Representing Medical Images With Encoded Local Projections

Hamid R. Tizhoosh, *Senior Member, IEEE*, and Morteza Babaie

***Abstract*—This paper introduces the "encoded local projections" (ELP) as a new dense-sampling image descriptor for search and classification problems. The gradient changes of multiple projections in local windows of gray-level images are encoded to build a histogram that captures spatial projection patterns. Using projections is a conventional technique in both medical imaging and computer vision. Furthermore, powerful dense-sampling methods, such as local binary patterns and the histogram of oriented gradients, are widely used for image classification and recognition. Inspired by many achievements of such existing descriptors, we explore the design of a new class of histogram-based descriptors with particular applications in medical imaging. We experiment with three public datasets (IRMA, Kimia Path24, and CT Emphysema) to comparatively evaluate the performance of ELP histograms. In light of the tremendous success of deep architectures, we also compare the results with *deep features* generated by pretrained networks. The results are quite encouraging as the ELP descriptor can surpass both conventional and deep descriptors in performance in several experimental settings.**

***Index Terms*—Deep features, histopathology images, image classification, image retrieval, LBP, medical image retrieval, projections, radon transform.**

I. Introduction

DESPITE a large number of descriptors being available, the immensely diverse nature of digital images and different requirements of each application field necessitate continuous innovation and extension of existing search and recognition algorithms. One of the domains witnessing such extensive innovations is medical imaging in which search and classification have many applications [1]–[3]. While most well-known descriptors are based on robust local information in small windows, medical images require other operations in local windows to capture relevant anatomical primitives. Projection-based descriptors, for instance, may result in a higher level of discrimination if they are measured and encoded properly. The design of such descriptors becomes even more relevant when we recognize the fact that trainable feature extraction methods such as *deep networks* that may surpass handcrafted descriptors such as scale-invariant feature transform (SIFT) and local binary patterns (LBP) might not always be feasible. This is because we cannot always provide a balanced and large set of labeled images in the medical field [4]. However, using pre-trained deep networks as feature extractors is perhaps a more viable option in many applications including medical imaging, where, generally, only small to medium size data is available.

In this study, *Encoded Local Projections* (ELP), a new image descriptor, was developed based on projections that capture spatial patterns by encoding local changes in the shapes of specific projections of gray-level images. Our design process was motivated by the challenges that the research community had experienced in retrieving medical images using keypoint-based approaches compared to dense-sampling algorithms [5]–[7]. Specifically, we could not duplicate the same level of success for medical images that one generally expects from applying commonly used descriptors for face, scene, and object recognition. Moreover, we realized that one may combine the essential traits of existing powerful descriptors to design a new descriptor that may be more suitable for medical images.

From the application perspective, the focus of this study is on image search and classification to verify the expressiveness of the proposed ELP descriptor. Based on the literature on the effectiveness of such methods [6], we investigate *dense-sampling* and *histogram-based* descriptors of short length that can be employed for tagging big image data. We compare our proposed descriptor with LBP and histogram of oriented gradients (HOG). In addition, we experiment with deep features in light of their immense success for non-medical cases although deep features are, unlike LBP and HOG, not handcrafted and are generally high-dimensional. Three publicly available image datasets (x-rays, CT, and histopathology) were used to validate the performance of the proposed ELP descriptor. The results confirm the potential of the ELP to be a consistently accurate image descriptor.

II. Related Literature

A. Image Descriptors

Conventional or handcrafted descriptors have been used for quite some time [8], [9]. A major group is "keypoint"-based descriptors [10]. SIFT and speeded-up robust features (SURF) [11], [12] belong to the most commonly used keypoint detectors and feature descriptors for various applications [13],

Manuscript received October 17, 2017; revised December 13, 2017; accepted January 3, 2018. Date of publication January 10, 2018; date of current version September 18, 2018. This work was supported by the Natural Sciences and Engineering Research Council of Canada in form of a Discovery Grant. *(Corresponding author: Hamid R. Tizhoosh.)*

H. R. Tizhoosh is with the Laboratory for Knowledge Inference in Medical Image Analysis, Faculty of Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: tizhoosh@uwaterloo.ca).

M. Babaie is with the Department of Mathematics and Computer Science, Amirkabir University of Technology.

Digital Object Identifier 10.1109/TBME.2018.2791567

[14]. While most keypoint detection schemes extract real-valued feature vectors, algorithms such as binary robust invariant scalable keypoints (BRISK) [15] use binary feature descriptors for which image search queries need to be performed within a short time [16].

A second group of handcrafted image descriptors is the "histogram"-based operators that generally extract a compact image representation in the form of a histogram assembled by counting local patterns or gradient directions. LBP [17], [18] and HOG [19]–[21] are perhaps the most prominent among these. LBP was designed with texture classification in mind, whereas HOG originally targeted human recognition, particularly pedestrians. Different versions of LBP histograms have demonstrated extremely high power of discrimination for a range of applications [22]–[24]. The most recent extension of LBP, median robust extended LBP (MRELBP), delivers very impressive results on texture patterns [25].

Employing *deep features* as image descriptors is a rather recent development, predominantly based on convolutional neural networks (CNNs), which are trained from scratch or used after training for classification to extract high-dimensional vectors embedded in the pooling or fully connected layers [26]–[28]. CNNs and other discriminative deep architectures require a large volume of labeled (and balanced) data to be optimally trained without the drawback of overfitting [29]–[31].

In this paper, we propose a projection-oriented and histogram-based scheme that operates in local windows of gray-level images. Therefore, we also examine the literature on Radon transform.

B. Radon Transform

In order to capture the patterns in an image $\mathbf{I}$ as a 2D function $f(x, y)$, one can project $f(x, y)$ along a number of parallel projections (in contrast to fan beam projections) at different angles θ. A projection is the sum (integral) of $f(x, y)$ values along the parallel lines constituted by each angle θ to create a new image $R(\rho, \theta)$ with $\rho = x\cos\theta + y\sin\theta$. The Radon transform can be given as

$$R(\rho, \theta) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y)\delta(\rho - x\cos\theta - y\sin\theta)dxdy, \tag{1}$$

where $\delta(\cdot)$ represents the Dirac delta function.

Radon transform can generally be used for reconstructing objects/scenes from parallel projections. However, there are many other applications reported in literature [32]–[34]. Radon composite features have been used to transform binary shapes into 1D representations for feature calculation [35]. Tabbone *et al.* proposed a histogram of the Radon transform invariant to geometrical transformations [36]. However, the histogram was restricted to counting the length of binary shapes. Daras *et al.* generalized the Radon transform to radial and spherical integration to search for 3D models of diverse shapes [37]. Trace transform is also a generalization of Radon transform [38] for invariant features using tracing lines applied on shapes with complex textures on a uniform background to detect change. Heutte *et al.* used the projections of binary images for character recognition [39]. A different usage of parallel projections, the variance of Radon projections, has been applied to register texture images to subsequently extract wavelet features [40]. The idea of Radon barcodes was introduced recently to binarize all projections (lines) in individual directions using either a "local" threshold for that angle, or based on the rise and fall of the projection amplitude (called the *MinMax* method) [41]–[43].

Obviously, most Radon-based approaches have been applied to binary images (i.e., binary shapes). However, we need to extract features from gray-level images. The latest developments using projections as descriptors have been restricted to "global" projections (applied to the entire image) incapable of recording the spatial patterns. One needs to zoom into image details to capture rich features.

The *novelty* of the descriptor proposed in this study is its ability to capture local projections in gray-level images in the form of a descriptive histogram. The challenge, hence, is to apply projections on gray-level images to local neighborhoods and generate short-length descriptors by counting the frequency of a suitable quantity.

III. Encoded Local Projections – ELP

Many different local descriptors have used the method of spatial binning and stacking of local histograms [44], [45]. The phrase "histogram of projections" has already been used in literature [39]. However, it has always meant to count some sort of *black and white frequencies* when examining binarized shapes, without any systematic relationship to binning any type of local gray-level patterns and gradients. Applying projections in small local windows followed by counting a meaningful quantity to assemble a histogram is a challenge that we address in this section.

By examining an image $\mathbf{I}$, based on projections $\mathbf{p}_\theta$, extracted using Radon transform $R(\theta, \rho)$ along parallel lines ρ and at certain angles θ, we aim to generate a histogram $\mathbf{h}$ that captures significant image attributes in (small) local neighborhoods $\mathbf{W}_{ij}$.

First, we establish that a maximum of 180 projections are sufficient for our purpose as $\mathbf{p}_\theta = \text{flip}(\mathbf{p}_{\theta+\pi})$. Second, we know that every projection $\mathbf{p}_\theta$ captures a pattern arrangement along ρ at a certain angle $\theta \in \{0^\circ, 1^\circ, \ldots, 179^\circ\}$. Third, it must be obvious that to efficiently create a histogram $\mathbf{h}$ of projections, we must select a relatively small number of projections in each neighborhood $\mathbf{W}_{ij}$. Fourth, not all spatial windows may have something to offer (because of reasons such as the image background). Hence, one may need to skip *homogenous* windows. Finally, gradient information of some sort must be collected and used during this process to capture the change in projection shapes.

Based on the aforementioned elaborations and thoughts, we must find specific projections that can uniquely describe the gradient of patterns (i.e., projections) in local neighborhoods. Intuitively, we know that flat or homogenous projections may be of less interest to us as they do not contain significant primitives such as edges and corners. Hence, we calculate the homogeneity

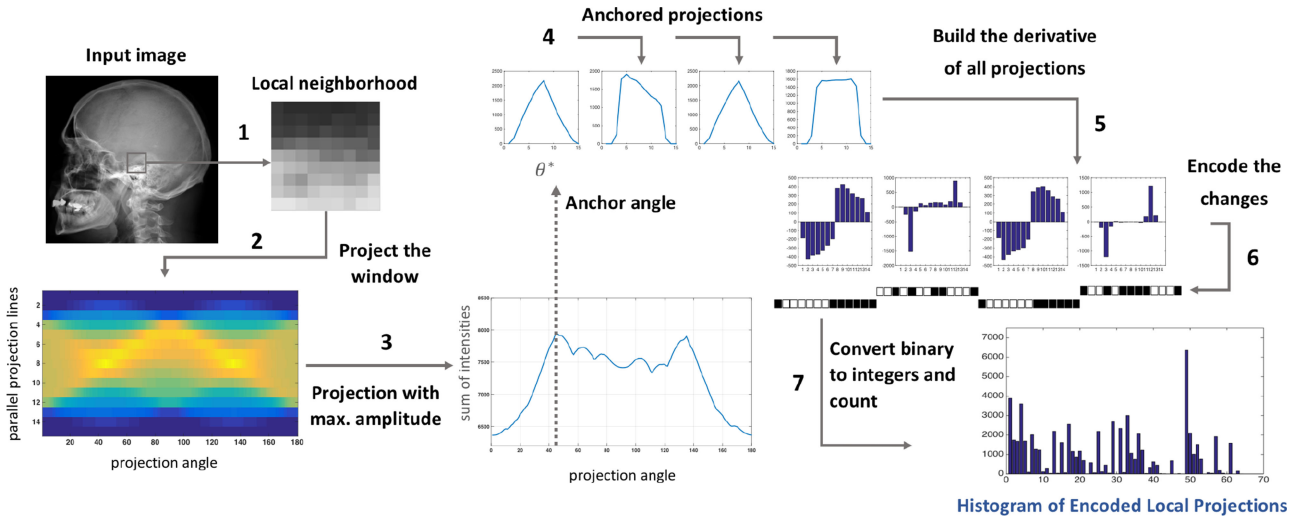

Fig. 1. The steps to extract the ELP histogram: 1) Local windows are selected, 2) 180 projections of each window are calculated (this is called a *sinogram*), 3) the anchor projection θ^* is located using maximum amplitude detection, 4) anchored projections $\theta^* + \alpha_1^\circ, \theta^* + \alpha_2^\circ, \theta^* + \alpha_3^\circ$ are isolated (x-axes depict ρ, and y-axes depict $R(\rho, \theta)$ in (1)), 5) the derivatives of projections are calculated (x axes depict ρ, and y axes depict $\frac{\partial}{\partial \rho} R(\rho, \theta)$), 6) projection gradients are encoded (see Fig. 2), and 7) binary numbers are converted to integers and counted to construct a histogram.

of each window $\mathbf{W}_{ij}$ according to [46]

$$H = 1 - \frac{1}{2^{n_{\text{bits}}}} \sqrt{\sum_i \sum_j (\mathbf{W}_{ij} - m)^2}, \tag{2}$$

where $m = \text{median}_{i,j} \mathbf{W}_{ij}$ and n_{bits} denote the number of bits used to encode the image (e.g., $n_{\text{bits}} = 8, 12, 16$). Only windows with sufficient heterogeneity will be processed. We are interested in finding projections that capture a significant change (assuming that impulsive noise does not play a dominant role, which can be easily ascertained by pre-filtering the image, or by downsampling). Therefore, we must find a special angle θ^* with respect to its gradient $\frac{\partial}{\partial \rho} R(\rho_i, \theta^*)$ across parallel lines ρ_i. Hence, we can find θ^* using

$$\theta^* = \underset{i}{\text{argmax}} \int_j \frac{\partial}{\partial \rho} R(\rho_j, \theta_i). \tag{3}$$

The sum of gradient values of a projection can be used as a measure of how much change that projection is representing. Discretely, we get:

$$\theta^* = \underset{i}{\text{argmax}} \sum_j \left[R(\rho_{j+s}, \theta_i) - R(\rho_j, \theta_i) \right], \tag{4}$$

where $s \in 1, 2, 3, \ldots$ is a proper (small) step. One may, as a more efficient compromise, simply search for the angle θ^*, whose projection, $\mathbf{p}$, has the maximum amplitude:

$$\theta^* = \underset{i}{\text{argmax}} \max[R(\rho_1, \theta_i), \ldots, R(\rho_{|\mathbf{p}|}, \theta_i)]. \tag{5}$$

We call the projection $\mathbf{p}_{\theta^*}$ at θ^* the *anchor projection*. Obviously, the anchor projection gives us a certain level of robustness against rotation because, regardless of the orientation of the local window, we would find the same projection based on its amplitude. However, we must extract more information from the anchor projection obtain have sufficient quantities to count for forming a histogram $\mathbf{h}$. For instance, we can allow a small number of *anchored* equidistant projections starting at θ^*.

For $n \times n$ neighborhoods $\mathbf{W}$, we have $|\mathbf{p}| = n$, provided we neglect zero-padding for diagonal projections, which is equivalent to projecting only within a circle inscribed in the image/window. Subsequently, we need some type of *encoding* to facilitate meaningful and efficient frequency recording. For this purpose, we employ "MinMax" encoding [42] but we apply it on the gradient of the anchored projections. Given the projection vector $\mathbf{p}$ of size n and its derivative $\mathbf{p}' = \frac{\partial}{\partial \rho}\mathbf{p}$, the binary encoding $\mathbf{b}$, $\forall i \in \{1, 2, \ldots, n-1\}$, can be given as

$$\mathbf{b}(i) = \begin{cases} 1 & \text{if } \mathbf{p}'(i+1) > \mathbf{p}'(i), \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Each spatial window $\mathbf{W}$ can produce several binary vectors $\mathbf{b}$ if we anchor equidistant projections at θ^*. For four anchored projections, for instance, that can be counted after conversion into decimal numbers, we receive four integers to count. Fig. 1 illustrates all major steps involved in generating the ELP descriptor. Fig. 2 illustrates how the *MinMax* scheme works.

After binary encoding of the derivative of the anchor projection $\mathbf{p}'_{\theta^*}$, we convert $\mathbf{b}$ to an integer d, such that we can increment $\mathbf{h}(d)$. To collect more information, we obtain three additional projections anchored to $\mathbf{p}_{\theta^*}$ (starting at θ^*): $\Theta = \{\theta^*, \theta^* + \alpha_1, \theta^* + \alpha_2, \theta^* + \alpha_3\}$. These could be equidistant projections: $\Theta = \{\theta^*, \theta^* + \pi/4, \theta^* + \pi/2, \theta^* + 3\pi/4\}$. All projections are encoded, converted, and counted in the same manner.

The counting to generate the histogram $\mathbf{h}$ can occur in two different ways: 1) *Merged* histogram (counts the decimals of all binarized derivatives of projections of Θ in one histogram with $|\mathbf{h}| = 2^{|\mathbf{p}|}$), 2) *Detached* histogram (counts the decimals of each binarized derivative of projections in a separate histogram and then concatenates them into one longer histogram $\mathbf{h}$ with $|\mathbf{h}| = |\Theta| \times 2^{|\mathbf{p}|}$). The detached version of the ELP descriptor has a slightly higher discrimination power as we will report in the results section. The histogram is L_2 normalized at the

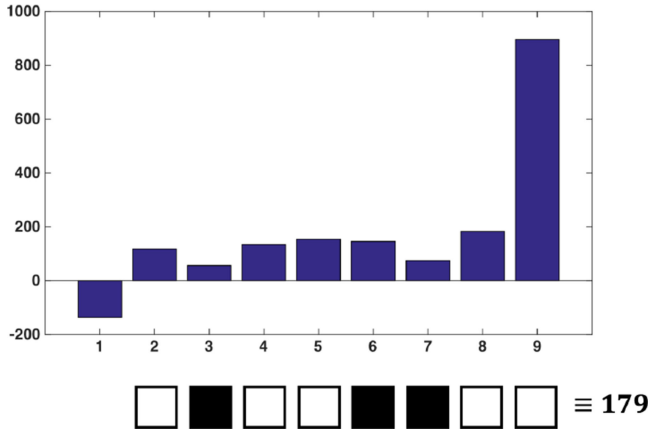

Fig. 2. Encoding the derivative of a projection $\mathbf{p}$ of length n using (6) to create a binary code of length $n-1$. Converted to decimals, a histogram can then be constructed.

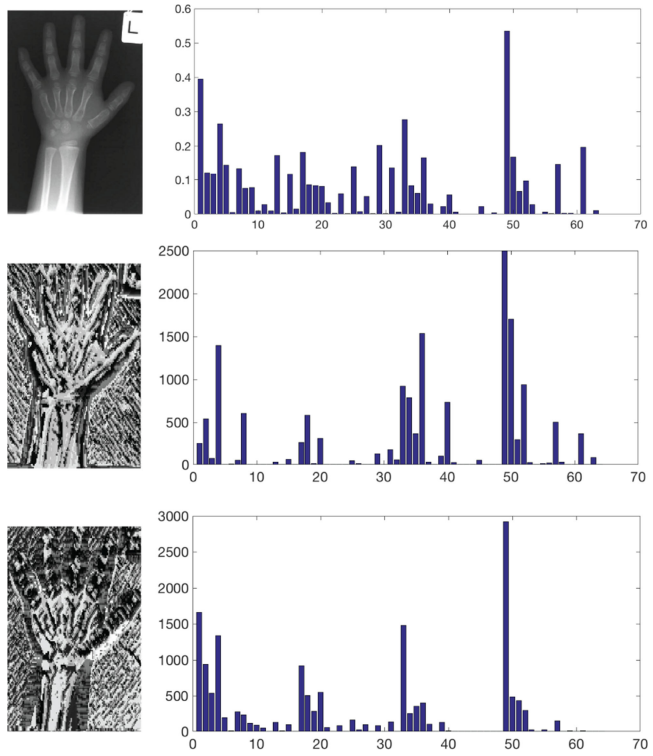

Fig. 3. Visualization of ELP descriptor. Top row: original image with its merged (normalized) ELP histogram (all patterns counted together). Middle row: ELP features for θ^* and its (unnormalized) histogram. Bottom row: ELP features for $\theta^* + 90^\circ$ and its (unnormalized) histogram.

Algorithm 1: Algorithm to extract the ELP histogram.

input : An image $\mathbf{I}$, window size n, homogeneity threshold T_H, type of histogram S
output: A histogram $\mathbf{h}$

[Rows,Cols]$\leftarrow$ `FindDimensions` $(\mathbf{I})$;
$\mathbf{h} \leftarrow \emptyset$
scan through the entire image
for $i \leftarrow n/2$ **to** *Rows*$-n/2$ **do**
  for $j \leftarrow n/2$ **to** *Cols*$-n/2$ **do**
    get a window $\mathbf{W} \leftarrow$`GetWindow` $(\mathbf{I}, i, j, n, w)$;
    calculate the homogeneity
    $H \leftarrow$ `CalcHomogenity` $(\mathbf{W})$;
    if heterogenous, then process the window
    if $H < T_H$ **then**
      get all projections;
      $\mathbf{R} \leftarrow$`ApplyRadon` $(\mathbf{W})$;
      find the anchor projection
      $\theta^* \leftarrow$`FindMaxAmplitude` $(\mathbf{R})$;
      update the histogram
      for $\alpha \in \Theta = \{0^\circ, \alpha_1^\circ, \alpha_2^\circ, \alpha_3^\circ\}$ **do**
        $\theta = \theta^* + \alpha$;
        get the gradient of the projection;
        $\mathbf{p}'_\theta \leftarrow$`GetGradient` $(\mathbf{p}_\theta)$;
        binarize the projection
        $\mathbf{b} \leftarrow$`GetBinary` $(\mathbf{p}'_\theta)$;
        convert binary to decimal
        $d \leftarrow$`ConvertToDecimal` $(\mathbf{b})$;
        $\mathbf{h}_\alpha(d) = \mathbf{h}_\alpha(d) + 1$;
switch S **do**
  case *'Merged':* **do**
    for $k \leftarrow 1$ **to** $|\Theta|$ **do**
      $\mathbf{h}(k) = \sum_\alpha \mathbf{h}_\alpha(k)$
  case *'Detached'* **do**
    $\mathbf{h} \leftarrow$ `ConcatHist` $(\mathbf{h}_0, \mathbf{h}_{\alpha_1}, \mathbf{h}_{\alpha_2}, \mathbf{h}_{\alpha_3})$
  normalize the histogram
  $\mathbf{h} \leftarrow \mathbf{h}/||\mathbf{h}||_2$

end. As the ELP descriptor counts local changes of projections anchored at a characteristic projection (namely, the maximum projection θ^*), the ELP descriptor can be invariant to rotation. Algorithm 1 summarizes all the steps carried out to extract the ELP descriptor from an image. One can display the ELP descriptor by visualizing the decimal values (converted from binary patterns) as pixel values (see Fig. 3).

On Intuition Behind the ELP Histogram – Why should ELP work? We know that generally projections are very useful. They can provide directional silhouettes of scenes, objects, and in our case, organs. Examining several projections from different directions can help assemble a "complete picture", which is how computed tomography works. However, *global* projections (applied across the entire image) are arguably of little use as features because they smash many regions and boundaries together to the limit of indiscriminability, especially if the image is rich in details and only a few directions are used for projection. Hence, it must be obvious-based on the empirical evidence obtained from computer vision- that if there is any discrimination power in projections for gray-level images, it has to be sought in small spatial windows. However, when a descriptor zooms into the details of an image (around key points or randomly), it has to either provide a large number of features (like SIFT and SURF), or it has to encode local information in a suitable manner in order to count patterns and assemble a histogram (like LBP and

HOG). The latter is apparently more desirable considering that, anticipating big image data, we desire to keep our descriptors compact (i.e., of short length). Therefore, we need to ascertain the quantity to count in spatial windows when we are using projections. Each projection is simply a function. One can quantify the changes in these functions using gradient calculations. Flat regions will have no change in their projections and edges and corners will be reflected in some slope changes of the projection. Derivative calculation, hence, can expectedly capture the change. From here, all we need is to (somehow) encode (binarize) this change, convert it to a decimal (as many existing techniques do), and assemble a histogram. Such a descriptor will capture the local projection change into a histogram and is expected to demonstrate a reasonable level of identification capability.

IV. Experiments

For the validation of the proposed ELP descriptor, we focus on image search and classification tasks and use three image datasets: a collection of 14,400 x-ray images (IRMA), a set of 168 CT patches (CT Emphysema), and a set of more than 28,000 histopathology images (KIMIA Path24). Our target is to measure the discrimination power of ELP, in comparison with other dense sampling histograms. As techniques like LBP have been successfully used in conjunction with powerful learning algorithms to deliver the best results for medical image search, (e.g., for IRMA [47]), it is expected that their standalone usage should also provide competitive results. LBP's discrimination power is being increasingly applied for medical image analysis [48], [49].

Sections IV-A, IV-B and IV-C describe the three image datasets that we have used. Section IV-D provides an overview of the comparison with other algorithms. Section IV-E discusses the parameter settings for all experiments. Subsequently, we run *two series of experiments*: 1) We test image retrieval using distance measurements and analyze the results for all datasets (Sections IV-F, and IV-G) 2) We test image classification using SVM and analyze the results for the largest dataset (Sections IV-H and IV-I). A note on pre-trained versus fine-tuned deep networks is provided in Section IV-J.

A. IRMA X-Ray Dataset

IRMA is a collection of several datasets. We use one of them, which is a collection of 14,410 x-ray images specifically collected and marked for CBIR tasks. It has been created from clinical cases at the Department of Diagnostic Radiology of the RWTH Aachen University [50].[1] Each image in the dataset is tagged with an IRMA code comprised of four mono-hierarchical axes with three to four positions each: the technical code (T) for imaging modality, directional code (D) for body orientations, anatomical code (A) for the body region being imaged, and biological code (B) for the biological system examined. The IRMA code, therefore, is a string of 13 characters TTTT-DDD-AAA-BBB, each in $\{0, 1, \ldots, 9; a, b, \ldots, z\}$ [50]. The IRMA error is defined as [51]

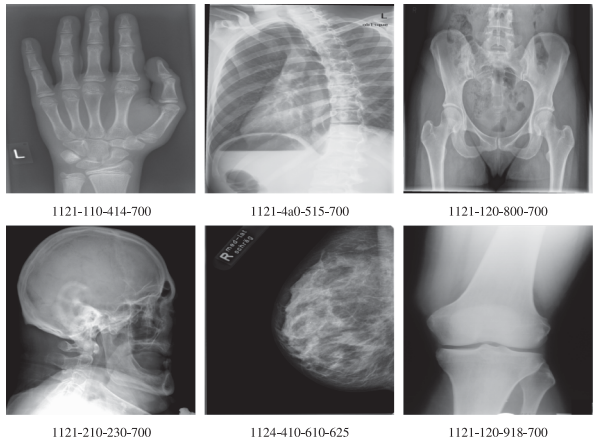

Fig. 4. Sample images from IRMA dataset (14,400 x-ray images with their IRMA codes).

$$\text{error} = \sum_{i=1}^{n_{\text{char}}} \frac{1}{b_i} \frac{1}{i} g(l_i, \hat{l}_i), \tag{7}$$

where l_i is the code of the query image, $\hat{l}_i$ the code of the retrieved image, b_i the number of possible states for each position, n_{char} is the number of characters on the axis, and $g(\cdot)$ is a function that delivers a number in $[0, 1]$ for correct/wrong matchings (hence, the error is a partial value and not Boolean). The total error E is then calculated over all axes and accumulated over all 1,733 test images. The errors are subsequently normalized; completely wrong axis decisions are assigned an error of 0.25 and a correct axis an error of 0. Thus, an image in which all positions in all axes are wrong has an error count of 1, and an image in which all positions in all axes are correct has an error count of 0. The accuracy for IRMA retrieval can be calculated with:

$$A_{\text{IRMA}} = 1 - \frac{1}{1733} \sum_{i=1}^{n_{\text{char}}} \frac{1}{b_i} \frac{1}{i} g(l_i, \hat{l}_i). \tag{8}$$

We used the Python implementation provided by *ImageCLEFmed09* to compute the errors based on the aforementioned definitions.[2] Fig. 4 shows some sample images along with their corresponding IRMA codes. We resized the images to 200 $\times$ 200 for all methods.

B. Kimia Path24 Dataset

The Kimia Path24 dataset has been created from 350 whole scan images (WSIs) depicting diverse body parts [52]. As reported in the initial paper, the images have been captured by *TissueScope LE 1.0*[3] in the bright field using a 0.75 NA lens. A total of 24 WSIs have been selected based on visual distinction for non-clinical experts. This means that a subset of the WSIs,

[1]To download the IRMA dataset, visit https://ganymed.imib.rwth-aachen.de/irma/datasets_en.php

[2]http://www.imageclef.org/

[3]http://www.hurondigitalpathology.com/

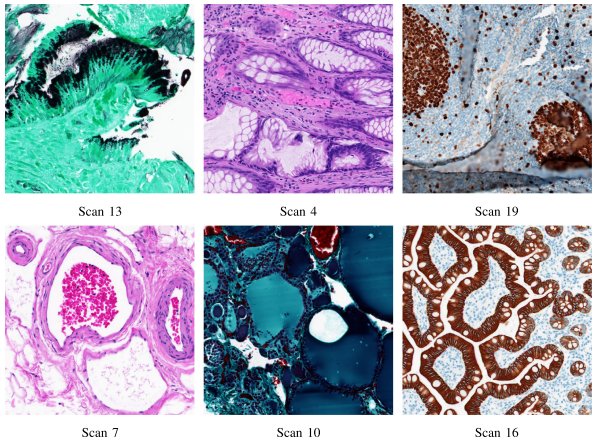

Fig. 5. Sample images from KIMIA Path24 dataset (more than 28,000 images with their classes).

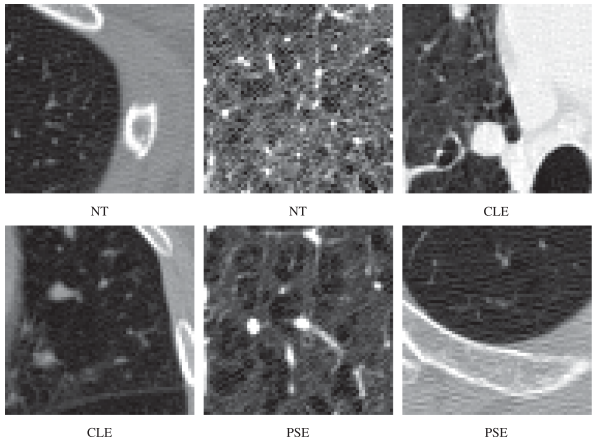

Fig. 6. Sample images from CT Emphysema dataset (168 patches with their classifications).

chosen from the 350 scans, has been selected so that the images clearly represent different *texture* patterns [52].

The patches extracted from the scans are 1000 × 1000 pixels that correspond to 0.5 mm × 0.5 mm. Background pixels are ignored by analyzing the homogeneity and the gradient change for each patch and a threshold is used to exclude background patches (which are widely homogenous and do not exhibit significant gradient information). A total of 1,325 patches have been manually selected to make sure that the patches are representing the dominant WSI textures. Each selected patch has been removed from the scan and saved separately as a testing patch. The remaining parts of the WSI can be used to construct a training dataset. We ensured that no overlap occurred and extracted 27,050 training patches of size 1000 × 1000.

The dataset has a total of 1,325 patches P_s^j that belong to 24 sets $\Gamma_s = \{P_s^i | s \in S, i = 1, 2 \ldots, n_{\Gamma_s}\}$ with $s = 0, 1, 2, \ldots, 23$. Looking at the set of retrieved images R for any experiment, the *patch-to-scan accuracy* can be given as:

$$\text{Accuracy}_{\text{patch-to-scan}} = \frac{1}{1325} \sum_{s \in S} |R \cap \Gamma_s|. \tag{9}$$

We calculate the *whole-scan accuracy* using the following equation:

$$\text{Accuracy}_{\text{whole-scan}} = \frac{1}{24} \sum_{s \in S} \frac{|R \cap \Gamma_s|}{n_{\Gamma_s}}. \tag{10}$$

Hence, the total accuracy η_{total} can be defined to take into account both patch-to-scan and whole-scan accuracies: $\eta_{\text{total}} = \text{Accuracy}_{\text{patch-to-scan}} \times \text{Accuracy}_{\text{whole-scan}}$. The dataset and the code for accuracy calculations can be downloaded from the web.[4] Fig. 5 shows sample patches from the KIMIA Path24 dataset obtained from different scans. We resized the images to 250 × 250 for all methods, except deep networks, for which the images were kept slightly smaller.

C. CT Emphysema Dataset

For this study, we also used the "Computed Tomography Emphysema Database" [53].[5] The database contains 115 high-resolution CT slices with 168 square patches that have been manually annotated in a subset of the slices with an in-plane resolution of 0.78 × 0.78 mm^2, slice thickness of 1.25 mm, a tube voltage of 140 kV, and a tube current of 200 mAs. The 512 × 512 pixel slices depict the upper, middle, and lower part of the lung of each patient. The 168 patches, of size 61 × 61 pixels, are from three different classes, NT (normal tissue, 59 observations), CLE (centrilobular emphysema, 50 observations), and PSE (paraseptal emphysema, 59 observations). The NT patches were annotated in *never smokers*, and the CLE and PSE ROIs were annotated in *healthy smokers* and *smokers with COPD* (chronic obstructive pulmonary disease) in areas of the leading pattern. Fig. 6 shows examples for NT, CLE and PSE classes from the CT Emphysema dataset. Given the set of correctly classified images $\mathcal{C}$, the accuracy A_{CT} can be calculated as

$$A_{\text{CT}} = \frac{|\mathcal{C}|}{168}. \tag{11}$$

D. Comparisons

Because we have been focusing on histogram-based descriptors emerging from "dense sampling", we will compare our results with LBP and HOG as two very successful representatives of such descriptors. As LBP turns out to be the most powerful histogram-based descriptor in many cases, we also conducted experiments with one of its most recent extensions, MRELBP. However, the results were not competitive, which is why we did not report them. Instead, we downloaded the MRELBP from CMVS website.[6] To demonstrate this, we only report the best MRELBP results for the IRMA dataset (8 neighbors, radius 8, radial differences).We used the following MATLAB 2016b

[4] http://kimia.uwaterloo.ca/

[5] http://image.diku.dk/emphysema_database/

[6] http://www.oulu.fi/cmvs/node/33019

implementations of LBP and HOG, and a self-implementation of ELP.[7]

Although our focus was on compact (short-length) descriptors, our recent success cases with deep learning encouraged us to generate *deep features* using a pre-trained *VGG*[8] network with 16 layers trained using the ImageNet database [29], [54], [55] (also see Section IV-J). The usage of such networks as feature extractors has become quite common, specially when there is not enough data to properly train a network from scratch. We also examined "*transfer learning*" [28], [56] to fine-tune the pre-trained VGG network for KIMIA Path24 images. It is argued in literature that this may increase accuracy [57]. After many additional hours of retraining the outer (classifying) layers, however, the results were on par with VGG net without any fine-tuning. Moreover, *AlexNet* was tested but could not surpass the VGG results.

Given two histograms $\mathbf{h}_1$ and $\mathbf{h}_2$, with $|\mathbf{h}_1| = |\mathbf{h}_2| = n$, we can measure their dissimilarity through d_{L_1}, d_{L_2}, and d_{χ^2} by using L_1 and L_2 norms, as well as χ^2 (Chi-squared) distance and cosine similarity. For classification, we used SVM.

Our focus in comparative experiments will be mainly on accuracy. However, we know that algorithms such as ELP and LBP have quadratic upper bounds, $\mathcal{O}(n^2)$, as they use nested loops. Such bounds are more difficult to establish for neural networks because they depend on several factors such as number of hidden layers, number of neurons, and the type of activation function [58]. Generally, neural networks may show exponential time complexity for training but deliver results in $\mathcal{O}(1)$.

E. Parameter Settings

For ELP, we run some experiments on the IRMA dataset and set most of the parameters for all other datasets. We use one pixel stride to shift local windows. We experimented with many different window sizes and found that small window sizes generally deliver good results. Therefore, we tested all methods for all datasets for windows sizes 8×8, 9×9, and 10×10. Based on these experiments, the size of spatial windows was set to 9×9 for IRMA and 10×10 for other datasets. Additionally, $\alpha = \{\theta^*, \theta^* + 45^\circ, \theta^* + 90^\circ, \theta^* + 125^\circ\}$ was used for all datasets (we ran a set of experiments and found that the slight deviation from equidistant projections, 125° instead of 135°, improved the results). For the IRMA dataset, we used nine sub-images. For other datasets, the entire image was processed. For Kimia Path24, the selection of θ^* through maximum detection obtained from the sinogram was replaced by gradient calculation of the window to save time. ELP results are reported as $\mathbf{ELP}_{(\mathbf{w},\mathbf{t}),\mathbf{D}}$ where w is the window size, t is the histogram type (m for merged and d for detached), and D is the distance measure. For the normalization of detached histograms, we experimented with two schemes: we normalized individual histograms and then concatenated them, and we concatenated them first and then normalized the compound histogram. The latter approach provides better results.

[7]The ELP MATLAB code is available at kimia.uwaterloo.ca

[8]Visual Geometry Group, University of Oxford

TABLE I
EFFECT OF WINDOW SIZE ON RETRIEVAL ERROR FOR IRMA DATASET FOR DIFFERENT DISTANCE MEASURES (NO SUB-IMAGES)

	L_2	L_1	χ^2
$\text{ELP}_{[7,m]}$	73.87%	75.70%	76.67%
$\text{ELP}_{[8,m]}$	75.68%	77.77%	78.56%
$\text{ELP}_{[9,m]}$	76.11%	78.52%	79.46%

We ran initial experiments on the IRMA dataset to find out which window sizes are suitable. We expected the projections in the local neighborhood to deliver better results compared to global projections (applied on entire image) or localized projections (applied on sub-images). We found out that a large number of projections may not be possible in commonly used window sizes such as 3×3, or 5×5 but we needed enough projection directions to detect a meaningful "anchor" projection. We tested the smallest windows sizes ($7 \times 7, 8 \times 8, 9 \times 9$) for which we could still project meaningfully and found out that 9×9 was a good choice (see Table I). Slightly larger windows may provide better result for a given image category; however, the computational expense also increases.

For LBP and HOG, we ran exhaustive experiments for each dataset to find the best results. For LBP, we tried all combinations of radii $r = \{1, 2, 3\}$, number of neighbors (8, 12, 16, 20, and 24), rotation invariance versions $\text{rot}_{\text{inv}} = \{\text{yes, no}\}$, and number of sub-images $n_{\text{sub-images}} = \{1, 4, 9, 16\}$. LBP results are reported as $\mathbf{LBP}_{(\mathbf{n},\mathbf{r}),\mathbf{D}}$ where n is the number of neighbors, r is the radius, and D is the distance measure.

For HOG, we tried all combinations of block size $b = \{1, 2, 3\}$, number of sub-images $n_{\text{sub-images}} = \{4, 9, 16, \dots\}$, and number of bins ($= 6, 9, 12, 18, 21$, etc.) as long as $|\mathbf{h}| = n_{\text{sub-images}} \times \text{number of bins} < 1500$.

Deep features, extracted from the pre-trained network, had no parameter to adjust. As mentioned earlier, neither trying a different network (VGG versus AlexNet) nor fine-tuning by slight re-training was able to significantly improve the results.

F. Results Obtained From Distance Calculations

We generated descriptors using LBP, HOG, ELP and a pre-trained VGG net for all images of all datasets. Our experiments started with investigations into LBP and ELP with respect to the effect of descriptor length on the classification accuracy. As Fig. 7 shows, the retrieval error for the IRMA dataset decreases as the descriptor length increases. This was the motivation for us to examine the concatenation of the histograms of multiple sub-images. We confirmed that for all handcrafted descriptors, the concatenation of several histograms of sub-images does in fact increase the discrimination power of the descriptor (HOG intrinsically looks at local blocks).

The purpose of the experiments was to compare the performance of all descriptors for different image categories using direct similarity measurements using distance calculations. We

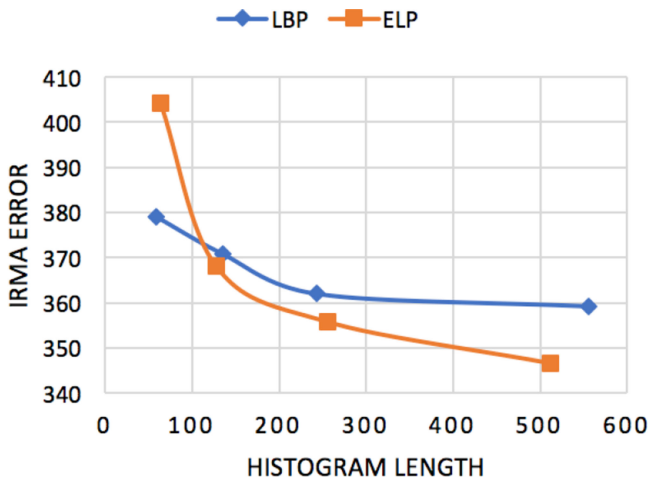

Fig. 7. Both LBP and ELP deliver accurate results when we increase the descriptor length and the motivation to concatenate histograms of multiple sub-images.

TABLE II
BEST RESULTS FOR ALL DATASETS AND DESCRIPTORS WHEN CITY BLOCK (L_1), EUCLIDEAN (L_2), CHI-SQUARED (χ^2) AND COSINE (cos) DISTANCES ARE USED FOR DIRECT SIMILARITY MEASUREMENTS

IRMA Dataset			
(12,677 images for indexing and 1,733 images for testing)			
Method	A_{IRMA}	$\lvert\mathbf{h}\rvert$	$n_{\text{sub-images}}$
$\text{LBP}^{u}_{(8,2),\chi^2}$	85.91%	531	$3 \times 3 = 9$
$\text{ELP}^{*}_{(9,m),L_1}$	85.20%	576	$3 \times 3 = 9$
$\text{ELP}_{(9,m),L_1}$	85.10%	1152	$3 \times 3 = 9$
HOG_{L_1}	84.66%	900	$6 \times 6 = 36$
VGG16-FC7_{L_1}	84.58%	4096	$1 \times 1 = 1$
MRELBP_{χ^2}	75.26%	200	$1 \times 1 = 1$
CT Emphysema Dataset			
(168 images for indexing/testing)			
Method	A_{CT}	$\lvert\mathbf{h}\rvert$	$n_{\text{sub-images}}$
$\text{ELP}_{(10,m),L_1}$	80.95%	256	$1 \times 1 = 1$
$\text{LBP}^{uri}_{(12,3),\chi^2}$	80.36%	18	$1 \times 1 = 1$
VGG16-FC7_{L_2}	69.64%	4096	$1 \times 1 = 1$
HOG_{L_2}	65.47%	1215	$9 \times 9 = 81$
Kimia Path24 Dataset			
(27,000 images for indexing and 1,325 images for testing)			
Method	Accuracy	$\lvert\mathbf{h}\rvert$	$n_{\text{sub-images}}$
$\text{ELP}_{(10,d),\chi^2}$	{71.16%, 68.05%}	1024	$1 \times 1 = 1$
$\text{ELP}_{(10,m),\chi^2}$	{70.70%, 67.93%}	256	$1 \times 1 = 1$
$\text{VGG16-FC7}_{\cos}$	{70.11%, 68.13%}	4096	$1 \times 1 = 1$
$\text{LBP}^{u}_{(24,2),L_1}$	{65.55%, 62.56%}	555	$1 \times 1 = 1$
HOG_{L_1}	{17.58%, 16.76%}	648	$6 \times 6 = 36$

For LBP and HOG, the best results were achieved for each dataset using an exhaustive parameter search. For Kimia Path24, we provide both patch-to-scan and whole-scan accuracies in this order.

did not attempt to beat state-of-the-art benchmarks for these datasets, although we achieved this for the largest one, namely, KIMIA Path24.

Table II shows the results for all three datasets. Over and above the accuracies for each dataset, we also report the length of the descriptor $|\mathbf{h}|$ and the number of sub-images $n_{\text{sub-images}}$.

G. Analysis of Distance-Based Results

Results for IRMA Dataset – As the upper section of Table II shows, although LBP (in uniform version with 8 neighbors and a radius of 2) delivers the highest accuracy, other descriptors are quite close, delivering comparable numbers. ELP (merged histograms) is, with an accuracy of 85.10%, very close to LBP. Reducing the length of ELP by eliminating least-frequent bins, resulting in ELP* improves the results further. LBP achieved its best results with χ^2 whereas all other descriptors achieved their highest accuracies with L_1. One may point out that the performance of deep features, considering their length ($|\mathbf{h}| = 4096$), may be regarded as being rather low. A critical observation was that in many cases, although LBP and ELP had the same IRMA error, ELP was delivering more consistent results with higher semantic matching. Figs. 8 and 9 show two examples of such cases.

Results for CT Emphysema Dataset – For CT patches, both LBP and ELP provided identically high accuracy levels. However, this time, ELP was found to be the most accurate descriptor with the LBP histogram having almost the same accuracy but 14 times shorter than ELP (18 versus 256). The performance of both deep features and HOG considerably dropped for CT patches. HOG's poor performance is perhaps due to the fact that there is no compact object(s) in the images. As for deep features, the patches depict a small part of a large CT scan with large (almost) flat regions so that the image primitives learned through training by ImageNet may not be enough.

Results for KIMIA Path24 Dataset – For histopathology images, both ELP and deep features provide the highest accuracies whereas ELP is, being the shorter descriptor, the better choice. As we may regard histopathology patterns, at least at some resolutions, as *anatomical textures*, and as LBP has been designed for texture recognition, the drop in accuracy of LBP for this dataset is quite surprising. HOG, perhaps expectedly, delivered extremely poor results (there is no *regular* gradient pattern to capture in pathology images).

Summary of Distance-Based Results – While the performance of LBP, HOG and deep features was subject to fluctuations when applied to different image categories, ELP consistently emerged as one of the best methods across the three different datasets. Additionally, visual inspections revealed that the ELP histogram was capable of providing anatomically more meaningful results for x-ray images, an advantage that is not captured by the common error calculations of public datasets. Moreover, the fact that ELP is not only shorter but also more accurate than deep features for histopathology images is an encouraging observation.

H. Results Obtained Through SVM Classification

In a second round of experiments, we examined the effect of classifying the images using a sophisticated classification algorithm like SVM. The IRMA dataset is a typical *retrieval* dataset in which every retrieval task is assigned a number between 0 and 1. Hence, classifying the descriptors for IRMA images may not necessarily provide any insight into the expressiveness of the descriptors. The CT Emphysema dataset is not large enough to draw any reliable conclusions. Hence, we focused on the last dataset (which is also the largest one) to repeat the experiments

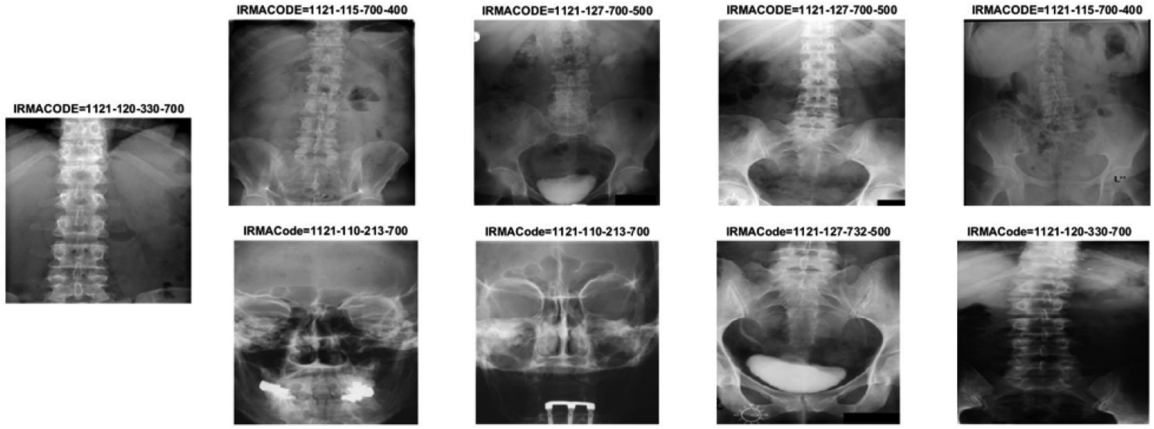

Fig. 8. Although the IRMA errors for ELP and LBP are the same but ELP results are anatomically more consistent. Left: Query, top row: ELP results, bottom row: LBP results.

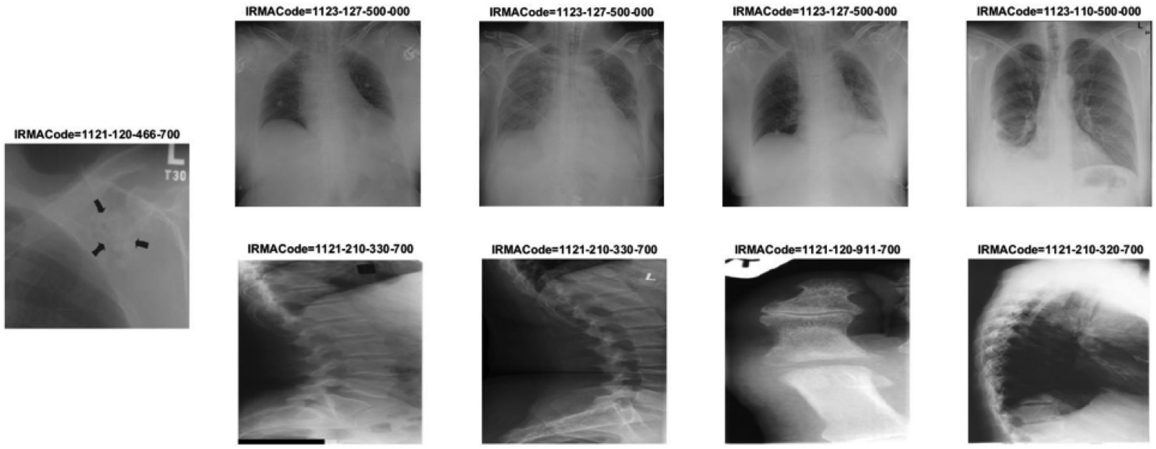

Fig. 9. IRMA errors for ELP and LBP are the same but the ELP descriptor recognizes that the query image is a part of a chest x-ray. Left: Query, top row: ELP results, bottom row: LBP results.

TABLE III
SVM CLASSIFICATION, SORTED BASED ON PATCH-TO-SCAN ACCURACY, OF ELP, DEEP FEATURES, AND LBP IN COMPARISON WITH RESULTS REPORTED FOR KIMIA PATH24 DATASET IN [52] (THE LATTER REPORTED IN THE LAST THREE ROWS)

Method	Accuracy {patch-to-scan, whole-scan}	$\lvert\mathbf{h}\rvert$
$ELP^{SVM}_{(10,d)}$	{82.7%, 79.9%}	1024
$ELP^{SVM}_{(10,m)}$	{82.3%, 79.3%}	256
VGG^{SVM}_{FC7}	{79.5%, 76.9%}	4096
$LBP^{SVM}_{(24,2)}$	{77.8%, 73.3%}	555
VGG^{SVM}_{Pool5}	{72.5%, 67.2%}	6272
$LBP_{(24,3)}$ [52]	{66.1%, 62.5%}	555
CNN [52]	{65.0%, 64.8%}	n.a.
BoVW [52]	{65.0%, 61.0%}	n.a.

with SVM classification of all descriptors instead of using distance calculations.

Table III shows the results for SVM classification of LBP, ELP, and deep features. SVM results for HOG were still found to be quite poor and have not been reported anymore. We experimented with all layers of the VGG network. The best result for a fully connected layer (FC7) was 79.5% for the patch-to-scan accuracy. The best result for a pooling layer (Pool5) was 72.5%. The number of features for Pool5 was 25088 elements and we resampled it to 6272 elements by taking an average of four consequent features. Several SVM configurations failed to generate satisfactory results. Only the linear SVM delivered reasonable results for Pool5.

We also list the benchmark results on Kimia Path24 dataset using CNN, LBP and BoVW (the lower part of Table III highlighted in gray) for comparison. These numbers are reported in [52], which uses a CNN consisting of 3 convolutional layers with 3×3 kernels, each with 2×2 max-pooling with 64, 128, and 256 filters, respectively. The output from the last convolution layer is fed into a fully-connected layer with 1,024 neurons and subsequently to 24 units with softmax activation for classification.

I. Analysis of the SVM Results

A CNN trained from scratch, an LBP with 24 neighbors in a radius of 3 pixels, and a trained BoVW, all deliver accuracies $\approx$65% for the KIMIA Path24 images [52]. SVM-ELP showed the highest patch-to-scan accuracy ($\approx$83%) followed by SVM-Deep with a pitch-to-scan accuracy of $\approx$80%. SVM-LBP

(with 24 neighbors in a radius of 2 pixels) results in an accuracy of ≈78%. ELP descriptor achieves the highest values for both sensitivity (84% ± 14%) and precision (84% ± 14%) with $|\mathbf{h}| = 1024$. Deep features exhibit a sensitivity of 80% ± 10% and a precision of 77% ± 16% with $|\mathbf{h}| = 4096$. Again, it is quite encouraging that ELP is not only 3% more accurate than deep features but also has a much shorter length.

J. A Note on Pre-Training vs. Fine-Tuning

Deep networks are quite powerful tools. A practical way of using them in medical image classification is to employ "pre-trained" networks as we have reported. This eliminates the challenge of not having a large, labelled and balanced image dataset. However, one may also "fine-tune" a pre-trained network in order to better adjust the weights to medical images (e.g., to compensate for lack of color information in our images in a VGG network that has been trained with natural color images). Kieffer *et al.* [57] recently reported that the fine-tuning of the VGG network for the Kimia Path24 actually reduces the patch-to-scan accuracy but slightly increases the whole-scan accuracy. Tajbakhsh *et al.* [4] examine the same question and conclude that this may be an application-based choice.

V. Summary and Conclusions

A new dense-sampling descriptor using parallel projections was introduced in this study. The histogram of *Encoded Local Projections* (also called ELP descriptor) is extracted from local neighborhoods when an anchor projection (i.e., with maximum amplitude) is detected, to which three equidistant projections are anchored. The gradient of these four characteristic projections in each neighborhood is then encoded using the MinMax procedure. We then count the frequency of these encodings when converted to integers. Every projection in small neighborhoods records multiple local (parallel) patterns. Moreover, it bases its discrimination on gradient information to examine how these patterns change.

Experiments on three publicly available datasets demonstrated that ELP has the potential to retrieve medical images with reliable accuracy. In our experiments, ELP, with an almost constant setting, consistently delivered the best results although we exhaustively fine-tuned the competing descriptors LBP and HOG for each dataset separately, and bearing in mind that deep features are the result of extensive training and optimization. Strikingly, we also observed that the ELP descriptor delivered semantically better matches (at least for one dataset), an effect that needs further validation in practice as public datasets do not provide any quantification schemes to capture this benefit for any descriptor.

References

[1] I. Diamant *et al.*, "Task-driven dictionary learning based on mutual information for medical image classification," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 6, pp. 1380–1392, Jun. 2017.

[2] S. Roy *et al.*, "Three-dimensional spatiotemporal features for fast content-based retrieval of focal liver lesions," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 11, pp. 2768–2778, Nov. 2014.

[3] M. Jiang *et al.*, "Computer-aided diagnosis of mammographic masses using scalable image retrieval," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 2, pp. 783–792, Feb. 2015.

[4] N. Tajbakhsh *et al.*, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1299–1312, May 2016.

[5] M. Kashif *et al.*, "Feature description with SIFT, SURF, BRIEF, BRISK, or FREAK? A general question answered for bone age assessment," *Comput. Biol. Med.*, vol. 68, pp. 67–75, 2016.

[6] D. Sargent *et al.*, "Feature detector and descriptor for medical images," *SPIE Med. Imag.*, vol. 7259, 2009, Art. no. 72592Z.

[7] S. Haas *et al.*, "Superpixel-based interest points for effective bags of visual words medical image retrieval," in *MICCAI Int. Workshop Med. Content-Based Retrieval Clin. Decis. Support*, 2011, pp. 58–68.

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

[9] Y. Uchida, "Local feature detectors, descriptors, and image representations: A survey," 2016, arXiv:1607.08368.

[10] S. Krig, "Interest point detector and feature descriptor survey," in *Computer Vision Metrics*. Berkeley, CA, USA: Apress, 2014, pp. 217–282.

[11] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[12] H. Bay *et al.*, "Speeded-up robust features (SURF)," *Comput. Vision Image Understand.*, vol. 110, no. 3, pp. 346–359, 2008.

[13] S. Juan *et al.*, "A scene matching algorithm based on surf feature," in *Proc. 2010 Int. Conf. Image Anal. Signal Process.*, 2010, pp. 434–437.

[14] Z. Zhou, X. Ou, and J. Xu, "SURF feature detection method used in object tracking," in *Proc. 2013 Int. Conf. Mach. Learn. Cybern.*, vol. 4, 2013, pp. 1865–1868.

[15] S. Leutenegger *et al.*, "BRISK: Binary robust invariant scalable keypoints," in *Proc. IEEE Int. Conf. Comput. Vision*, 2011, pp. 2548–2555.

[16] S. Choi and S. Han, "New binary descriptors based on BRISK sampling pattern for image retrieval," in *Proc. 2014 Int. Conf. Inf. Commun. Technol. Convergence*, 2014, pp. 575–576.

[17] T. Ahonen *et al.*, "Face description with local binary patterns: Application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 2037–2041, Dec. 2006.

[18] T. Ojala *et al.*, "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.

[19] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vision Pattern Recog.*, 2005, pp. 886–893.

[20] Q. Zhu *et al.*, "Fast human detection using a cascade of histograms of oriented gradients," in *Proc. 2006 IEEE Comput. Soc. Conf. Comput. Vision Pattern Recog.*, vol. 2, 2006, pp. 1491–1498.

[21] P. A. Torrione *et al.*, "Histograms of oriented gradients for landmine detection in ground-penetrating radar data," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 3, pp. 1539–1550, Mar. 2014.

[22] Y. Anavi *et al.*, "A comparative study for chest radiograph image retrieval using binary texture and deep learning classification," in *Proc. 2015 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2015, pp. 2940–2943.

[23] S. Brahnam *et al.*, *Local Binary Patterns: New Variants and Applications*. New York, NY, USA: Springer, 2014.

[24] S. Chakraborty *et al.*, "Performance enhancement of local vector pattern with generalized distance local binary pattern for face recognition," in *Proc. 2015 IEEE UP Sect. Conf. Electr. Comput. Electron.*, 2015, pp. 1–5.

[25] L. Liu *et al.*, "Local binary features for texture classification: Taxonomy and experimental study," *Pattern Recog.*, vol. 62, pp. 135–160, 2017.

[26] A. Babenko and V. Lempitsky, "Aggregating local deep features for image retrieval," in *Proc. IEEE Int. Conf. Comput. Vision*, 2015, pp. 1269–1277.

[27] K. Chatfield *et al.*, "Return of the devil in the details: Delving deep into convolutional nets," 2014, arXiv:1405.3531.

[28] M. Oquab *et al.*, "Learning and transferring mid-level image representations using convolutional neural networks," in *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, 2014, pp. 1717–1724.

[29] A. Krizhevsky *et al.*, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[30] J. Wan *et al.*, "Deep learning for content-based image retrieval: A comprehensive study," in *Proc. ACM Int. Conf. Multimedia*, 2014, pp. 157–166.

[31] A. Khatami *et al.*, "Parallel deep solutions for image retrieval from imbalanced medical imaging archives," *Appl. Soft Comput.*, vol. 63, pp. 197–205, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1568494617306877.

[32] T. V. Hoang and S. Tabbone, "Invariant pattern recognition using the RFM descriptor," *Pattern Recog.*, vol. 45, pp. 271–284, 2012.
[33] D. Jadhav and R. Holambe, "Feature extraction using radon and wavelet transforms with application to face recognition," *Neurocomputing*, vol. 72, pp. 1951–1959, 2009.
[34] W. Zhao *et al.*, "Retrieval of ocean wavelength and wave direction from SAR image based on radon transform," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2013, pp. 1513–1516.
[35] Y. Chen and Y. Chen, "Invariant description and retrieval of planar shapes using radon composite features," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4762–4771, Oct. 2008.
[36] S. Tabbone *et al.*, "Histogram of radon transform. A useful descriptor for shape retrieval," in *Proc. 19th Int. Conf. Pattern Recog.*, 2008, pp. 1–4.
[37] P. Daras *et al.*, "Efficient 3-d model search and retrieval using generalized 3-d radon transforms," *IEEE Trans. Multimedia*, vol. 8, no. 1, pp. 101–114, Feb. 2006.
[38] A. Kadyrov and M. Petrou, "The trace transform and its applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 8, pp. 811–828, Aug. 2001.
[39] L. Heutte *et al.*, "A structural/statistical feature based vector for handwritten character recognition," *Pattern Recog. Lett.*, vol. 19, no. 7, pp. 629–641, 1998.
[40] K. Jafari-Khouzani and H. Soltanian-Zadeh, "Radon transform orientation estimation for rotation invariant texture analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 6, pp. 1004–1008, Jun. 2005.
[41] H. Tizhoosh, "Barcode annotations for medical image retrieval: A preliminary investigation," in *Proc. 2015 IEEE Int. Conf. Image Process.*, 2015, pp. 818–822.
[42] H. R. Tizhoosh *et al.*, *MinMax Radon Barcodes for Medical Image Retrieval*. New York, NY, USA: Springer, 2016, pp. 617–627.
[43] M. Babaie *et al.*, "Retrieving similar x-ray images from big image data using radon barcodes with single projections," in *Proc. Int. Conf. Pattern Recog. Appl. Methods*, Porto, Portugal, 2017, pp. 557–566.
[44] K. Grauman and T. Darrell, "The pyramid match kernel: Discriminative classification with sets of image features," in *Proc. 10th IEEE Int. Conf. Comput. Vision*, vol. 2, 2005, pp. 1458–1465.
[45] S. Lazebnik *et al.*, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proc. 2006 IEEE Comput. Soc. Conf. Comput. Vision Pattern Recog.*, vol. 2, 2006, pp. 2169–2178.
[46] A. Jurio *et al.*, "New measures of homogeneity for image processing: An application to fingerprint segmentation," *Soft Comput.*, vol. 18, no. 6, pp. 1055–1066, 2014.
[47] Z. Camlica *et al.*, "Medical image classification via SVM using LBP features from saliency-based folded data," in *Proc. 14th Int. Conf. Mach. Learn. Appl.*, 2015, pp. 128–132.
[48] Y.-Y. Liu *et al.*, "Automated macular pathology diagnosis in retinal OCT images using multi-scale spatial pyramid and local binary patterns in texture and shape encoding," *Med. Image Anal.*, vol. 15, no. 5, pp. 748–759, 2011.
[49] S. Wan *et al.*, "Integrated local binary pattern texture features for classification of breast tissue imaged by optical coherence microscopy," *Med. Image Anal.*, vol. 38, pp. 104–116, 2017.
[50] T. Lehmann *et al.*, "The IRMA code for unique classification of medical images," *Proc. SPIE*, vol. 5033, pp. 440–451, 2003.
[51] H. Mueller *et al.*, *ImageCLEF—Experimental Evaluation in Visual Information Retrieval*. Berlin, Heidelberg: Springer, 2010.
[52] M. Babaie *et al.*, "Classification and retrieval of digital pathology scans: A new dataset," in *Proc. IEEE Conf. Comput. Vision Pattern Recognition Workshops (CVPRW)*, 2017, pp. 760–768.
[53] L. Sørensen *et al.*, "Quantitative analysis of pulmonary emphysema using local binary patterns," *IEEE Trans. Med. Imag.*, vol. 29, no. 2, pp. 559–569, Feb. 2010.
[54] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, arXiv:1409.1556.
[55] O. Russakovsky *et al.*, "Imagenet large scale visual recognition challenge," *Int. J. Comput. Vision*, vol. 115, no. 3, pp. 211–252, 2015.
[56] J. Yosinski *et al.*, "How transferable are features in deep neural networks?" in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 3320–3328.
[57] B. Brady Kieffer *et al.*, "Convolutional neural networks for histopathology image classification: Training vs. using pre-trained networks," in *Proc. 7th Int. Conf. Image Process. Theory, Tools Appl.*, 2017.
[58] M. Bianchini and F. Scarselli, "On the complexity of shallow and deep neural network classifiers." in *Proc. Eur. Symp. Artif. Neural Netw., Comput. Intell. Mach. Learn.*, 2014, pp. 371–376.