Evaluation of Gist Operator for Document Image Retrieval

Fahimeh Alaei
School of ICT, Griffith
University, Australia
fahimeh.alaei@griffithuni.edu.au

Alireza Alaei
School of ICT, Griffith
University, Australia
alireza20alaei@yahoo.com

Umapada Pal
CVPR Unit, Indian
Statistical Institute, India
umapada@isical.ac.in

Michael Blumenstein
University of Technology
Sydney, Australia
Michael.Blumenstein@uts.edu.au

Abstract—**As digitised documents normally contain a large variety of structures, a page segmentation- and layout-free method for document image retrieval is preferable. In this research work, therefore, wavelet transform as a transform-based approach is initially used to provide different under-sampled images from the original image. Then, Gist operator, as a feature extraction technique, is employed to extract a set of global features from the original image as well as the sub-images obtained from the wavelet transform. Moreover, the column-wise variances of the values in each sub-image are computed and they are then concatenated to obtain another set of features. Considering each feature set, locality-sensitive hashing is employed to compute similarity distances between a query and the document images in the database. Finally, a classifier fusion technique using the mean function is taken into account to provide a document image retrieval result. The combination of these features and a clustering score fusion strategy provides higher document image retrieval accuracy. Two different databases of the document image are considered for experimentation. The results obtained from the experimental study are detailed and the results are encouraging.**

Keywords: document image retrieval; wavelet transform; Gist operator; locality-sensitive hashing

I. INTRODUCTION

With the evolution of technology, a huge volume of document images has been produced by different devices. The volume of document images is progressively growing with the availability of new technology and devices. Nowadays, most digital documents including journal/conference papers, theses, books, and reports, are published only in image format. Although document image processing technologies, such as Optical Character Recognition (OCR), can be used to convert documents images automatically to a machine-readable text format suitable for recognition-based document image retrieval systems, it may not always be a practical and cost-effective solution for dealing such a huge volume of document images. Reasons can be mentioned include: i) available layout analysis and segmentation techniques may not provide satisfactory results on document images with complicated layouts; and ii) OCR technology cannot provide highly accurate recognition results, especially when it is applied on degraded and poor quality document images [1].

As a result, various recognition-free document image retrieval methods have been proposed in the literature [2-4]. There are document image retrieval methods based on handwritten signature, logo, trademark and shape coding [5-7]. The use of layout-specific structures, content-specific features [8, 9], key-region detection, and bag-of-words features for document image retrieval process have also been reported in the literature [4].

Texture-based features and their effectiveness for classification and segmentation of historical document images have recently been explored in [10, 11]. However, the previous studies of document image retrieval have not paid much attention to texture features for document image retrieval [2, 3]. Moreover, there is a high demand for an automatic technique to extract features similar to human perception that leads to an efficient retrieval system [12].

Gist descriptor [13, 14] and wavelet transform [16], as two appearance-based texture features, have traditionally been used in the field of computer vision and image processing for natural scene image analysis. In [12], Gist features were used for detection and classification of the target in satellite images. The image was divided into small images, and two sets of the feature as saliency and Gist were extracted from the images for their research objective. In [15], angular radial partitioning Gist descriptor was used for scene classification. To capture the finer details of scene image, the effectiveness of angular partitioning was considered. Multi-scale Gist features were used to represent the structural information of the buildings in the image for building recognition [16]. The wavelet transform as a texture-based feature has been used for different applications of signal and image analysis [17]. In [18], wavelet transform was used for off-line signature verification. The global influence of ink-type was estimated using single-level 2-D wavelet decomposition method. In [19], wavelet transform was used for decomposing the image and generated sub-images to form a new time series. Then the energy vector was extracted by time-frequency domain analysis for their research work. In [20], wavelet transform was also used for generating sub-images and then, the variance of each sub-image was extracted for the retrieval process.

Since our interest is in retrieving similar document images to a given query, we propose to employ Gist descriptor [12] as the basis of an appearance-based feature for document image retrieval in this research. Therefore, a combination of two different texture-based features, Gist and wavelet, is considered to propose a recognition-free approach for retrieving document images. Gist refers to the significant information that a viewer can detect from a glance at an image [12]. The property of the wavelet transform is also its ability to decompose the image into orthogonal components, so-called image details.

The present research work explores for the first time the effects of Gist descriptor for document image retrieval. In the proposed system, wavelet transform is initially employed on the original image to provide different under-sampled images.

Then, Gist operator is employed to extract a set of global features from the original image as well as the sub-images obtained from the wavelet transform. Moreover, the column-wise variances of the values in each sub-image are computed and they are then concatenated to obtain another set of features. Locality-sensitive hashing is employed to compute similarity distances between a given query and document images in the database, using each feature set. In the last step, a classifier fusion technique using the mean function is taken into account to provide document image retrieval results.

The rest of the paper is organized as follows. In Section II, the proposed document image retrieval approach is explained. Section III discusses experimental results and comparative analysis. Finally, conclusion is presented in Section IV.
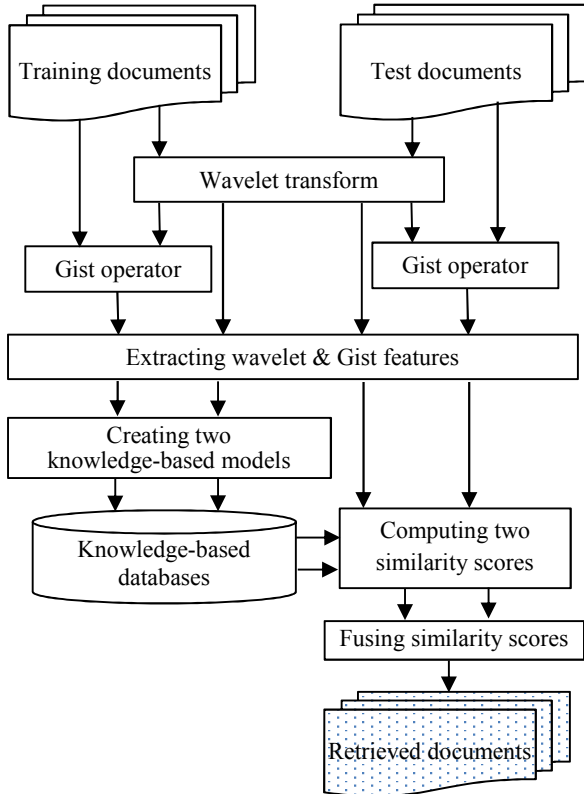

Fig. 1. Block diagram of the proposed method

II. PROPOSED METHOD

A block diagram of the proposed document image retrieval method is demonstrated in Fig. 1. In the training phase of the proposed method, the wavelet transform is applied on the document image to decompose the original document image into under-sampled document images [20, 21]. The three detailed generated document images, horizontal, vertical, and diagonal, are taken into account, and the approximation image is ignored. Gist operator is employed once on the original image as well as on the three generated under-sampled images. The Gist features extracted from the original image and under-sampled images are being concatenated to build the Gist features [13, 15, 16] for the image. The column-wise variance

of the values in each sub-image are also computed, and these are concatenated to obtain wavelet features. These feature vectors are kept separately as two knowledge-based databases. In the testing phase of the proposed method, the same feature extraction techniques used in the training phase are employed to characterise a given query image. Two similarity distances between the extracted wavelet and Gist features from the query image and the knowledge-based databases are then computed. A weighted average fusion strategy is finally applied on the obtained similarity scores, and image(s) with maximum visual similarity to the query image are retrieved accordingly. Details of each step are further illustrated in the following subsections.

A. Feature Extraction

Gist descriptor [13, 14] and wavelet transform [16], as two appearance-based texture features, were commonly applied on natural scene images in the field of computer vision and image processing. Both the Gist and wavelet transform can provide different appearance-based features, such as the particular information that a human can detect from an image. Since we are also interested in some appearance-based features for recognition free document image retrieval, two different feature extraction methods are proposed in this research work.

1. Wavelet Transform

Wavelet transform has been applied in the different fields of machine learning and image processing in the literature [18, 20, 22]. The attribute of the wavelet transform is to represent the image in multi-resolution forms, which provides more information compared to the original image. There are three different operators to generate multi-resolution of the image, sub-band coding, image pyramids, and the Haar transform. In this research work, the Haar transform [21], as the simplest orthonormal wavelet, is chosen [23].

A wavelet $\psi_{us}(x)$ is a function where $L^2(R)$ satisfy $\int |f(x)|^2 \, dx < \infty$, and a function produced by two parameters as u and s. The basic function is called mother wavelet $\psi(x)$, which translated by parameter u and scaled by s can be represented as:

$$\psi_{us}(x) = \frac{1}{\sqrt{s}} \psi\left(\frac{x-u}{s}\right) \qquad (1)$$

The discrete wavelet transform (DWT) coefficients are used to decompose the image into under-sampled images. In other words, transforming an image from the spatial domain into frequency domain DWT is considered. However, continuous wavelet transform $W_\psi(u, s)$ is not supported in practical application.

$$W_\psi(u, s) = \int_{-\infty}^{+\infty} f(x) \, \psi_{u,s}(x) dx \qquad (2)$$

$$u = (k - l), s = (k - 1)2, k = 1, 2,$$

where l and k are identified indices denoted by ψ_{lk} in these wavelets. To compute the wavelet coefficients of a discrete signal all integrals are replaced by sums. The s and u parameters are used to discretise the wavelet transform. Equation (1) remains integer, for the argument of the scaled

and transformed wavelet, and, an arbitrary signal $f(x)$ of finite energy using an orthonormal wavelet can be written as:

$$f(x) = \sum_k \sum_l d_l^k \psi_l^k(x) \qquad (3)$$

$$d_l^k = \int_{-\infty}^{\infty} f(x) \psi_l^k(x) dx \qquad (4)$$

The two-dimensional DWT is computed by applying a filter bank to the image [24]. The wavelet transform computation of a two-dimensional image is also a multi-resolution approach, which applies recursive filtering and sub-sampling. By considering DWT, the given image at each scale is decomposed into four frequency sub-bands. An approximation coefficients matrix and three-detail coefficients matrices are generated as follows:

$$L_n(u_i, u_j) = \left[H_x * \left[H_y * L_{n-1} \right]_{\downarrow 2,1} \right]_{\downarrow 1,2} (u_i, u_j) \qquad (5)$$

$$D_{n1}(u_i, u_j) = \left[H_x * \left[G_y * L_{n-1} \right]_{\downarrow 2,1} \right]_{\downarrow 1,2} (u_i, u_j) \qquad (6)$$

$$D_{n2}(u_i, u_j) = \left[G_x * \left[H_y * L_{n-1} \right]_{\downarrow 2,1} \right]_{\downarrow 1,2} (u_i, u_j) \qquad (7)$$

$$D_{n3}(u_i, u_j) = \left[G_x * \left[G_y * L_{n-1} \right]_{\downarrow 2,1} \right]_{\downarrow 1,2} (u_i, u_j) \qquad (8)$$

where low pass and high pass filters are illustrated by H and G, and '*' indicates the convolution operator. Sub-sampling along the rows and columns are represented by $\downarrow 2,1$ and $\downarrow 1,2$, and L_n is the low resolution image at scale n. An example of the original image and three detailed sub-images obtained using DWT is shown in Fig. 2.
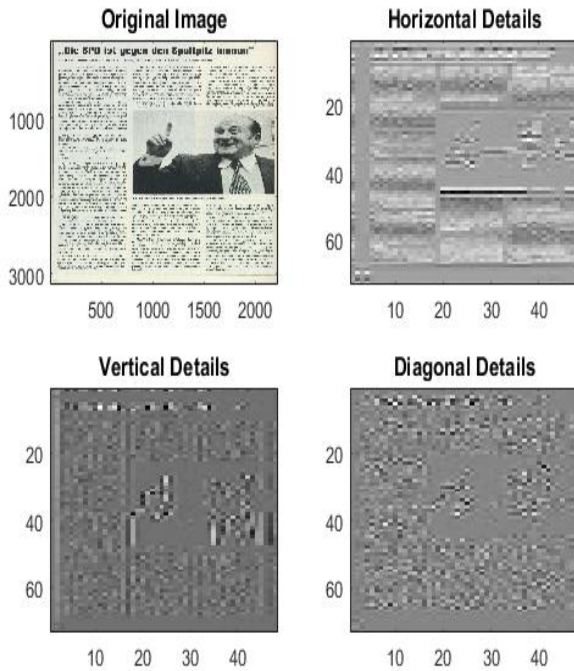

Fig. 2. Example of an original image and three sub-images generated using wavelet transform.

In our experimentation before applying a wavelet transform on document images, the images are normalised to 140×90 pixels. This is the minimum size of document images in the MTDB database, so we use this size for size normalisation in our experimental analysis. The under-sampled, horizontal, vertical, and diagonal images of size 71×45 pixels obtained by employing the wavelet transform on the given image are considered for wavelet feature extraction. The column-wise variance of the values in each under-sampled image is then computed. As a result, a variance wavelet feature vector of size 45 features is obtained from every under-sampled image. Feature vectors extracted from all three under-sampled images are finally concatenated to obtain the wavelet-based feature vector of size 135 = 3×45 for each input image.

2. Gist Operator

The Gist descriptor was initially introduced to make a holistic and low-dimensional representation of the structure of an image [13]. Gist descriptor has been used for other applications such as scene recognition and classification [12, 15, 16]. The Gist assigns to the meaningful information that a viewer can identify from a glimpse at an image [25]. The Gist descriptor usually consists of the semantic label and the spatial layout of the image. In this work, global features of document images measured orientations and spatial frequencies of image components. The colour image with $M \times N$ resolution is converted to an intensity image $i(x, y)$, and Fourier transform is applied on the intensity image, which is defined as:

$$I(f_x, f_y) = \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} i(x, y) h(x, y) e^{-j2\pi(f_x x + f_y y)} \qquad (9)$$

where f_x and f_y are the spatial frequency variables. The circular Hanning window is presented by $h(x, y)$ to reduce boundary affects [16]. The Gabor filter is used to represent the energy spectrum. In other words, the Gabor filter is employed to generate spectrum maps that are sampled with coarse resolution. The maps are generated at four different angles ($\theta_i = 0, 45, 90, 135$) and at four spatial scales ($c = 0,1,2,..,7$), which can be presented as:

$$M_i(c) = Gabor(\theta_i, c) \qquad (10)$$

Each one of these maps can represent one orientation in one scale of the input image that is decomposed into a fixed 4×4 grid of subregions. So, to compute 16 raw gist features $G_i^{k,l}(c, s)$ per map, mean intensity is calculated in each subregion

$$G_i^{d,z}(c, s) = \frac{1}{16WH} \sum_{u=\frac{dW}{4}}^{\frac{(d+1)W}{4}-1} \sum_{u=\frac{zH}{4}}^{\frac{(z+1)H}{4}-1} \lfloor M_i(c, s) \rfloor (u, v) \qquad (11)$$

where d and z are the horizontal and vertical direction, and W and H are the sizes of the map. Finally, 16 mean values of all 32 feature maps are linearly combined to produce a Gist feature set of size 16×32 = 512 features.

Since the Gist operator is employed on the original image as well as three different sub-images generated by the wavelet transform, four Gist feature vectors extracted from the original image and three sub-images are concatenated. As a result, the

final Gist feature vector of size 2048 = 4×512 features for each input image is constructed.

B. Creation of Knowledge-based Feature

Since two types of features, wavelet and Gist features, are extracted from the document image, two separate knowledge-based databases are created accordingly. In the Gist knowledge-based database all the Gist feature vectors are retained. Similarly, all the wavelet-based feature vectors are saved as the wavelet knowledge-based database to be used in the testing phase.

C. Similarity Measure and Fusion Strategy

To have a fast and efficient retrieval system, a fast measuring distance is required. In the previous studies, the k-nearest neighbor (k-NN) method has usually been used to find the similarity distance between a query image and trained document images [20, 26]. In this work, locality-sensitive hashing (LSH) [27] is taken into account. The LSH creates a group of buckets which contain similar data elements; therefore, time complexity is highly reduced compared to the k-NN.

1. Locality-Sensitive Hashing

For large databases of high dimension, locality-sensitive hashing [27] is the most valuable method using probability distributions, to retrieve similar images to a given query. Thus, the privilege of this method is to reduce the high dimensional feature space into remarkably small places, to search efficiently for the nearest neighbor.

In LSH method, a set of hash functions $(h_1, ..., h_{n_h})$ are selected randomly for adjusting LSH, in a metric space (s, d). In the metric space, s is defined as database and d as metric. Considering $P(A)$ as a probability of event A, the function family H is a (c, r, p_1, p_2) LSH family. Then, if and only if, r is given a positive, real number $c > 1$ and $x, y \in s$, any $h_i \in H$ satisfies the following conditions:

$$\begin{cases} p(h_i(x) = h_i(y)) \geq p_1, & if\ d(x,y) \leq r \\ p(h_i(x) = h_i(y)) \leq p_2, & if\ d(x,y) > c \times r \end{cases} \quad (12)$$

$$where\ p_1 > p_2$$

To index the data points in the database, LSH hashes all the points in the database with the hash functions. If $x, y \in s$ are close together by definition of the metric d, then k_x and k_y are identical. Each hash bucket includes data points with the same hash key that are located in a connecting area of the physical memory. So, the hash key is used to address the hash bucket. For orderly saving of the hash keys, a hash table is used that links to a non-empty hash bucket [28].

Giving a query point, LSH retrieves all the data points from the hash bucket to find the nearest point to a query. As in our proposed method, two knowledge-based databases are created for the Gist and wavelet features, and for a given query two different ranked lists of document images are obtained based on similarity scores computed by employing the LSH method.

2. Fusion Strategy

By considering the LSH method, document images with maximum similarity to a given query are ranked according to each feature extraction method. The classifier fusion method [29, 30] is a practical solution to consider properties of both methods to make a better decision. The following equation considers the achieved probability as

$$d'_a(y) = F\left(d_{a,b}(y), ... d_{a,b}(y)\right),\ b = 1, 2 \quad (13)$$

where a and b are the number of samples used for training and the number of classifiers respectively. The term F stands as the fusion function that can be the average, median, minimum, maximum or majority vote. For the proposed methods, the average function performed more accurate retrieval results among the other functions, so the average function is preferred. The average fusion method provides

$$P_a = \frac{1}{b}\sum_{m=1}^{b} p_{a,m} \quad (14)$$

where the output of the classifier d_b for a sample a is $p_{a,b} = d_{a,b}(y)$.

In the proposed method, the measured distances using Gist and the wavelet transform features for a given query image and the training sample a are called $p_{a,\text{Gist}}$ and $p_{a,\text{WT}}$. To combine the classifiers, a weighted average function is proposed to get better performance as follows:

$$P_a = \alpha * p_{a,\text{Gist}} + \beta * p_{a,\text{WT}} \quad (15)$$

$$\beta = 1 - \alpha$$

where α and β are demonstrated weights for the calculated distances according to the classifiers.

In the proposed method, the weights are determined by experiment and given as $\alpha = \frac{1}{3}$ and $\beta = \frac{2}{3}$. However, only slight variations in the results can occur by altering the value of weights. Considering weighted average function, new scores are measured and sorted in a descending order to rank the scores. As a final point, the most similar image with highest similarity to the given query is chosen out of the training database.

III. EXPERIMENTAL ANALYSIS

A. Database and Evaluation Metric

Two databases consisting of high dimensional images were considered for evaluation of the proposed method. These two different databases include a great diversity of page layouts.

The first database was the Media Team Document Database (MTDB) from the University of Oulu [31]. The MTDB includes three different resolutions; the size of high-resolution document images in MTDB is approximately 2800 × 1700 pixels. Eleven classes among 19 classes of the MTDB were considered for the experiments. In the experiments employing MTDB, one image from each category (154 images) was used for training, and the remaining 1,168 images were considered for testing.

The other database was ITESOFT [32]. The ITESOFT database is a collection of several scanned official documents in 12 different classes. The samples had not been equally distributed in all the classes. The ITESOFT database includes two different resolutions; high-resolution document images considered for our experiments are 3700 × 2500 pixel size. In the experiments, only 10 samples from each class were considered for training the method, and the remaining 996 images were used for testing.

The percentage of accuracy was considered for the evaluation metric. The document images in a database were ranked based on the maximum visual similarity to a given query image. The accuracy was measured based on the correct ranking of the document images. The results are displayed in the first position (Top-1), Top-3, Top-5, and Top-10. The Top-1 value indicates the percentage of the correct retrieval result at first place with respect to a given query. Subsequently, the Top-10 value represented the accuracy of the system when the system could rank the correct document image at the first 10 places with respect to a given query.

B. Results and Discussion

The high-resolution images of both the databases were considered for the experiments. The results of our proposed method are demonstrated in Table I. By applying the proposed method on MTDB, 66.3% and 96.7% of the document images were retrieved correctly at Top-1 and Top-10, respectively. Similarly, the proposed method respectively provides 70% and 90.2% correct retrieval accuracies in Top-1 and Top-10 for ITESOFT database.

TABLE I. THE RESULTS OBTAINED USING OUR PROPOSED METHOD

Method	Database	Top-1 (%)	Top-3 (%)	Top-5 (%)	Top-10 (%)
Gist & Wavelet	MTDB	66.3	87.5	93.2	96.7
	ITESOFT	70.0	80.0	85.7	90.2

TABLE II. THE RESULTS OBTAINED USING GIST DESCRIPTOR AND WAVELET TRANSFORM

Method	Database	Top-1 (%)	Top-3 (%)	Top-5 (%)	Top-10 (%)
Only Gist features from original image	MTDB	55.3	75.8	83.3	91.8
	ITESOFT	69.0	75.5	79.8	85.7
Gist features from original image & under-sampled	MTDB	59.7	**80.9**	**87.9**	**95.3**
	ITESOFT	64.6	**76.0**	**81.2**	**87.5**
Wavelet features	MTDB	61.5	77.5	82.9	88.1
	ITESOFT	50.2	67.9	74.5	86.2

To demonstrate the significance of our proposed method of using two different features along with a fusion strategy, the results obtained by the Gist from the original image and wavelet transform alone, as well as Gist from the original image and under-sampled images using the MTDB and ITESOFT datasets for document image retrieval, are shown in Table II. From Table II, it can be noted that the Gist features extracted from the original images and under-sampled images obtained from wavelet transform provided higher results on both databases compared to the other features (Gist from only original image and wavelet features). It is also apparent from Table I and Table II that the proposed method could provide higher retrieval accuracy compared to the Gist and wavelet transform features alone. The results shown in Tables I and II are significant in at least two major aspects. The first point is that employing Gist operator on the under-sampled images can provide more information regarding the image result for higher retrieval accuracy. The other point is the significance of classifier fusion which considerably increases the accuracy.

C. Comparative Analysis

To make a direct comparison of the results obtained from the proposed method with the state-of-the-art results, the methods that used the MTDB data for experimentation are considered. The methods which used Gabor filter [33], local binary pattern (LBP) [26], and the combination of LBP and wavelet transform [20] for document image classification and retrieval were considered for comparison. The retrieval results obtained from different state-of-the-art methods are shown in Fig. 3. As is demonstrated in Fig. 3, the results obtained from the proposed method outperform the state-of-the-art results.

Moreover, the proposed method is more efficient regarding the retrieval time. In the methods [20, 26] that used the nearest neighbour method for similarity measure and document retrieval, the most similar document image to a given query was obtained in 0.018 second. However, in our proposed method, by considering LSH method, the retrieval time is reduced to 0.008 second.
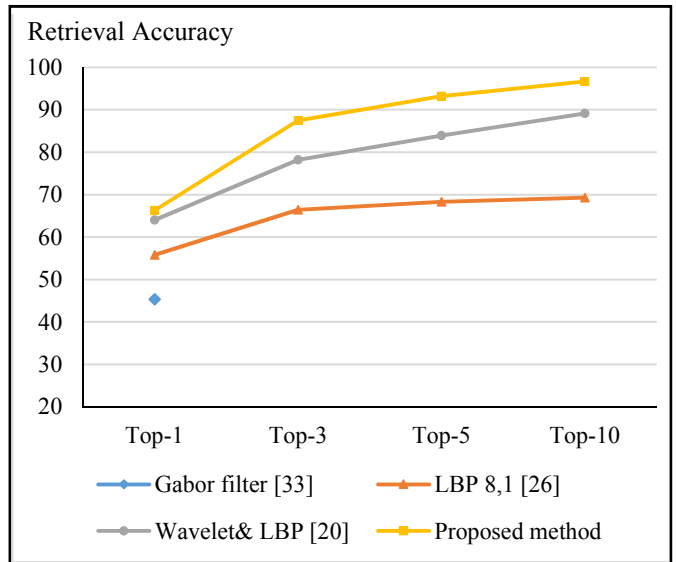

Fig. 3. Comparison of the results obtained from the proposed method and state-of-the-art methods.

IV. CONCLUSION

In the present study, an efficient document image retrieval method based on two different texture features along with a fusion strategy is presented. The wavelet transform is taken into account and under-sampled images are generated from the original image. Gist features are then extracted from under-sampled images as well as the original image. By applying Gist descriptor, orientations and spatial frequencies of the document images are measured. Moreover, statistical features from under-sampled images are obtained. The similarity between a given query image and document images in the database are measured separately, using LSH. Finally, a classifier fusion with the mean function is considered to obtain document image retrieval results. Experiments on two databases show that the proposed approach can provide promising results for document image retrieval.

REFERENCES

[1] K. Takeda, K. Kise, and M. Iwamura, "Real-Time Document Image Retrieval for a 10 Million Pages Database with a Memory Efficient and Stability Improved LLAH," In Proceedings of the International Conference on Document Analysis and Recognition, pp. 1054-1058, 2011.

[2] J. Kumar, P. Ye, and D. Doermann, "Structural similarity for document image classification and retrieval," Pattern Recognition Letters, vol. 43, pp. 119-126, 2014.

[3] S. Marinai, "Page Similarity and Classification," Handbook of Document Image Processing and Recognition, pp. 223-253, 2014.

[4] H. Gao, M. Rusinol, D.Karatzas, J. Lladós, T. Sato, M. Iwamura, et al., "Key-Region Detection for Document Images--Application to Administrative Document Retrieval," In Proceedings of the International Conference on Document Analysis and Recognition, pp. 230-234, 2013.

[5] J. Li, Z.G. Fan, Y. Wu, and N. Le, "Document Image Retrieval with Local Feature Sequences," In Proceedings of the International Conference on Document Analysis and Recognition, pp. 346-350, 2009.

[6] G. Zhu, Y. Zheng, and D. Doermann, "Signature-Based Document Image Retrieval," In Proceedings of the International Conference on Image Processing, Computer Vision & Pattern Recognition, pp. 752-765, 2008.

[7] A. Alaei, P. P. Roy, and U. Pal, "Logo and seal based administrative document image retrieval: a survey," Computer Science Review, vol. 22, pp. 47-63, 2016.

[8] C. Shin and D. Doermann, "Document Image Retrieval Based on Layout Structural Similarity," In Proceedings of the International Conference on Image Processing, Computer Vision & Pattern Recognition, pp. 606-612, 2006.

[9] G. Zhu and D. Doermann, "Logo matching for document image retrieval," In Proceedings of the International Conference on Document Analysis and Recognition, pp. 606-610, 2009.

[10] M. Mehri, P. Héroux, P. Gomez-Krämer, and R. Mullot, "Texture feature benchmarking and evaluation for historical document image analysis," International Journal on Document Analysis and Recognition, pp. 1-35, 2017.

[11] N. Journet, J.Y. Ramel, R. Mullot, and V. Eglin, "A proposition of retrieval tools for historical document images libraries," In Proceedings of the International Conference on Document Analysis and Recognition, pp. 1053-1057, 2007.

[12] Z. Li and L. Itti, "Saliency and gist features for target detection in satellite images," IEEE Transactions on Image Processing, vol. 20, pp. 2017-2029, 2011.

[13] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," International Journal of Computer Vision, vol. 42, pp. 145-175, 2001.

[14] A. Oliva and A. Torralba, "Building the gist of a scene: The role of global image features in recognition," Progress in Brain Research, vol. 155, pp. 23-36, 2006.

[15] W. Liu, S. Kiranyaz, and M. Gabbouj, "Robust scene classification by gist with angular radial partitioning," In Proceedings of the International Symposium on Communications Control and Signal Processing, pp. 1-6, 2012.

[16] C. Zhao, C. Liu, and Z. Lai, "Multi-scale gist feature manifold for building recognition," Neurocomputing, vol. 74, pp. 2929-2940, 2011.

[17] Z. Xizhi, "The application of wavelet transform in digital image processing," In Proceedings of the International Conference on Multimedia and Information Technology, pp. 326-329, 2008.

[18] J. Vargas, C. M. Travieso, J. B. Alonso, and M. A. Ferrer, "Off-line signature verification based on gray level information using wavelet transform and texture features," In Proceedings of the International Conference on Frontiers in Handwriting Recognition, pp. 587-592, 2010

[19] M. Zhao, Q. Chai, and S. Zhang, "A method of image feature extraction using wavelet transforms," In Proceedings of the Emerging Intelligent Computing Technology and Applications, pp. 187-192, 2009.

[20] F. Alaei, A. Alaei, M. Blumenstein, and U. Pal, "Document image retrieval based on texture features and similarity fusion," In Proceedings of the International Conference on Image and Vision Computing New Zealand, pp. 1-6, 2016.

[21] A. Haar, "On the theory of orthogonal function systems," Mathematische Annalen, vol. 69, pp. 331-371, 1910.

[22] M. Singha, "Content Based Image Retrieval using Color and Texture," Signal & Image Processing: An International Journal, vol. 3, pp. 39-57, 2012.

[23] R. C. Gonzalez and R. E. Woods, "Book on "Digital Image Processing"," ed: Prentice-Hall of India Pvt. Ltd, 2005.

[24] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," Pattern Analysis and Machine Intelligence, vol. 11, pp. 674-693, 1989.

[25] M. Douze, H. Jégou, H. Sandhawalia, L. Amsaleg, and C. Schmid, "Evaluation of gist descriptors for web-scale image search," In Proceedings of the International Conference on Image and Video Retrieval, pp. 19-26, 2009,

[26] F. Alaei, A. Alaei, U. Pal, and M. Blumenstein, "Document Image Retrieval Based on Texture Features: A Recognition-Free Approach," In Proceedings of the International Conference on Digital Image Computing: Techniques and Applications, pp. 1-7, 2016.

[27] A. Gionis, P. Indyk, and R. Motwani, "Similarity search in high dimensions via hashing," The International Journal on Very Large Data Bases, pp. 518-529, 1999.

[28] Y. Zhou, C. Liu, N. Li, and M. Li, "A novel locality-sensitive hashing algorithm for similarity searches on large-scale hyperspectral data," Remote Sensing Letters, vol. 7, pp. 965-974, 2016.

[29] L. I. Kuncheva, "A theoretical study on six classifier fusion strategies," Pattern Analysis & Machine Intelligence, vol. 24, pp. 281-286, 2002.

[30] J. Kittler, M. Hatef, R. P. Duin, and J. Matas, "On combining classifiers," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 20, pp. 226-239, 1998.

[31] J. Sauvola and H. Kauniskangas, "MediaTeam document database II," A CD-ROM collection of document images, University of Oulu Finland, 1999.

[32] F. Alaei, N. Girard, S. Barrat, and J.-Y. Ramel, "A New One-Class Classification Method Based on Symbolic Representation: Application to Document Classification," In Proceedings of the International Workshop on Document Analysis Systems, pp. 272-276, 2014.

[33] C.-F. Tsai, "On Classifying Digital Accounting Documents," International Journal of Digital Accounting Research, vol. 7, pp. 53-71, 2007.