better results than those of the original LBP method for this database, and the performance of LTCoP and LMeP is comparable. Among all local patterns, LWP gets obvious improvements in contrast with other approaches. The proposed HCSC improves LWP by a large margin, i.e., $\{7.60\%, 10.92\%, 10.53\%\}$, using the corresponding criteria.

E. Experiments on the NEMA-CT Database

We further investigate the retrieval performance of the proposed method using the NEMA-CT database. Table V shows the ARP, ARR, and F_{score} of all methods. We can observe that these methods yield close performance for this database. LWP and LDP gain more satisfactory results among all

TABLE V
COMPARISON RESULTS OF DIFFERENT METHODS ON THE NEMA-CT DATABASE

Methods	Evaluation Criteria		
	ARP	ARR	F_{score}
LBP	0.9055	0.2933	0.4431
LTP	0.9200	0.3023	0.4551
LDP	0.9422	0.3108	0.4674
LTrP	0.9369	0.3096	0.4654
LTCoP	0.9215	0.3031	0.4562
LMeP	0.9309	0.3062	0.4608
SS-3D-LTP	0.9224	0.3026	0.4557
LWP	0.9532	0.3133	0.4716
HCSC	**0.9833**	**0.3364**	**0.5013**

Best results are in bold.

comparative methods, outperforming LBP by about 0.05 via ARP measurement. The proposed HCSC features surpass all the other features, and it gets improvements of $\{3.16\%, 7.37\%, 6.30\%\}$ by comparing with the results of the second best feature LWP in terms of ARP, ARR, and F_{score}.

F. Discussion

As a deep convolutional network whose filters are wavelet operators, the scattering transform explores the informative and discriminative characteristics of the CT images. The projection and BoW model, involved in HCSC, further provide a compact representation of numerous scattering coefficients. These merits ensure the representation capacity of HCSC. The experimental results also demonstrated that HCSC achieved satisfactory performance in medical CT image retrieval compared with several state-of-the-art features.

On the other hand, the implementation of HCSC may consume more time in contrast with several hand-crafted local features. For HCSC extraction, the most time-consuming step lies in the scattering transform, which takes more computational cost because it is a particular type of deep convolutional networks. Nevertheless, for medical retrieval, especially for diagnostic purposes, the retrieval accuracy is relatively more important than the implementation efficiency. Therefore, the proposed HCSC features are promising features for CT image retrieval.

IV. CONCLUSION

In this paper, we developed the HCSC as a novel feature for MIR. First of all, we performed a scattering transform to a medical image to achieve its translation invariant representations. After the compression operation, a histogram was derived as the feature vector using the BoW framework. In order to evaluate the HSCS features, we conducted MIR experiments using three CT image databases, namely EXACT09-CT, TCIA-CT, and NEMA-CT, respectively. Experimental results demonstrated that the proposed features obtained the state-of-the-art performance in comparison with several existing features, such as LBP, LMeP, LWP, and SS-3D-LTP features. There are several interesting directions deserving further studies, such as developing more discriminative projection methods to compress the scattering coefficients, applying more powerful frameworks to CSC for feature extraction, and improving the scattering transform for image representation.

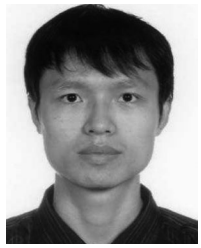
ACKNOWLEDGMENT

The authors would like to thank the editor and anonymous reviewers for their insightful comments and constructive suggestions, and to sincerely thank the authors of [20] for providing the NEMA-CT database.

REFERENCES

[1] D. L. Rubin, H. Greenspan, and J. F. Brinkley, "Biomedical imaging informatics," in *Biomedical Informatics*. New York, NY, USA: Springer, 2014, pp. 285–327.
[2] H. Müller, N. Michoux, D. Bandon, and A. Geissbuhler, "A review of content-based image retrieval systems in medical applicationsclinical benefits and future directions," *Int. J. Med. Informat.*, vol. 73, no. 1, pp. 1–23, 2004.
[3] X. Zhang, W. Liu, M. Dundar, S. Badve, and S. Zhang, "Towards large-scale histopathological image analysis: Hashing-based image retrieval," *IEEE Trans. Med. Imag.*, vol. 34, no. 2, pp. 496–506, Feb. 2015.
[4] S. Antani, L. R. Long, G. R. Thoma, and D.-J. Lee, "Evaluation of shape indexing methods for content-based retrieval of X-ray images," *Proc. SPIE*, vol. 5021, pp. 405–416, 2003.
[5] X. Xu, D.-J. Lee, S. Antani, and L. R. Long, "A spine X-ray image retrieval system using partial shape matching," *IEEE Trans. Inf. Technol. Biomed.*, vol. 12, no. 1, pp. 100–108, Jan. 2008.
[6] H. Muller, A. Rosset, J.-P. Vallee, and A. Geissbuhler, "Comparing features sets for content-based image retrieval in a medical-case database," *Proc. SPIE*, vol. 5371, pp. 99–109, 2004.
[7] H. C. Akakin and M. N. Gurcan, "Content-based microscopic image retrieval system for multi-image queries," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 4, pp. 758–769, Jul. 2012.
[8] J. C. Felipe, A. J. Traina, and C. Traina, Jr, "Retrieval by content of medical images using texture for tissue identification," in *Proc. 16th IEEE Symp. Comput.-Based Med. Syst.*, 2003, pp. 175–180.
[9] A. J. Traina, C. A. Castañón, and C. Traina, Jr, "Multiwavemed: A system for medical image retrieval through wavelets transformations," in *Proc. 16th IEEE Symp. Comput.-Based Med. Syst.*, 2003, pp. 150–155.
[10] A. Quddus and O. Basir, "Semantic image retrieval in magnetic resonance brain volumes," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 3, pp. 348–355, May 2012.
[11] A. B. L. Larsen, J. S. Vestergaard, and R. Larsen, "HEp-2 cell classification using shape index histograms with donut-shaped spatial pooling," *IEEE Trans. Med. Imag.*, vol. 33, no. 7, pp. 1573–1580, Jul. 2014.
[12] T. Ojala, M. Pietikainen, and T. Maenpaa, "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.
[13] D. Unay, A. Ekin, and R. S. Jasinschi, "Local structure-based region-of-interest retrieval in brain MR images," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 4, pp. 897–903, Jul. 2010.
[14] L. Sørensen, S. B. Shaker, and M. De Bruijne, "Quantitative analysis of pulmonary emphysema using local binary patterns," *IEEE Trans. Med. Imag.*, vol. 29, no. 2, pp. 559–569, Feb. 2010.
[15] S. Murala, R. Maheshwari, and R. Balasubramanian, "Directional binary wavelet patterns for biomedical image indexing and retrieval," *J. Med. Syst.*, vol. 36, no. 5, pp. 2865–2879, 2012.
[16] B. Li and M. Q.-H. Meng, "Tumor recognition in wireless capsule endoscopy images using textural features and SVM-based feature selection," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 3, pp. 323–329, May 2012.
[17] S. Murala and Q. J. Wu, "Local ternary co-occurrence patterns: A new feature descriptor for MRI and CT image retrieval," *Neurocomputing*, vol. 119, pp. 399–412, 2013.
[18] S. Murala and Q. J. Wu, "Spherical symmetric 3D local ternary patterns for natural, texture and biomedical image indexing and retrieval," *Neurocomputing*, vol. 149, pp. 1502–1514, 2015.
[19] S. R. Dubey, S. K. Singh, and R. K. Singh, "Local diagonal extrema pattern: A new and efficient feature descriptor for CT image retrieval," *IEEE Signal Process. Lett.*, vol. 22, no. 9, pp. 1215–1219, Sep. 2015.

[20] S. R. Dubey, S. K. Singh, and R. K. Singh, "Local wavelet pattern: A new feature descriptor for image retrieval in medical CT databases," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5892–5903, Dec. 2015.
[21] S. R. Dubey, S. Singh, and R. Singh, "Local bit-plane decoded pattern: A novel feature descriptor for biomedical image retrieval," *IEEE J. Biomed. Health Informat.*, vol. 20, no. 4, pp. 1139–1147, Jul. 2016, doi: 10.1109/JBHI.2015.2437396.
[22] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, and S. Mougiakakou, "Lung pattern classification for interstitial lung diseases using a deep convolutional neural network," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1207–1216, May 2016.
[23] Z. Yan *et al.*, "Multi-instance deep learning: Discover discriminative local anatomies for bodypart recognition," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1332–1343, May 2016.
[24] K. Sirinukunwattana, S. E. A. Raza, Y.-W. Tsang, D. R. Snead, I. A. Cree, and N. M. Rajpoot, "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1196–1206, May 2016.
[25] S. Mallat, "Recursive interferometric representation," in *Proc. 18th Eur. Signal Process. Conf.*, 2010, pp. 716–720.
[26] J. Bruna, "Scattering representations for recognition," Ph.D. dissertation, Ecole Polytechnique, Palaiseau, France, 2013.
[27] J. Bruna and S. Mallat, "Classification with scattering operators," in *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, 2011, pp. 1561–1566.
[28] J. Bruna and S. Mallat, "Invariant scattering convolution networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1872–1886, Aug. 2013.
[29] E. Tola, V. Lepetit, and P. Fua, "DAISY: An efficient dense descriptor applied to wide-baseline stereo," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 5, pp. 815–830, May 2010.
[30] Y. Y. Tang, B. F. Li, H. Ma, and J. Lin, "Ring-projection-wavelet-fractal signatures: A novel approach to feature extraction," *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 45, no. 8, pp. 1130–1134, Aug. 1998.
[31] Y. Y. Tang, Y. Tao, and E. C. Lam, "New method for feature extraction based on fractal behavior," *Pattern Recognit.*, vol. 35, no. 5, pp. 1071–1081, 2002.
[32] G. Liu, Z. Lin, and Y. Yu, "Radon representation-based feature descriptor for texture classification," *IEEE Trans. Image Process.*, vol. 18, no. 5, pp. 921–928, May 2009.
[33] R. Lan and J. Yang, "Orthogonal projection transform with application to shape description," in *Proc. 17th IEEE Int. Conf. Image Process.*, 2010, pp. 281–284.
[34] R. Lan, J. Yang, Y. Jiang, C. Fyfe, and Z. Song, "Whitening central projection descriptor for affine-invariant shape description," *IET Image Process.*, vol. 7, no. 1, pp. 81–91, 2013.
[35] H. M. Wallach, "Topic modeling: Beyond bag-of-words," in *Proc. 23rd Int. Conf. Mach. Learn.*, 2006, pp. 977–984.
[36] E. Nowak, F. Jurie, and B. Triggs, "Sampling strategies for bag-of-features image classification," in *Proc. 9th Eur. Conf. Comput. Vision*, 2006, pp. 490–503.
[37] T. Li, T. Mei, I.-S. Kweon, and X.-S. Hua, "Contextual bag-of-words for visual categorization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 4, pp. 381–392, Apr. 2011.
[38] P. Lo *et al.*, "Extraction of airways from CT (EXACT'09)," *IEEE Trans. Med. Imag.*, vol. 31, no. 11, pp. 2093–2107, Nov. 2012.
[39] K. Clark *et al.*, "The Cancer Imaging Archive (TCIA): Maintaining and operating a public information repository," *J. Digit. Imag.*, vol. 26, no. 6, pp. 1045–1057, 2013.
[40] NEMA-CT Image Database. (2012). [Online]. Available: ftp://medical.nema.org/medical/Dicom/Multiframe/
[41] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognit.*, vol. 29, no. 1, pp. 51–59, 1996.
[42] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1635–1650, Jun. 2010.
[43] B. Zhang, Y. Gao, S. Zhao, and J. Liu, "Local derivative pattern versus local binary pattern: Face recognition with high-order local pattern descriptor," *IEEE Trans. Image Process.*, vol. 19, no. 2, pp. 533–544, Feb. 2010.
[44] S. Murala, R. Maheshwari, and R. Balasubramanian, "Local tetra patterns: A new feature descriptor for content-based image retrieval," *IEEE Trans. Image Process.*, vol. 21, no. 5, pp. 2874–2886, May 2012.
[45] S. Murala and Q. Wu, "Local mesh patterns versus local binary patterns: Biomedical image indexing and retrieval," *IEEE J. Biomed. Health Informat.*, vol. 18, no. 3, pp. 929–938, May 2014.



Rushi Lan received the B.S. and M.S. degrees from Nanjing University of Information Science and Technology, Nanjing, China, in 2008 and 2011, respectively, and the Ph.D. degree from the University of Macau, Macau, China, in 2016.

He is currently a Research Assistant in the Department of Computer and Information Science, University of Macau. His research interests include image classification, image denoising, and metric learning.

Yicong Zhou (M'07–SM'14) received the B.S. degree from Hunan University, Changsha, China, in 1992, and the M.S. and Ph.D. degrees from Tufts University, Medford, MA, USA, in 2006 and 2008, respectively, all in electrical engineering.

He is currently an Associate Professor and Director of the Vision and Image Processing Laboratory, Department of Computer and Information Science, University of Macau, Macau, China. His research interests include chaotic systems, multimedia security, image processing and understanding, and machine learning.

Dr. Zhou is an Associate Editor for the *Journal of Visual Communication and Image Representation* and a leading chair of IEEE Systems, Man, and Cybernetics Society Technical Committee on Cognitive Computing. He was a recipient of the third price of Macau Natural Science Award in 2014.