Journal of Medical Imaging

MedicalImaging.SPIEDigitalLibrary.org

MIRank-KNN: multiple-instance retrieval of clinically relevant diabetic retinopathy images

Parag Shridhar Chandakkar
Ragav Venkatesan
Baoxin Li

MIRank-KNN: multiple-instance retrieval of clinically relevant diabetic retinopathy images

Parag Shridhar Chandakkar,* Ragav Venkatesan, and Baoxin Li

Arizona State University, School of Computing, Informatics and Decision Systems Engineering, Tempe, Arizona, United States

Abstract. Diabetic retinopathy (DR) is a consequence of diabetes and is the leading cause of blindness among 18- to 65-year-old adults. Regular screening is critical to early detection and treatment of DR. Computer-aided diagnosis has the potential of improving the practice in DR screening or diagnosis. An automated and unsupervised approach for retrieving clinically relevant images from a set of previously diagnosed fundus camera images for improving the efficiency of screening and diagnosis of DR is presented. Considering that DR lesions are often localized, we propose a multiclass multiple-instance framework for the retrieval task. Considering the special visual properties of DR images, we develop a feature space of a modified color correlogram appended with statistics of steerable Gaussian filter responses selected by fast radial symmetric transform points. Experiments with real DR images collected from five different datasets demonstrate that the proposed approach is able to outperform existing methods. © 2017 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: [10.1117/1.JMI.4.3.034003](https://doi.org/10.1117/1.JMI.4.3.034003)]

Keywords: diabetic retinopathy; image retrieval; multiple-instance framework; computer-aided diagnosis.

Paper 17093PR received Apr. 13, 2017; accepted for publication Jul. 11, 2017; published online Sep. 1, 2017.

1 Introduction

Diabetic retinopathy (DR) is a consequence of diabetes and it is one manifestation of a systemic disease which can affect a myriad of organ systems. It can cause vision loss if not treated early.¹ Studies showed that 381.8 million patients worldwide were diagnosed with diabetes in 2013.² In 2010, 10.9 million US residents, aged 65 or older, were suffering from diabetes.¹ According to the recent estimates, the number of diabetic patients will rise to 591.9 million by 2035, which indicates a 55% increase in the number of adults with diabetes.² Recent reports show that close to 25,000 people suffering from diabetes become blind every year in the US due to DR.³ It is estimated that DR is the cause for 5% of the world's blind population.⁴ The risk of vision loss due to DR can be significantly reduced by early screening and treatment.^{5,6} Increasing population, cost, limitations of health-care facilities, lack of enough providers in densely populated areas, lack of awareness on the patients' side, and other factors are constraints for regular screening of every diabetic patient. Thus, to identify and treat DR in its early stages requires a new perspective on the problem.

Two important stages of DR are nonproliferative diabetic retinopathy (NPDR) and proliferative diabetic retinopathy (PDR). Microaneurysm (MA) is a sign of DR. MAs rarely cause notable harm but their presence indicates an underlying systematic disorder. MAs can usually be detected by detailed eye examinations. Neovascularization (NV) is a proliferation of functional blood vessels and is a symptom of PDR. Other complications such as vitreous hemorrhages (extravasation of blood around the vitreous body) and/or tractional retinal detachment follow. Timely and effective DR treatment to patients demands reliable detection. Assistance in the form of effective

computer-aided technologies may help improve the sensitivity, consistency, and efficiency of DR severity detection.⁷ Recent advancements in computer-aided diagnosis and medical imaging have propelled the development of automated image analysis techniques to solve the DR problem.⁷ It was observed that a content-based image retrieval (CBIR) approach can help new and experienced ophthalmologists⁸ in a diagnosis. For example, a CBIR system can provide them with both standard reference images and previously diagnosed images containing similar lesions, so that a more precise grading may be achieved. Conventional CBIR techniques may not be effective when directly applied on DR images. Figure 1 show typical fundus images diagnosed with PDR and NPDR against normal ones. Symptoms of NPDR include yellow and waxy exudates. It is noteworthy that very little difference can be observed between the normal images and the NPDR images. Most parts of an affected image would appear to be no different from a normal image as the lesions are localized. A global representation of the image, which is often used in conventional CBIR, may not capture its discriminative characteristics. Properly designed feature extractors evaluated on small regions would be needed to capture lesion-defining information, although such regions need to be determined. Therefore, to achieve the goal of retrieving DR images that are relevant to a given image, we would need a retrieval framework that can take into consideration unknown, localized pathologies. Further, new features need to be designed so that lesion-specific characteristics of an image may be captured.

In this paper, we propose a multiclass multiple-instance clinically relevant retrieval framework called “multiple-instance-rank k -nearest neighbor (MIRank-KNN)” that addresses the problem of localized pathologies in unknown regions. Our

*Address all correspondence to: Parag Shridhar Chandakkar, E-mail: pchandak@asu.edu

Fig. 1 Fundus image of eyes: (a) normal, (b) NPDR, and (c) PDR.

approach is based on the observation that both color and gradient features occurring in small, localized regions characterize the lesions of interest. Thus, for feature extraction, we propose to use spectrally tuned color correlogram (CC) and statistics of steerable Gaussian filter (SGF) response of the points selected by fast radial symmetric transform (FRST). The clinical relevance in our approach is attributed to the multi-instance retrieval framework—MIRank-KNN—and the feature design. For example, MIRank-KNN allows us to retrieve images with similar localization and the number of lesions. The features determine the type of lesions to be retrieved. To facilitate the evaluation of the proposed method, the database and the source code used in this paper are posted on the author's web page.⁹ Upon the decision of the paper, these will be made publicly available to other researchers for further improvement and evaluation.

2 Related Work

Several CBIR systems were previously used in the areas of dermatology, radiology, pathology, and ophthalmology. Research groups have used CBIR for computed tomography, magnetic resonance imaging, positron emission tomography, and retinal images.^{8–21} In the field of positron emission tomography, physiological kinetic features were used to build a CBIR-based system.¹⁰ Hierarchical, spatiotemporal, and evolutionary semantics of neural images were captured by Chu et al.¹⁸ to develop a semantic model for CBIR purposes. Textures and histograms of pathologies were combined in a signature, which was formed on a per-image basis, in a system developed by Kelly et al.¹⁹ This system made use of query-by-example techniques to retrieve images. Korn et al.²⁰ used fast query using nearest-neighbor search to retrieve medical tumor images.

Two of the major works in previous CBIR investigations on ophthalmology include the Structured Analysis of the Retina (STARE) project and the CBIR system developed in 1996.^{12,22} STARE was developed for performing automatic diagnosis of images, annotation of image contents, and searching for similar images. STARE used basic image features to define the similarity metric. A recent work by Quellec et al.⁸ illustrated the importance of CBIR in DR diagnosis. It provided statistical support for the claim that CBIR systems will assist experienced as well as relatively inexperienced ophthalmologists in DR diagnosis. A supervised learning approach that employed features extracted on segmented lesions and macula was used for CBIR of DR images.¹¹ A classifier is trained for detecting lesions with the help of ground-truth data of manually segmented lesions. Certain similar supervised learning approaches use adaptive variants of wavelet transform.^{8,16,17} The DR image retrieval problem was broken into two steps, namely, image background learning and feature extraction.¹⁵ Background of a normal (unaffected) DR image was first learned and then nearest-neighbor retrieval was performed with the help of intensity and texture features. Most of the above state-of-the-art DR-CBIR frameworks rely on a training stage followed by a nearest-neighbor retrieval to give them accurate results.

Lesion detection is handled as a separate problem by many since it is an important step in DR image classification and retrieval. Many solutions have emerged for detecting retinal landmarks, such as MAs, NV, and hemorrhages, from developments in the area of image processing. A bag-of-words scheme was first employed to train individual lesion detectors, and then a high-level classifier was used to determine if patients met the criteria for referral-warranted DR.²³ Interest point detection and visual dictionary were also used²⁴ for individual DR lesion detection. Multiple-instance learning (MIL) was used to find relevant patterns in a DR image after being trained on a database consisting of relevant and irrelevant images.²⁵ Active learning has been employed to select the most informative examples for labeling, thus reducing the load on human experts and obtaining the optimal classification accuracy at the same time.²⁶ Recent machine learning frameworks, such as multiclass MIL frameworks, have shown promising results with images without lesion-level labels.^{13,27,28} A promising fact about all these pioneering works is that all of them suggest that the clinicians will benefit from CBIR systems or the introduction of advanced vision-based approaches to improve the basic systems.

Color and textures are proven to be competitive features in spite of their simplicity.²⁹ CCs tried to model the color correlation in an image.^{30,31} Correlograms are considered better than color histograms and most other color features.³⁰ They incorporate spatial correlation of image pixels that gives them an edge over histogram features. They also describe the global correlation of local spatial correlation of colors. Correlograms are a strong representation of texture with considerably small dimensionality. Gabor features³² and histogram of neighborhood mean moments (HNM)¹⁴ are popular alternatives for describing texture and color of an image, respectively. HNM and Gabor features are widely used approaches in medical image retrieval, but both of them may not produce discriminative features for DR images due to their unique color spectrum. Scale invariant feature transform (SIFT) keypoint detection and description has been shown to be effective for retrieval of radiograph images.²⁹

Though there has been a significant amount of progress in the area of CBIR of ophthalmological images, the lack of a widely

accepted ophthalmological system remains. It shows that the central problems are still unsolved. Some of the shortcomings of the mentioned DR-CBIR retrieval systems are as follows. Most of the systems require a training stage and thus labeled data. For example, manually segmenting lesions requires expertise, a lot of time, and intensive labor. Parameters tuning may not be consistent across different types of images and datasets. To characterize localized lesions, these systems either use local features or extract features only on the detected lesions (usually through another classifier). Though local features are good at characterizing small regions, features of a small but relevant region may get suppressed when all region features are bundled into a long one-dimensional vector. Such a vector is necessary if a nearest-neighbor retrieval scheme is to be used. Another option is to use MIL but that also requires a training stage. Our approach is motivated from the need to overcome these shortcomings. To this end, our contributions are as follows:

1. We develop a completely unsupervised DR image retrieval system. We use the labeling information from the public DR datasets only to calculate retrieval metrics.
2. We develop an indexing and retrieval framework, which considers multiple regions of an image simultaneously. It outperforms nearest-neighbor and citation-KNN retrieval schemes by a large margin, even when state-of-the-art local features are used.
3. We develop a spectrally tuned color quantization scheme aimed toward DR images. We show its superiority over the original correlogram feature through entropy analysis and experiments.
4. Our approach works across five datasets, and its performance varies only slightly over a wide range of parameters.
5. Through our study on DR images and extensive experiments, we propose a unique combination of color and texture features. The proposed features outperform several other state-of-the-art features used in natural and medical image retrieval.

A detailed study of emerging trends in automated analysis of DR images and a discussion on the need for better clinically relevant CBIR systems by Li and Li⁷ clearly shows the space and scope for this research article.

3 Proposed Approach

The proposed retrieval approach can be broken up and studied in two parts.

1. The feature extraction process consisting of color and texture features.
2. MIRank-KNN—the multiple-instance retrieval framework.

3.1 Feature Extraction

We consider a holistic representation of a DR image as opposed to a cascade representation, which usually first identifies individual lesions and then forms the entire feature.^{23,24} Clinicians

Fig. 2 Visualizing the necessity of multiple-instance framework. (a) NPDR image, instances marked in red are the lesions, and (b) normal image.

also work with global representation of an image and perform a diagnosis only after considering all the lesions. We use color and texture properties of an image, which will be detailed in this section.

Color is a distinguishing characteristic for DR images. However, accurate modeling of the color spectrum of DR images can be challenging due to the variation in lighting in the images of the same class. This is illustrated in Fig. 1. In spite of belonging to the same class, the images in the first and last rows have a large difference in their color spectrum. On the other hand, images belonging to different classes may have a very small difference in their color spectrum and often that difference is localized. Figure 2 illustrates this perfectly.

The color feature should be invariant to light color change and shift. It should encode local correlation of colors to model small and localized differences in the color spectrum of two images. Light color change and shift are mathematically represented as

$$\begin{pmatrix} R' \\ G' \\ B' \end{pmatrix} = \begin{pmatrix} cR & 0 & 0 \\ 0 & cG & 0 \\ 0 & 0 & cB \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} + \begin{pmatrix} sR \\ sG \\ sB \end{pmatrix}, \quad (1)$$

where R, G, B and R', G', B' are the current and modified pixel value, respectively. cR, cG, cB and sR, sG, sB are the factors contributing to the lighting change and shift, respectively. The color feature is called invariant if it produces identical feature vectors for both the pixel sets— (R, G, B) and (R', G', B') .

We use CC to represent DR images as it encodes global distribution of local spatial correlation of colors. We modify the correlogram feature to be invariant to the lighting changes, shifts, and most importantly, the unique color spectrum of DR images. They have an almost always saturated red channel as observed in Fig. 1. Additionally, DR images of different classes may have similar histograms, making it difficult for traditional color-based features to produce discriminative feature vectors. Effective retrieval of DR images, therefore, demands spectrally tuned color features. Section 3.1.1 describes in detail the spectrally tuned color correlogram.

3.1.1 Spectrally tuned color correlogram

CC was proposed as an effective color feature which encodes global distribution of local spatial correlation of colors.³⁰ The

CC of an image is a table indexed by color pairs such that the k 'th entry for the color pair (i, j) gives the probability of finding a pixel of color j at a distance k from color i . We quantize the image into $(m =)$ 16 colors and use $k = 1$. We then calculate color correlation of the center pixel with all other pixels in a 3×3 block. A two-dimensional histogram is then defined over the entire image for $i, j \in [m]$ as

$$h_{c_i, c_j}(I) = \Pr_{p_1 \in c_i(p_1), p_2 \in I} [p_2 \in c_j(p_2) | d(p_1, p_2) = 1]. \quad (2)$$

This gives the probability that a pixel belonging to quantized color c_1 has another pixel of color c_2 at a unit distance, where $c_1, c_2 \in [m]$. This models the global distribution of local correlation of colors. The dimensionality of CC is $O(m^2k)$. We quantize the image into 16 colors which gives us a feature vector of 256 dimensions.

The choice of the quantization scheme is crucial to the performance of the CC features. A popular quantization scheme proposed by Li³¹ was designed based on human vision and the stimulus reactions of human vision to various colors. From Fig. 1, it can be easily observed that the spectrum of DR images is entirely different from that of natural images. Therefore, a new quantization scheme has to be designed which is spectrally tuned toward DR images.

We proposed a spectrally tuned quantization scheme for the CC features tuned toward DR images.²⁸ We modify it here to be robust to light color change and shift. We create a transformed color space where all three channels have zero mean and unit variance. Let the new color space be represented by $(R^T, G^T, \text{ and } B^T)$

$$\begin{pmatrix} R^T \\ G^T \\ B^T \end{pmatrix} = \begin{pmatrix} \frac{R - \mu_R}{\sigma_R} \\ \frac{G - \mu_G}{\sigma_G} \\ \frac{B - \mu_B}{\sigma_B} \end{pmatrix}. \quad (3)$$

The transformed color space is invariant against arbitrary light color changes and shifts due to normalization of all channels. Its invariance properties have also been analyzed.³³ This transformed color space will henceforth be used for all operations. The unique shades (i.e., $\langle R^T, G^T, B^T \rangle$ triplets) in all the images are now extracted and arranged in an $M \times 3$ matrix. k -means clustering is performed on the matrix to generate the

Fig. 3 Three-dimensional visualization of Li's quantization and the spectrally tuned quantization schemes for DR image color space. The centroids are indicated by spheres, and points associated with each centroid are shown with cross marks in appropriate colors. (Please zoom in for better viewing.)

spectrally tuned quantization bins (i.e., centroids). Once the bins are obtained, the CC features are calculated for all possible color pairs (i, j) where $i, j \in [1, \dots, 16]$. It is empirically observed that modeling the correlation of all the possible pairs gives better retrieval results. However, changing the number of bins affects the performance only by a small amount. The evaluation of the spectrally tuned CC as well as its comparison with the original CC is given in Sec. 5.

While the spectral tuning of the quantizer attempts to maximize the entropy by keeping the density of points (shades) associated with each centroid (quantized color bin) uniform, the one proposed by Li³¹ has highly varying density of points associated with each centroid. A three-dimensional (3-D) visualization of both the quantization schemes are shown in Fig. 3. It can be seen that while in Li's quantization the distribution of points is non-uniformly associated, the ones in the proposed approach are reasonably well-distributed. The effect clustering has on this can be noticed better in Fig. 4. The spectrally tuned CC makes better utilization of the bandwidth provided by the 16 color bins than the other quantization scheme where most of the shades fall under 4 out of 45 bins, thereby heavily affecting the ability of the CC feature to encode the spatial and global correlation of colors in DR images.

Entropy analysis: We now provide a measure for the amount of information encoded in both approaches using the given number of bins. We further show the superiority of our approach using the developed entropy measure. The entropy is a measure of randomness in the data which in turn corresponds to the amount of information present in it. Entropy of a random variable X consisting of values $(x_1, x_2, \dots, x_n)$ is given as

$$H(X) = -\sum_i P(x_i) \log_b P(x_i), \quad (4)$$

$$\text{where } P(x_i) = \frac{\text{\#shades falling under bin } i}{\text{Total\# of shades in the database}}.$$

Fig. 4 Histogram of quantized shades for Li's and the proposed quantizers, respectively.

We use logarithm base 2 and define $0 * \log 0 = 0$. Let us denote the distribution on the left in Fig. 4 by D_1 and the other distribution by D_2 . The entropies of both distributions are $H(D_1) = 2.9201$ and $H(D_2) = 3.7222$ bits. Though the proposed quantization scheme provides an increase of 27.47%, both the values cannot be compared due to different-sized distributions (16 versus 44 bins). To tackle this problem, various methods such as scaled entropy, normalized entropy, and sliding window approach have been proposed.^{34,35} Instead of using a normalizing/scaling approach, we compare the two distributions by using an indirect method that uses a reference distribution possessing maximum entropy with the given number of bins. Uniform distribution is chosen as reference since it provides the maximum entropy among all discrete distributions supported on a finite set $(x_1, x_2, \dots, x_n)$.³⁶ We measure the percentage difference in the entropy values of the original and its corresponding uniform distribution. Let us define dE_1 as the percentage difference between D_1 and its corresponding uniform distribution, denoted by UD_1 . dE_2 and UD_2 can be similarly defined. By performing entropy calculations, values of UD_1 and UD_2 are found to be 5.4594 and 4 bits, respectively. By comparing dE_1 and dE_2 , it is possible to evaluate the similarity of two different-sized distributions. Since this approach initially compares a distribution with its corresponding uniform distribution, the comparison between two different-sized distributions always happens on a standard scale as desired. The values of dE_1 and dE_2 indicate 46.51% and 7.46% decrease in entropy, respectively. Thus the right-hand side distribution (proposed) packs in much more information with fewer number of bins and thus provides a better feature representation of a DR image as shown in Fig. 5. Though the image on the right does not explicitly capture lesions, it better represents the DR image by packing in more information with the help of spectrally tuned bins and in turn produces a more discriminative feature space.

Fig. 5 Quantization of a DR image using AutoCC³¹ (middle) and the proposed approach (right).

3.1.2 Steerable Gaussian filter

To retrieve clinically relevant images for all the three classes, color features alone are not sufficient. Since the shape and texture of DR lesions are major factors in deciding the severity level of the disease, we make use of SGF coupled with FRST.

SGFs are oriented filters and are used in many computer vision and image processing tasks, such as edge detection and classification of lines, edges, and contours. They are filters of desired orientation obtained from a linear combination of basis filters.³⁷ This is called steering. The ability of SGF to model edges and contours makes it useful for DR images, since edges and contours at various orientations are distinguishing features for different severity levels of DR. The filters shown in Fig. 6 are separated by 45 deg. These filters are used as kernels to

calculate directional derivatives of an image, which help to model contours at different orientations.

In the case of DR images, it means blood vessels, lesions, optic disks, etc. will have different filter responses since each of them has a different structure. Figure 7 shows the filter response of a typical DR image to a filter orientated at 225 deg. The peaks in the SGF response can be attributed to the textural discontinuity at those particular locations. The statistical variation in its values can be modeled by taking the standard deviation, skewness, and kurtosis. This produces a 3-D vector for each angle. For a total of eight angles, this gives a 24-D feature vector. In Fig. 7, though the SGF response rightly peaks at the location DR lesions, many blood vessels are also unnecessarily highlighted. This generates false positives in the feature space, thus SGF response alone cannot be used as

Fig. 6 Some of the bases of SGFs.

Fig. 7 SGF filter response to an NPDR image: (a) input NPDR image and (b) filter response.

Fig. 8 Interest point detection using FRST on (a) normal, (b) NPDR, and (c) PDR images. (d) the FRST interest points superimposed on the SGF response shown in Fig. 7. (Please zoom in for better viewing.)

features. To filter out unwanted high SGF response regions, a second complementary stage is needed which can select regions of interest. To detect such regions, we use FRST.³⁸

3.1.3 Fast radial symmetric transform

Numerous context-free point of interest operators have been proposed in the literature. These work on the principle that points which possess local radial symmetry are the ones which draw human attention. The principle is based on the psychological findings proposed by researchers in the past.^{39–41} Use of FRST is preferred over other point of interest operators because of its superior performance and its speed. It works in linear time. The comparison of FRST with other state-of-the-art interest point detectors is given in Sec. 5.

FRST (implemented using publicly available code⁴²) uses local radial symmetry to highlight interest points. It accounts the contribution of radial symmetry of gradients that are at a distance d from the point under consideration. The pixel at a distance d to the gradient points is called a positive pixel, and the pixel from which the gradient points away is called a negative pixel. The coordinates of these positive and negative pixels are obtained as follows:

$$p_{+ve}^{(-)}(p) = p_{(-)}^{(+)} \text{round}\left(\frac{g(p)}{\|g(p)\|} n\right), \quad (5)$$

where $g(p)$ is the gradient of pixel p and n is the radius. Therefore, $p_{+ve}^{(-)}(p)$ [$p_{-ve}^{(+)}(p)$] gives coordinates of a positive (negative) pixel at a distance n from pixel p . The transform estimates the contribution of each pixel to the symmetry in its neighborhood using the concept of positive and negative pixels instead of calculating the contribution of the local neighborhood to a central pixel. We have used a set of radii $\{1, 3, 5\}$ to robustly detect interest points and to negate the presence of nerves and other spurious components that may lead to noisy feature space. Quantitative analysis shows that examining a small subset of radii gives good approximation to the output.³⁸ The values of radii should be small enough to capture the smallest of lesions. The retrieval performance worsens as the values of radii increase. The effect of parameters of FRST on retrieval performance has been analyzed in Sec. 5.

In Fig. 8, though most of the detected points are lesions in images of affected eyes, false detections do exist. FRST works on local maxima detection. The interest points are then generated by performing nonmaxima thresholding on the image

created by FRST. The detected interest points are superimposed on the original image for visualization purposes (see Fig. 8). Figure 8 shows that the FRST points (shown in red marks) are complementary to the SGF response, thereby forming an effective combination. The procedure to couple the SGF response and FRST is described below.

Once the interest points have been obtained through FRST, an image is divided into 64 blocks by an 8×8 grid. Blocks containing the majority of the black background are removed by thresholding. Every image block is considered independently and a region of 15×15 is selected around each interest point. A 24-D feature vector is calculated for the corresponding 15×15 region in the filtered image (Fig. 7). Considering that there are n number of interest points detected on a given image block, they would correspond to n feature vectors. We take the mean over each dimension to reduce the n feature vectors to a single 24-D vector. The same action is carried out when there are multiple interest points lying in the 15×15 region. If there are no interest points detected in the region, then the feature vector is all zeros. For PDR images, more interest points are likely to be found in the small region as compared to NPDR and normal images. By taking the average of the features of all the interest points in an image block, the NPDR image features obtain fewer values while normal image features get even fewer values. Therefore, instead of removing a region containing zero interest points, replacing its feature vector with zeros produces a more discriminative feature space. The feature space is now suitable for multiple-instance retrieval framework.

3.2 MIRank-KNN

The retrieval scheme we propose here is called MIRank-KNN and it produces a ranked list of all the images in an archived dataset by considering all the blocks in a given image simultaneously. Thus, the proposed scheme emphasizes even a small block of an image containing a lesion. To the best of our knowledge, this is the first attempt at developing an unsupervised approach that retrieves images based on multiple blocks (instances) of the query image. Supervised multiple-instance retrieval methods have been developed that borrow concepts from MIL.^{43–46}

MIL was first introduced by Dietterich et al.⁴⁷ It is a form of supervised learning where the data are in the form of labeled bags. Each bag contains a variable number of instances. The labels are provided on a bag level. A positive bag has at least

one positive instance. In a negative bag, all the instances are negative. The main goal is to classify each bag as positive or negative and if possible, infer instance-level labels. This is a difficult problem due to lack of knowledge of instance labels. There has been a plethora of approaches developed to solve MIL problems. We list only those approaches that laid the foundation and are relevant in this context.

Diverse density (DD) was proposed as a general purpose solution for the MIL problems.⁴⁸ It attempts to find a concept point in the feature space which is close to at least one instance from a positive bag and is away from all the instances in a negative bag. It is also called an instance prototype and has maximum DD. With the knowledge of these instance prototypes, each bag can now be classified based on its distance from the prototype. Expectation-maximization procedure was coupled with DD (EM-DD) to solve MIL problems.⁴⁹ Support vector machines (SVMs) were also employed for MIL.⁵⁰ A KNN-based MIL method called citation-KNN was developed.⁵¹ It uses the concept of references and citers of a bag for its classification. References of a bag (say bag B) are its nearest-neighbors and a citer is a bag which considers bag B among one of its neighbors. Table 1 shows the concept of references and citers of a bag. Two nearest-references of bag B_1 are B_2, B_4 while its two nearest-citers are B_4, B_2 . Assume that the query image is B_2 , then the nearest-neighbor rank of B_1 is 2 (since B_1 is the second nearest-neighbor of B_2) and the citer rank of B_1 is 1 (since B_2 is the nearest-neighbor of B_1). For a more formal definition of references and citers, readers are pointed to Wang and Zucker.⁵¹ MIL techniques, such as DD and EM-DD, are used for multiple-instance CBIR.^{43,45} EM-DD was used by Rahmani et al.⁴⁶ One-class SVM was also used.⁴⁴ All the aforementioned methods use a learning stage before starting the process of retrieval. We now describe the proposed unsupervised retrieval framework—MI-RankKNN—in detail.

We divide each image into 64 blocks on an 8×8 grid. CC features are extracted for each block separately. The SGF and FRST features are also extracted for these blocks. This gives us a 280-D (256-CC features + 24-statistics of SGF features) feature vector for each block. It should be noted that the number of blocks per image may vary as the number depends on the thresholding scheme. Recall that citation-KNN⁵¹ is one of the most popular KNN-based methods used to solve multiple-instance classification problems. It uses a modified Hausdorff distance. Hausdorff distance is used to calculate the distance between two subsets of metric space. In short, it gives a measure of dissimilarity between the two metric spaces. Since the Hausdorff distance is very sensitive to outliers, minimal Hausdorff distance was proposed⁵¹ which is defined as

$$H(A, B) = \min_{b \in B} \min_{a \in A} \|a - b\|, \quad (6)$$

where $H(A, B)$ denotes the Hausdorff distance between two bags (nonempty subsets of a metric space) A and B whereas a and b are the instances in bags A and B , respectively. When applied to DR images, this may not give the desired results. For example, in Fig. 2, only two instances (marked by red borders) are totally different from the instances in the normal image. Thus, the minimal Hausdorff distance between an NPDR and a normal image will almost be equal to the distance between two normal images. This calls for a new way of measuring the distance between images which will effectively capture small, localized lesions.

Consider two images X and Y with m and n blocks each. The minimum distance between features of i 'th block of X and all blocks of Y is given as

$$D_{(i)}(X, Y) = \min_{y \in Y} d(x_i, y_j) \forall j \in \{1, 2, \dots, n\}, \quad (7)$$

where $d(\cdot, \cdot)$ represents the Euclidean distance between features of two blocks. D is an m -dimensional vector, which records the distance between each block of X and its closest match in Y . The distances used in this article are Euclidean unless otherwise specified. We use the above equation to calculate the D vector for each block in the query image with every image in the database. We match each block of the query image to its closest image in the database. A simple sorting of all the matches for each block in the query image gives the list of best-matched blocks for each query block in the dataset and is called the similarity ranked list. An aggregated ranked list is created for each image in the database by averaging the similarity rank for each block in every image in the database. The sorted list of aggregated ranks of images gives the m -rank list which is to be treated as the final ranking. The algorithm is to obtain the m -rank in Algorithm 1.

The nearest-neighbor metric may not be sufficient to always generate optimum retrieval results for DR images.⁵¹ This problem is avoided by incorporating the citer rank to the framework of m -rank. The citer rank is calculated as mentioned in Sec. 3.2. This is accomplished by obtaining a final “meanRank,” which is the average of m -rank and citer rank of Q . The top k retrieved images are then based on the meanRank list.

Since the distances in MIRank-KNN are calculated on the block level, it is more sensitive to the presence of localized lesions. For example, the distance between an affected image that contains a small block of lesion and a normal image will

Table 1 Nearest-references and citers of four bags.

	$N = 1$	$N = 2$	$N = 3$
B_1	B_2	B_4	B_3
B_2	B_3	B_1	B_4
B_3	B_2	B_1	B_4
B_4	B_1	B_2	B_3

Note: The bold represents that the two nearest neighbors of B_1 are B_2 and B_4 . B_1 is a citer of B_2 and B_4 .

Algorithm 1 m -rank algorithm

1. Calculate D' between Q (query image) and every image in the database.
2. Sort D' along columns and its indices to give a similarity list. $SL(x, y) = z$ represents the y 'th best match for the x 'th block of Q and the best match is a block belonging to image z .
3. Calculate an aggregated ranked list of all images in the database by averaging the similarity rank for each block.
4. Sort aggregated ranked list to obtain the m -rank list.

produce a vector D that has one large value. This improves the overall retrieval ranking. By storing the distances between blocks of all the images in the databases (and thereby continuously updating that distance matrix), the process of retrieval can be made significantly faster.

4 Experimental Setup

The dataset used to evaluate the proposed set of features consists of 493 images, assembled from four well-known and publicly available databases including DIARETDB0,⁵² DIARETDB1,⁵³ STARE,⁵⁴ and Messidor (kindly provided by the Messidor program partners, see Ref. 55.) and an annotated dataset of 84 PDR images from Jaeb Center for Health Research. For reviewing purposes, we have posted the dataset and the source code used in this study (the implementation and the database are available in Ref. 9), which will be made public upon the decision of the paper. There are 164 normal images, 161 NPDR images, and 168 PDR images. Labeling of the first three datasets was unambiguous since the labels were well-defined and had three categories as desired. The Messidor dataset has three categories apart from the normal condition. They are defined by the amount of presence of MAs, hemorrhage, and NV. We classified the images from the Messidor dataset as NPDR if annotations indicated an absence of NV in them, and accordingly, we also obtained PDR images if their annotations indicated a strong presence of MAs and hemorrhages as well as NV. It should be noted that our database consists of DR images from five different sources. The retrieval results show that results are consistent over all images. Therefore, our approach is robust in handling variations in color and brightness of images as well as other conditions while capturing them. Due to the adaptive nature of our approach, it can be applied in various places to get effective results.

The proposed approach was evaluated against prior state-of-the-art vision methods that use Gabor features³² and semantic of neighborhood color moment histogram features.¹⁴ A short description for each approach is as follows:

1. Gabor feature-based image retrieval: Gabor features have been used for textured image retrieval.³² Gabor wavelet transform of an image is calculated over four scales and six orientations. Mean and standard deviation of the transform are calculated over all orientations and scales. This acts as the final feature vector.
2. HNM: HNM has been used for retrieval of gastroscopic images that are predominantly red in color.¹⁴ The image is first transformed into hue-saturation-value color space and then quantized. Low-order moments are known to express the color distribution well. Therefore, the first three central moments for each pixel in its 3×3 neighborhood are calculated. After operating on each pixel, three distinct histograms are calculated from the matrices of central moments. These histograms are concatenated to form the final feature vector.

These two approaches use k -nearest-neighbor-based retrieval systems. Distances providing the best retrieval results are selected. All 493 images were queried one-by-one and the top ($k =$) five images were retrieved using the approach. Popular evaluation metrics were adopted and used Refs. 56 and 57.

5 Results and Analysis

This section presents the result of the proposed multiple-instance retrieval approach on our DR image dataset. Our approach produces state-of-the-art results even when there is a wide variation in the lighting of images. There are two main parameters in our approach: number of bins in CC features and FRST radii. We show the effect of parameter tuning on the results. We compare our approach with two state-of-the-art image retrieval systems: textured image retrieval using Gabor features³² and medical image retrieval using HNM.¹⁴ The proposed approach consists of several important components, namely, CC features, FRST, SGF, and multiple-instance retrieval framework—MIRank-KNN. We analyze the effect of individual components by comparing them with other widely used features. We compare FRST with SIFT,⁵⁸ FAST,^{59,60} and Harris corner with color saliency boosting⁶¹ and show that FRST produces better results. We also compare CC features with transformed color histogram,³³ color moments,³³ and original AutoCC features.³¹ MIRank-KNN framework considers image blocks (instances) while retrieving images. To show its superiority, we replace it with the citation-KNN retrieval framework (explained later) that works on multiple-instance feature space but does not concentrate on all the blocks of an image simultaneously. Finally, to justify the choice of multiple-instance-based retrieval framework, we compute our features as well as a set of state-of-the-art local features on the entire image and compare the results.

5.1 Results of the Proposed Approach

In this section, we present statistical analysis of the proposed retrieval approach and compare it with the three aforementioned retrieval approaches. We present four evaluation metrics: (1) $\geq k$ hit-rate (HR), (2) mean accuracy at k 'th rank, (3) precision at k 'th ($P@k$) rank, and (4) mean average precision (MAP). $P@k$ and MAP metrics were proposed for document retrieval⁵⁷ but were also used effectively for image retrieval in Ref. 62. The $\geq k$ HR is defined as the percentage of images for which at least k relevant images were retrieved. Mean accuracy at k 'th rank denotes the percentage of relevant images retrieved at that particular rank. $P@k$ rank measures the relevance of top k images in the ranking result with respect to the query image. It is also the probability of finding a relevant image in the top k images, given as

$$P@k = \frac{\# \text{relevant images in top } k \text{ images}}{k}. \quad (8)$$

We average $P@k$ values for all the queries to get a single $P@k$ value. Average precision (AP) is defined as the average of $P@k$ values for all relevant queries

$$AP = \frac{\sum_{k=1}^N [P@k * \text{rel}(k)]}{\# \text{total relevant images for the query}}, \quad (9)$$

where N is the number of retrieved images and $\text{rel}(k)$ is an indicator function on the relevance of the n 'th image given as

$$\text{rel}(k) = \begin{cases} 1, & \text{if the } k' \text{th image is relevant} \\ 0, & \text{otherwise} \end{cases}. \quad (10)$$

MAP is obtained by averaging AP values for all the queries. It is clear that $P@n$ and MAP tend to emphasize the quality at

Fig. 9 Precision–recall curves for five methods when five images are retrieved.

higher ranks. But MAP is less affected by this since it depends on the entire list of retrieved images.

Five ROC curves are shown in Fig. 9, each one corresponding to one of the five retrieval approaches, namely, color name and SIFT on the entire image as well as blocks, Gabor features, HNM, and the proposed approach. Each curve was calculated as the average of precision–recall curves for normal, NPDR, and PDR images as follows. We retrieve five images and calculate the precision and recall values in response to each query image. To get precision and recall for the entire category, we average over all images of that category. To calculate precision at all recall values, we use the concept of “interpolated precision” (IP). IP p_{interp} at a recall level r is defined as the maximum precision obtained for any recall level $r' \geq r$. Mathematically

$$p_{\text{interp}}(r) = \max_{r' \geq r} p(r'). \quad (11)$$

We get IP at recall values ranging from 0 to 1 in steps of 0.2. For more details about precision–recall curve calculation, we refer the reader to Manning et al.⁶³

Mean accuracy at the k 'th rank and $P@k$ rank is given in Table 2. Table 3 provides values of MAP, overall retrieval accuracy, and the HRs at all ranks. Mean accuracy produces consistently higher values which suggests that the proposed approach retrieves clinically relevant images at all ranks. The ratio of relevant to irrelevant retrieved images is high in the proposed approach as compared to the other methods. The proposed approach produces higher precision at every rank and higher MAP value than other approaches. Therefore, the proposed approach has produced better results by retrieving clinically relevant images at all ranks instead of just at top ranks.

We conduct a visual inspection of the results produced by HNM. It shows that HNM did not quite retrieve images with clinically similar lesions, particularly images containing hemorrhages. HNM retrieved only those images that have similar brightness conditions as the query image. This reduced the clinical relevance and the generalization ability. Since HNM models the global distribution of patch-level statistics, it produces a higher performance than Gabor features. When compared to the results of the proposed approach, the reduction in performance can be attributed to the following factors:

Table 2 Mean accuracy and $P@k$ rank (in %).

	Acc@1/ $P@1$	Acc@2/ $P@2$	Acc@3/ $P@3$	Acc@4/ $P@4$	Acc@5/ $P@5$
Gabor	73.43/73.43	72.62/73.02	68.76/71.60	67.75/70.64	70.39/70.59
HNM	77.69/77.69	74.04/75.86	72.01/74.58	73.83/74.39	69.57/73.43
Proposed	84.38/84.38	81.54/82.96	83.37/83.10	79.51/82.20	80.93/81.95

Note: Best results are in bold.

Table 3 $\geq k$ HR (in %).

	≥ 1 HR	≥ 2 HR	≥ 3 HR	≥ 4 HR	≥ 5 HR	Mean accuracy	MAP
Gabor	93.91	86.61	76.06	60.24	36.10	70.59	79.68
HNM	94.52	87.22	77.28	64.50	43.61	73.43	82.49
Proposed	96.55	91.48	87.42	76.88	57.40	81.95	87.6

Note: Bold indicates the best results across all algorithms in this table.

1. HNM quantization scheme is not tuned to DR images.
2. HNM fails to model the global distribution of local spatial correlation of colors unlike CC features.
3. HNM feature is calculated on an image level, whereas CC features in the proposed approach are calculated on a block level. This makes CC features more “descriptive.”

We perform two experiments to show (1) the compatibility and reproducibility of our approach and (2) adaptability to various databases with minor updates. In the first experiment, we perform 20 iterations, and on each iteration, a random subset of 95% of total images is selected and all the images are queried one-by-one. In the second experiment, the quantization scheme is designed by using a randomly selected database of 95% of the images, which ensures that algorithm performance does not heavily degrade by adding a few images. In both experiments, accuracy and HRs remained consistent as desired.

A mean confusion matrix is created to better understand the results of our approach. It is shown in Table 4. The following example illustrates the process of constructing a confusion matrix. Suppose all images are queried one-by-one resulting in n retrieved images. The first row of the confusion matrix shows that 81.59% of the n images were normal, when a normal image was queried. Similarly, 13.78% images were NPDR and 4.63% images obtained were PDR, while querying a normal image. The second and third rows can be similarly explained. The rate of retrieving a normal image in response to a queried PDR image (and vice versa) is less. This is exactly in accordance with the requirements of the algorithm. Retrieving a normal image for a PDR image can be quite harmful. Similarly, a retrieved PDR image in response to a normal query image might mislead the ophthalmologist.

Images retrieved by our approach are shown in Fig. 10. The top three rows contain successful retrievals for normal, NPDR, and PDR images, respectively. In spite of varying illumination and exposure, our approach yields perfect results. However, there are some cases when our approach fails. In the fourth row of Fig. 10, the query image is NPDR where the lesions span only a few pixels. Our approach retrieves a normal image on the fourth rank. There are subtle differences in the retinal blood vessels of the query NPDR and the retrieved normal image, which our method fails to capture. In the fifth row, a PDR image is retrieved in response to a normal image. In that PDR image, no lesions are clearly visible and even the optic disc is hardly visible, which makes the retrieval difficult. In the last row, a normal image is retrieved when a PDR image is queried. The normal image is extremely red, and the retinal blood vessels have a different appearance as compared to the other normal images. The goal of perfect retrieval on a variety of fundus images is still far from achieved, but we think our results are a promising step in that direction.

Table 4 Mean confusion matrix (in %).

	Normal	NPDR	PDR
Normal	81.59	13.78	4.63
NPDR	9.81	80.50	9.69
PDR	8.45	7.86	83.69

Our approach is robust to exposure and lighting changes in an image to some extent. Today’s cameras are high quality and usually do not produce noisy or bad exposure images. In rare cases, clinicians generally prefer to retake the picture so that their diagnosis is not affected. However, we acknowledge that small changes in exposure can happen and may be acceptable to an ophthalmologist as long as a correct diagnosis can be made. In the literature, many approaches perform histogram equalization to try to correct effects of bad exposure. Our spectral tuning method is better since it adapts to the color spectrum of retinal images. The robustness of our system can be assessed by the fact that we get 87.6% MAP on a dataset of varying illumination and exposure retinal images.

5.2 Effect of Varying Illumination

Through our interactions with ophthalmologists, we have observed that if the quality of the retinal image captured is unacceptable, then another photo is captured. Thus, it is reasonable to assume that the photos will always be of high quality. However, photos captured in different labs or settings can have different contrasts and illuminations. We show that our algorithm can handle these changes well through an experiment described as follows. We change (i.e., increase or decrease, which is randomly chosen) the brightness of each image by a random amount, ranging anywhere from 15% to 25% of the original brightness. We change the brightness by mapping the values of the “V” channel of an image to a new intensity curve. We also introduce a nonlinear effect by controlling the shape of the new intensity curve using γ . The value of γ is also randomly chosen from the interval, [0.6,1.4]. We pick an image from the original set of images and retrieve from the other set, having modified values of brightness. We repeat this for each image and calculate the same metrics as done previously. It is to be noted that we do not spectrally tune our CC feature using the modified images, which is needed for a fair assessment of the effect of varying brightness on the performance. Table 5 shows the results of both the experiments. It is interesting to see that the images with modified intensity values get higher results. Since our original database has a lot of intensity variation, it is possible that the images which were too dark actually benefited from this experiment and hence the performance improvement. The results of both the experiments show that our approach can handle intensity variations with a reasonable tolerance.

5.3 Effect of Different Feature Combinations in the Proposed Approach

We analyze the contribution of color features and texture features in the proposed approach. Images are retrieved using two separate sets of features: (1) CC features and (2) FRST + SGF features. By using texture features in set 2 alone, we get 55.25% accuracy whereas by using only color features of set 1, 77.32% can be obtained. The features are indeed complementary since the results improve considerably after using both feature sets.

5.4 Effect of Parameter Tuning

The proposed approach has two main parameters: (1) number of bins in CC features and (2) set of radii in FRST. We analyze the effects of those parameters on the retrieval accuracy.¹

Fig. 10 Retrieved images using the proposed approach. Each row contains a query image (leftmost) and five retrieved images. A retrieved image belongs to the same category as the query image unless a red bar over it. Query images and the corresponding retrieved images in top three rows belong to normal, NPDR, and PDR category, respectively. In the fourth row, query image belongs to NPDR, fourth retrieved image is normal. Fifth row contains a normal query image, fifth retrieved image has PDR. In the sixth row, the query image has PDR, second retrieved image is normal. Please view in color.

1. Number of bins in CC features: The dimensionality of CC features is a function of the number of bins— $O(m^2)$. We choose to quantize each image into 16 bins. However, we show that the retrieval performance does not vary by a large amount with respect to number of bins.
2. FRST radii: As mentioned previously, the values of FRST radii should be small enough to capture the smallest of lesions. A set containing small, arbitrary values of radii produces good results. As the values of radii increase, the retrieval performance drops. If the

lesion size is known beforehand, values of radii can be set accordingly. We use three radii {1, 3, 5} in our implementation.

The results of parameter tuning are given in Table 6.

5.5 Comparison with Other State-of-the-Art Interest Point Detectors

FRST is a crucial component of our system. Its main advantages are speed and quality of detected interest points. By adjusting its parameters, it is possible to detect even the smallest of lesions.

Table 5 $\geq k$ HR (in %) with images of varying intensity.

	≥ 1 HR	≥ 2 HR	≥ 3 HR	≥ 4 HR	≥ 5 HR	Mean accuracy	MAP
No variation	96.55	91.48	87.42	76.88	57.40	81.95	87.6
Linear variation	97.36	94.73	90.47	82.76	64.71	86.00	90.85
Nonlinear variation	99.19	95.74	90.67	82.56	63.69	86.37	92.32

Table 6 Effect of parameter tuning on retrieval accuracy.

Parameters		Overall accuracy	MAP
FRST radii	{1, 3, 5}	81.95	87.60
	{5,7,9}	80.57	87.17
	{10,15,20}	78.34	86.33
Number of bins in CC	8	79.95	86.08
	32	80.16	87.78
	64	80.12	88.02

Interest point detectors are bound to produce false detections. The SGF stage acts as a complementary stage to FRST. We compare three other state-of-the-art interest point detectors, namely, SIFT,⁵⁸ FAST,^{59,60} and Harris corner with color saliency boosting (implemented from publicly available codes at Refs. 60, 64, and 65).⁶¹ Results in Table 7 show the superiority of FRST.

5.6 Comparison with Other State-of-the-Art Color Features

Color is a distinguishing characteristic for DR images, and hence spectrally tuned CC features play a major role in the retrieval process. The CC features are invariant to lighting color changes. We compare them with three other color features that are also invariant to lighting color changes. The three color features are: (1) transformed color histogram,³³ (2) color moments,¹⁴ and (3) original AutoCC features.³¹

5.6.1 Transformed color histogram

A transformed color space is created as shown in Eq. (1). Histograms of individual color channels are then concatenated to obtain the final feature vector. This feature, though invariant to light color changes, does not encode local spatial correlation of colors.

5.6.2 Color moments

Image is quantized by using the proposed quantization approach. Then, the color moment feature vector is calculated in an identical manner to HNM. Since the proposed quantization scheme is used, the pros and cons of color moments can be analyzed.

5.6.3 Original AutoCC features

Original AutoCC features were used for CBIR. We use their human-vision-based quantization scheme and calculate the

AutoCC features for all possible color pairs (i, i) , where $i \in [1, \dots, 64]$. We get a 64-D feature vector for each image. From the results in Table 7, the ineffectiveness of this quantization scheme in case of DR images is observed.

5.7 Effect of MIRank-KNN on Retrieval Performance

MIRank-KNN considers features of all the image blocks simultaneously in the retrieval process. The existing multiple-instance retrieval algorithms^{43–46} for natural images and DR-CBIR algorithms^{8,16,17} require training and some of them even require the user to specify the region of interest. On the other hand, our algorithm does not require labeled data for training, and the entire approach is automated. The labeled data are only used for calculating retrieval accuracy and other metrics. We show the superiority of MIRank-KNN through a threefold experiment.

First, we replace MIRank-KNN with a conventional k -nearest-neighbor approach while keeping the same feature space. We calculate the CC features for the entire image instead for individual blocks. Averaging of the feature vectors of the interest points given by FRST is done for the entire image instead on a block level to produce the final features. As expected, this fails to encode characteristics of small lesions and produces poor results. The results are given in Table 7 (last row).

Second, we introduce another baseline framework—citation-KNN retrieval—based on the lazy-learning framework.⁵¹ In the citation-KNN retrieval framework, the image is first divided into 64 blocks as done previously. The same feature set is calculated for each block. Consider two images I_1 and I_2 consisting of n_1 and n_2 blocks, respectively. We measure the similarity between these two images by calculating the minimal Hausdorff distance between them as shown in Eq. (6). This distance measure was also used by Wang and Zucker.⁵¹ Though this distance can work with multiple blocks, it cannot capture the appearances of localized lesions in its distance computation. For example, minimal Hausdorff distance between a normal image and an MA image, shown in Fig. 2, will be very small. Using maximal Hausdorff distance is also a bad option, since it is sensitive to outliers. On the other hand, MIRank-KNN will output a large distance for at least two blocks in its distance vector, thus not recognizing the other image as a clinically relevant image. Table 7 contains the results of citation-KNN retrieval framework.

Finally, we show the importance of multiple-instance framework. We show that simply using state-of-the-art local features are not enough to characterize localized lesions. We do so in two parts as follows:

1. We first assess the performance of local features alone, without involving a multiple-instance framework. We use SIFT and color name descriptors which are excellent local shape and color features, respectively.^{29,66}

Table 7 Analysis and comparison of the proposed approach with other state-of-the-art methods.

Method	Parameters	Overall accuracy	MAP
Proposed approach	FRST radii = {1, 3, 5} and number of bins = 16	81.95	87.6
FRST replaced with other interest point detectors			
SIFT	Peak threshold = 1	79.07	86.27
Harris corner with boosted color saliency	Top 150 interest points	77.04	84.29
FAST	Threshold = 7.5	80.08	86.79
CC replaced with other color features			
Transformed color histogram	Number of bins = 32	74.85	83.15
	Number of bins = 64	73.51	80.98
	Number of bins = 128	68.36	78.00
Color moments	Number of bins = 16	75.74	83.31
	Number of bins = 24	75.45	83.46
	Number of bins = 32	76.51	83.15
Original AutoCC features	Number of bins = 64	53.79	68.99
MIRank-KNN replaced with citation-KNN retrieval framework			
Citation-KNN retrieval	FRST radii = {1, 3, 5} and number of bins = 16	65.60	74.38
	FRST radii = {5, 7, 9} and number of bins = 16	65.23	73.81
	FRST radii = {10, 15, 20} and number of bins = 16	65.19	74.22
	FRST radii = {1, 3, 5} and number of bins = 8	67.99	76.00
	FRST radii = {1, 3, 5} and number of bins = 24	66.41	75
	FRST radii = {1, 3, 5} and number of bins = 32	69.70	79.27
Results without the use of multiple-instance framework			
Whole image features	FRST radii = {1, 3, 5} and number of bins = 16	58.83	72.02

Note: Bold indicates highest results.

Color name descriptor has been recently shown to be effective for image retrieval.⁶⁷ We follow a similar procedure for feature extraction. We extract SIFT keypoints and their descriptors to characterize the shape of lesions. Color name descriptors are extracted around SIFT keypoints and are appended to the SIFT descriptors. We then use nearest-neighbor retrieval.

2. We now show the effectiveness of multiple-instance framework. We divide the image into 64 patches (8×8 grid). We extract the same features as described above for each patch. MIRank-KNN is used for retrieval.

The results in Table 8 show that the combination of local features and MIRank-KNN produces a better performance

than local features alone. We also show three visual examples in Fig. 11. It shows that using local features on the entire image can yield incorrect results when the lesions are small and localized. Thus, normal images may be retrieved in response to affected (NPDR or PDR) images and vice versa.

Table 8 Performance of local features.

Method	Overall accuracy	MAP
Color name + SIFT + KNN on the entire image	75.29	83.06
Color name + SIFT + MIRank-KNN on the image blocks	78.86	85.87

Fig. 11 Left column shows query image. Middle and right columns show retrieved images by using local features with k -nearest-neighbor retrieval and local features with MIRank-KNN retrieval, respectively. In the top two rows, left and right images are PDR whereas the middle image is normal. In the last row, left and right images are normal and the middle one is PDR.

6 Conclusion

This paper presents an approach for retrieving clinically relevant DR images. The approach consists of a feature space that is spectrally tuned to the DR spectrum. Feature space makes optimal utilization of the quantization scheme and thus produces a better representation of the image even with a fewer number of bins. It makes sure that shades in DR images are almost uniformly spread across all the quantization bins, thereby creating a feature space with much higher entropy. Our approach also consists of a multiclass multiple-instance retrieval framework called MIRank-KNN, which uses minimal Hausdorff distance. The results using the proposed approach are reported and compared with other state-of-the-art retrieval frameworks in the literature. The ability of a multiple-instance framework to capture localized lesions was also analyzed over state-of-the-art local features. It was found that in all of the cases, our multiple-instance retrieval framework provides a boost to the results. The results based on our DR image dataset suggest that the proposed method can give good performance on different datasets and is robust against varying illuminations and exposures. It is also invariant to small additions or removal of data.

Disclosures

No conflicts of interest, financial or otherwise, are declared by the authors.

Acknowledgments

The work was supported in part by a grant from the Agency for Healthcare Research and Quality (Grant No: 1R21HS019792-01A1). Any opinions and findings expressed in this material are those of the authors and do not necessarily reflect the view of the Agency for Healthcare Research and Quality.

References

- Centers for Disease Control and Prevention and Others, "National diabetes fact sheet: national estimates and general information on diabetes and prediabetes in the United States, 2011," US Department of Health and Human Services, Centers for Disease Control and Prevention, Atlanta (2011).
- L. Guariguata et al., "Global estimates of diabetes prevalence for 2013 and projections for 2035," *Diabetes Res. Clin. Pract.* **103**(2), 137–149 (2014).

3. M. D. Abràmoff et al., "Evaluation of a system for automatic detection of diabetic retinopathy from color fundus photographs in a large population of patients with diabetes," *Diabetes Care* **31**(2), 193–198 (2008).
4. S. R. Salomão, M. R. Mitsuhiro, and R. Belfort Jr., "Visual impairment and blindness: an overview of prevalence and causes in Brazil," *An. Acad. Bras. Cienc.* **81**(3), 539–549 (2009).
5. S. Garg and R. M. Davis, "Diabetic retinopathy screening update," *Clin. Diabetes* **27**(4), 140–145 (2009).
6. X. Zhang et al., "Prevalence of diabetic retinopathy in the United States, 2005–2008," *J. Am. Med. Assoc.* **304**(6), 649–656 (2010).
7. B. Li and H. K. Li, "Automated analysis of diabetic retinopathy images: principles, recent developments, and emerging trends," *Curr. Diabetes Rep.* **13**, 453–459 (2013).
8. G. Quéllec et al., "Automated assessment of diabetic retinopathy severity using content-based image retrieval in multimodal fundus photographs," *Invest. Ophthalmol. Visual Sci.* **52**(11), 8342–8348 (2011).
9. P. S. Chandakkar, R. Venkatesan, and B. Li, "Accompanying code for MIRank-KNN: multiple-instance retrieval of clinically relevant diabetic retinopathy images," <http://www.public.asu.edu/~bli24/DR-System-and-Data.html>.
10. W. Cai, D. Feng, and R. Fulton, "Content-based retrieval of dynamic PET functional images," *IEEE Trans. Inf. Technol. Biomed.* **4**(2), 152–158 (2000).
11. E. Chaum et al., "Automated diagnosis of retinopathy by content-based image retrieval," *Retina* **28**(10), 1463–1477 (2008).
12. A. Gupta et al., "Content-based retrieval of ophthalmological images," in *Proc. Int. Conf. on Image Processing*, Vol. 3, pp. 703–706, IEEE (1996).
13. P. S. Chandakkar, R. Venkatesan, and B. Li, "Retrieving clinically relevant diabetic retinopathy images using a multi-class multiple-instance framework," *Proc. SPIE* **8670**, 86700Q (2013).
14. Q. Chen et al., "Medical image retrieval based on semantic of neighborhood color moment histogram," in *2nd Int. Conf. on Bioinformatics and Biomedical Engineering (ICBBE 2008)*, pp. 2221–2224, IEEE (2008).
15. K. S. Deepak, G. D. Joshi, and J. Sivaswamy, "Content-based retrieval of retinal images for maculopathy," in *Proc. of the 1st ACM Int. Health Informatics Symp.*, pp. 135–143 (2010).
16. G. Quéllec et al., "Fast wavelet-based image characterization for highly adaptive image retrieval," *IEEE Trans. Image Process.* **21**(4), 1613–1623 (2012).
17. G. Quéllec et al., "Wavelet optimization for content-based image retrieval in medical databases," *Med. Image Anal.* **14**(2), 227–241 (2010).
18. W. W. Chu, I. T. Leong, and R. K. Taira, "A semantic modeling approach for image retrieval by content," *Int. J. Very Large Data Bases* **3**(4), 445–477 (1994).
19. P. M. Kelly, T. M. Cannon, and D. R. Hush, "Query by image example: the comparison algorithm for navigating digital image databases (CANDID) approach," *Proc. SPIE* **2420**, 238 (1995).
20. P. Korn et al., "Fast and effective retrieval of medical tumor shapes," *IEEE Trans. Knowl. Data Eng.* **10**(6), 889–904 (1998).
21. M. Lamard et al., "Content based image retrieval based on wavelet transform coefficients distribution," in *29th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBS 2007)*, pp. 4532–4535, IEEE (2007).
22. M. H. Goldbaum et al., "Image understanding for automated retinal diagnosis," in *Annual Symp. on Computer Application [sic] in Medical Care*, pp. 756–760, American Medical Informatics Association (1989).
23. R. Pires et al., "Assessing the need for referral in automatic diabetic retinopathy detection," *IEEE Trans. Biomed. Eng.* **60**, 3391–3398 (2013).
24. A. Rocha et al., "Points of interest and visual dictionaries for automatic retinal lesion detection," *IEEE Trans. Biomed. Eng.* **59**, 2244–2253 (2012).
25. G. Quéllec et al., "A multiple-instance learning framework for diabetic retinopathy screening," *Med. Image Anal.* **16**(6), 1228–1240 (2012).
26. C. I. Sánchez et al., "Active learning for an efficient training strategy of computer-aided diagnosis systems: application to diabetic retinopathy screening," in *Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI)* (2010).
27. X. Xu and B. Li, "Automatic classification and detection of clinically relevant images for diabetic retinopathy," *Proc. SPIE* **6915**, 69150Q (2008).
28. R. Venkatesan et al., "Classification of diabetic retinopathy images using multi-class multiple-instance learning based on color correlogram features," in *IEEE Int. Conf. of Engineering in Medicine and Biology Society*, pp. 1462–1465 (2012).
29. T. Deselaers, D. Keysers, and H. Ney, "Features for image retrieval: an experimental comparison," *Inf. Retr.* **11**(2), 77–107 (2008).
30. J. Huang et al., "Image indexing using color correlograms," in *Proc. of IEEE Computer Society Conf. on Computer Vision and Pattern Recognition* (1997).
31. M. Li, "Texture moment for content-based image retrieval," in *IEEE Proc. Int. Conf. on Multimedia and Expo*, pp. 508–511, IEEE (2007).
32. B. S. Manjunath and W.-Y. Ma, "Texture features for browsing and retrieval of image data," *IEEE Trans. Pattern Anal. Mach. Intell.* **18**(8), 837–842 (1996).
33. K. E. Van De Sande, T. Gevers, and C. G. Snoek, "Evaluating color descriptors for object and scene recognition," *IEEE Trans. Pattern Anal. Mach. Intell.* **32**(9), 1582–1596 (2010).
34. X. Liu, C.-T. Lu, and F. Chen, "An entropy-based method for assessing the number of spatial outliers," in *IEEE Int. Conf. on Information Reuse and Integration*, pp. 244–249, IEEE (2008).
35. H. Heikinheimo et al., "Finding low-entropy sets and trees from binary data," in *Proc. of the 13th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, pp. 350–359, ACM (2007).
36. S. Y. Park and A. K. Bera, "Maximum entropy autoregressive conditional heteroskedasticity model," *J. Econometrics* **150**(2), 219–230 (2009).
37. W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," *IEEE Trans. Pattern Anal. Mach. Intell.* **13**(9), 891–906 (1991).
38. G. Loy and A. Zelinsky, "Fast radial symmetry for detecting points of interest," *IEEE Trans. Pattern Anal. Mach. Intell.* **25**(8), 959–973 (2003).
39. P. J. Locher and C. Nodine, "Symmetry catches the eye," in *Eye Movements: From Physiology to Cognition*, pp. 353–361 (1987).
40. W. Richards and L. Kaufman, "Center-of-gravity tendencies for fixations and flow patterns," *Percept. Psychophys.* **5**(2), 81–84 (1969).
41. L. Kaufman and W. Richards, "Spontaneous fixation tendencies for visual forms," *Percept. Psychophys.* **5**(2), 85–88 (1969).
42. P. D. Kovesi, "MATLAB and octave functions for computer vision and image processing," Centre for Exploration Targeting, School of Earth and Environment, University of Western Australia, <http://www.csse.uwa.edu.au/pk/research/matlabfns/> (2000).
43. Q. Zhang et al., "Content-based image retrieval using multiple-instance learning," in *Proc. of the Nineteenth Int. Conf. on Machine Learning*, Vol. 2, pp. 682–689 (2002).
44. C. Zhang et al., "A multiple instance learning approach for content based image retrieval using one-class support vector machine," in *IEEE Int. Conf. on Multimedia and Expo*, pp. 1142–1145 (2005).
45. C. Yang and T. Lozano-Perez, "Image database retrieval with multiple-instance learning techniques," in *Proc. 16th Int. Conf. on Data Engineering*, pp. 233–243, IEEE (2000).
46. R. Rahmani et al., "Localized content based image retrieval," in *Proc. of the 7th ACM SIGMM Int. Workshop on Multimedia Information Retrieval*, pp. 227–236, ACM (2005).
47. T. G. Dietterich, R. H. Lathrop, and T. Lozano-Pérez, "Solving the multiple instance problem with axis-parallel rectangles," *Artif. Intell.* **89**(1), 31–71 (1997).
48. O. Maron and T. Lozano-Pérez, "A framework for multiple-instance learning," in *Advances in Neural Information Processing Systems*, pp. 570–576, MIT Press (1998).
49. Q. Zhang and S. A. Goldman, "EM-DD: an improved multiple-instance learning technique," in *Advances in Neural Information Processing Systems*, pp. 1073–1080 (2001).
50. S. Andrews, I. Tsochantaridis, and T. Hofmann, "Support vector machines for multiple-instance learning," in *Advances in Neural Information Processing Systems*, pp. 561–568 (2002).
51. J. Wang and J.-D. Zucker, "Solving multiple-instance problem: a lazy learning approach" in *Proc. of the Seventeenth Int. Conf. on Machine Learning*, pp. 1119–1126 (2000).
52. T. Kauppi et al., "DIARETDB0: evaluation database and methodology for diabetic retinopathy algorithms," in *Machine Vision and Pattern Recognition Research Group*, Lappeenranta University of Technology, Finland (2006).

53. T. Kauppi et al., "The DIARETDB1 diabetic retinopathy database and evaluation protocol," in *Proc. of the British Machine Vision Conf.*, pp. 1–10 (2007).
54. B. McCormick and M. Goldbaum, "STARE: Structured Analysis of the Retina: image processing of TV fundus image," in *USA–Japan Workshop on Image Processing*, Jet Propulsion Laboratory, Pasadena (1975).
55. E. Decencière et al., "Feedback on a publicly distributed database: the Messidor database," *Image Anal. Stereol.* **33**(3), 231–234 (2014).
56. B. Sigurbjörnsson and R. Van Zwol, "Flickr tag recommendation based on collective knowledge," in *Proc. of the 17th Int. Conf. on World Wide Web*, pp. 327–336, ACM (2008).
57. T.-Y. Liu et al., "LETOR: benchmark dataset for research on learning to rank for information retrieval," in *Proc. of SIGIR Workshop on Learning to Rank for Information Retrieval*, pp. 3–10 (2007).
58. D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision* **60**(2), 91–110 (2004).
59. E. Rosten and T. Drummond, "Fusing points and lines for high performance tracking," in *IEEE Int. Conf. on Computer Vision*, Vol. **2**, pp. 1508–1511 (2005).
60. E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *European Conf. on Computer Vision*, Vol. **1**, pp. 430–443 (2006).
61. J. Van De Weijer, T. Gevers, and A. D. Bagdanov, "Boosting color saliency in image feature detection," *IEEE Trans. Pattern Anal. Mach. Intell.* **28**(1), 150–156 (2006).
62. F. F. Faria et al., "Learning to rank for content-based image retrieval," in *Proc. of the Int. Conf. on Multimedia Information Retrieval*, pp. 285–294, ACM (2010).
63. C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*, Vol. **1**, Cambridge University Press, Cambridge, U. K. (2008).
64. A. Vedaldi and B. Fulkerson, "VLFeat: an open and portable library of computer vision algorithms," <http://www.vlfeat.org/> (2008).
65. J. Van De Weijer, T. Gevers, and A. Gijssenij, "Accompanying code for edge-based color constancy," *IEEE Trans. Image Process.* **16**(9), 2207–2214 (2007).
66. F. S. Khan et al., "Color attributes for object detection," in *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 3306–3313, IEEE (2012).
67. L. Zheng et al., "Packing and padding: coupled multi-index for accurate image retrieval," in *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 1947–1954, IEEE (2014).

Parag Shridhar Chandakkar is currently pursuing his PhD in computer science from Arizona State University (ASU) advised by Dr. Baoxin Li. He received his master's degree in electrical engineering from ASU in 2012. His research interests lie in the field of computer vision, specifically computational image aesthetics, and enhancement and no-reference image-quality estimation. During his PhD, he has interned with Intel and IBM T. J. Watson Research, where he worked on advanced driver-assistance systems and distributed deep learning systems.

Ragav Venkatesan is currently a PhD student at ASU. He is a research associate with the Visual Representation and Processing Group at ASU. Prior to this, he was a research assistant in the Image Processing and Applications Lab, School of Electrical and Computer Engineering, ASU. From 2013 to 2014, he was with Intel Corp. He was also a coauthor of the book *Convolutional Neural Networks in Visual Computing*.

Baoxin Li is a professor of computer science and engineering at ASU. He received his PhD from the University of Maryland in 2000. From 2000 to 2004, he was a senior researcher with SHARP Labs. He was an adjunct professor with Portland State University from 2003 to 2004. He won twice SHARP Labs' president awards and once inventor of the year award. He is a recipient of the National Science Foundation CAREER award.