Journal of King Saud University – Computer and Information Sciences xxx (2018) xxx–xxx

Contents lists available at ScienceDirect

Journal of King Saud University – Computer and Information Sciences

journal homepage: www.sciencedirect.com

# Content-based medical image retrieval by spatial matching of visual words

P. Shamna *, V.K. Govindan, K.A. Abdul Nazeer

*Department of Computer Science, National Institute of Technology Calicut (NITC), Calicut, Kerala 673 601, India*

ARTICLE INFO

*Article history:*
Received 18 July 2018
Revised 19 September 2018
Accepted 5 October 2018
Available online xxxx

*Keywords:*
Content-Based Medical Image Retrieval
Spatial location-based image retrieval
Bag of Visual Words
Visual word encoding techniques
Spatial pyramid

ABSTRACT

Content-Based Image Retrieval (CBIR) systems have recently emerged as one of the most promising and best image retrieval paradigms. To pacify the semantic gap associated with CBIR systems, the Bag of Visual Words (BoVW) techniques are now increasingly used. However, existing BoVW techniques fail to capture the location information of visual words effectively. This paper proposes an unsupervised Content-Based Medical Image Retrieval (CBMIR) framework based on the spatial matching of the visual words. The proposed method efficiently computes the spatial similarity of visual words using a novel similarity measure called the Skip Similarity Index. Experiments on three large medical datasets reveal promising results. The location-based correlation of visual words assists in more accurate and efficient retrieval of anatomically diverse and multimodal medical images than the state-of-the-art CBMIR systems.



## 1. Introduction

Medical image retrieval research has emanated as the chalk horse for doctors and radiologists, particularly due to the prodigious increase in the use of medical images in the recent past. The advancement in medical imaging techniques (Diekhoff et al., 2017; Gherase et al., 2017; Hussain et al., 2017; Lord et al., 2017; Miwa and Otsuka, 2017; Woźniak et al., 2018b; Xie et al., 2018) and automatic diagnosis systems (Muhammad et al., 2017; Wang et al., 2018; Woźniak et al., 2018a; Woźniak and Połap, 2018) cause the accumulation of an extensive collection of medical images in hospitals. Modern hospitals store medical images in DICOM (Digital Imaging and Communications in Medicine) format. Nowadays, physicians rely on the text-based DICOM attributes to retrieve images from medical image repositories. However, the DICOM attributes need to be hand-crafted and manually annotated – a process which is error-prone and expensive (Mustra et al., 2008). So a reliable Content-Based Medical Image Retrieval (CBMIR) system based on visual cues needs to be developed to retrieve images more efficiently.

The Bag of Words were first applied in document classification and later adapted to computer vision applications like image classification and retrieval. BoVW (Csurka et al., 2004) represents an image as a histogram of visual words based on the count of visual words in the images. Recently, BoVW-based image retrieval has gained much popularity because of its superior expressive power and semantic discriminative power compared to other visual feature selection methods. BoVW methods record the count of visual words to represent and retrieve images from the dataset. Spatial pyramid matching (Lazebnik and Schmid, 2006) is the most popular BoVW method, which incorporates spatial information into the unordered histogram. In spatial pyramid mapping an image is partitioned into smaller blocks, and the BoVW of each sub-region is computed. Recent image retrieval systems employ advanced visual word encoding (Huang et al., 2011; Sánchez et al., 2013; Suharjito and Santika, 2017; Wu et al., 2012) and topic modelling concepts (Blei et al., 2010; Li et al., 2015; Swamy and Holi, 2013) to improve the semantic representation of images. (Avni et al., 2011) employed BoVW generated from the local patches of medical images for CBMIR. (Haas et al., 2012) used super pixel based interest point using SIFT-based descriptor to create BoVW for CBMIR. (Yang et al., 2012) retrieved focal liver sections by employing BoVW of densely sampled raw intensity values in the Region Of Interest (ROI). (García Seco De Herrera et al., 2013) created Bag of Colors to classify and retrieve coloured medical images.

* Corresponding author.
*E-mail addresses:* shamnapootheri@gmail.com (P. Shamna), vkg@nitc.ac.in (V.K. Govindan), nazeer@nitc.ac.in (K.A. Abdul Nazeer).

Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.jksuci.2018.10.002
1319-1578/

(Foncubierta-Rodríguez et al., 2013) implemented topic modelling technique (Hofmann, 1999) to remove trivial visual words from the dictionary, and retrieve images effectively using more meaningful BoVW. (Huang et al., 2014) proposed a partitioning learning method to retrieve images by extracting BoVW from the tumour region marked by radiologists. (Srinivas et al., 2015) proposed CBMIR using BoVW and sparse coding method by partitioning images into different regions. (Cao and Cao, 2016) represented images by combining BoVW and topic modelling (Hofmann, 1999) features to classify medical images.

CBMIR methods relying on the count based BoVW (Cao and Cao, 2016; Foncubierta-Rodríguez et al., 2013; García Seco De Herrera et al., 2013; Haas et al., 2012) ignores the discriminating location information of visual words, to compute the similarity between the images. Only a few methods consider the spatial information by partitioning the images into different regions (Huang et al., 2014; Srinivas et al., 2015; Yang et al., 2012). Partitioning methods are rotation variant and incapable of analysing and comparing global spatial details of visual words - an essential feature to evaluate the anatomical or structural similarity in medical images. It seems that, the recent CMBIR methods employing BoVW fail to integrate and compare the spatial information of visual words efficiently.

Retrieving medical images from the anatomically diverse multimodal dataset is a challenging task. Medical images are mostly grey coloured, and the count of visual words alone is incapable of representing or analysing the anatomical structure of the medical images properly. The use of spatial information of visual words would further help to enhance the representation of medical images. In this paper, we present a location-based correlation technique for retrieving medical images from datasets consisting of different modality and body regions. The main contributions of the paper are

- *Location features:* The spatial location of each visual word is represented as the distance of the visual word from the image centre, and the angle of the visual word from a reference axis. Arranging the location features in clockwise order incorporates rotation invariant spatial information of the visual words in the images.
- *Skip Similarity Index:* We propose a novel correlation measure, Skip Similarity Index to compute the location similarity of variable-length location feature vector. The count of a visual word in medical images belonging to the same class may differ due to the difference in the anatomical structure of each person or difference in imaging conditions. Instead of comparing the count of visual words, Skip Similarity Index effectively measures the sequential location similarity between each visual word by skipping the unmatched location features.

The remainder of this paper is organised as follows: Section 2 describes the proposed methodology. Section 3 provides the experimental results and comparison of the proposed method with state-of-the-art CBIR systems. Section 4 concludes the paper by highlighting the significant contributions of this work, and scope for future research.

## 2. Materials and method

The proposed image retrieval system has two phases: 1) Offline Location feature extraction and 2) Online Retrieval. Fig. 1 illustrates the workflow of the proposed CBMIR framework. In the location feature extraction phase, the dictionary and visual words are generated from the pre-processed (resizing, contrast enhancement etc.) images in the database. The locations of each visual word are calculated, and then arranged in clock-wise sequential order starting from the first quadrant. These location features are stored in the feature database. In the retrieval phase, the locations of visual words in the query image are estimated. The sequentially ordered location features in the query image are compared with that of the ones in database. The retrieval system computes the location-wise similarity of each visual word in the query and database images (test images) using the proposed Skip Similarity Index. The images are finally ranked based on the location Skip Similarity Index. Now, when a query is to be processed, only the top-ranked images are retrieved from the database.

### 2.1. Visual word generation process

Visual words are meaningful visual representations of patches around significant image regions (key-points). The four main steps involved in visual word generation processes are as given below.

1. *Key-point detection:* We used the scale invariant blob detector – Difference of Gaussian (DoG) filter given in Eq. (1) to detect key-points in the images.

$$DoG(x,y,\sigma) = (G(x,y,k\sigma) - G(x,y,\sigma)) * I(x,y) \quad (1)$$

where $G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}$ represent Gausian filter,$I(x,y)$ represents the input image and $*$ represents the convolution operation at x and y (Lowe, 2004). Seven different scales are used starting from initial scale $\sigma_1 = 2$, scaling factor $k = \sqrt{2}$, and final scale $\sigma_7 = 16$, to detect the key-points in the images. The DoG filter detects edges, by removing high-frequency components representing noise and low-frequency components representing the homogeneous areas in the images. The initial key-points ((x,y) image pixel coordinates) are selected as the extrema (maxima or minima) of DoG at two nearby scales. The final key-points are selected by discarding low-contrast key-points and eliminating the unstable key-points located along the edges from the initial key-points (Lowe, 2004). Fig. 2 portrays the detected key-points and patches around the key-points in a chest x-ray image.

2. *Key-point description*: We described the patches around the key-points by the Scale Invariant Feature Transform (SIFT) descriptor (Lowe, 2004). SIFT is one of the successful local feature descriptor used in the recent medical image applications (Cao et al., 2015; García Seco De Herrera et al., 2013; García Seco de Herrera and Müller, 2014; Markonis et al., 2014; Simpson et al., 2015; Villegas et al., 2015). SIFT features are the 4X4 histogram of 8 orientations (128 elements of feature vector) in the neighbourhood of each key-point.
3. *Visual word generation:* The visual words are generated by clustering the SIFT descriptors using K-Means clustering. Each cluster centre represents a visual word in the dictionary. To avoid poor clustering in K-means, which leads to form visual words with similar meaning, we initialised the cluster centres with a minimum Euclidean distance (Distance Threshold) from each other using Simple Cluster Seeking (SCS) algorithm (Tou and Gonzalez, 1974). As the user provides the number of clusters in K-Means, Distance Threshold must be provided as an input in the SCS algorithm. The optimum Distance Threshold is selected by evaluating the clusters formed by randomly chosen distances by the Calinski Harabasz variance ratio criterion (Calinski and Harabasz, 1974).
4. *Quantisation:* The patches around the key-points in the images are matched to the closest visual word in the dictionary. BoVW (Csurka et al., 2004) represents an image as a histogram of visual words based on the frequency of visual words in the images. Fig. 3 depicts the different steps in the visual words generation process and BoVW representation of images.

Please cite this article in press as: Shamna, P., et al. Content-based medical image retrieval by spatial matching of visual words. Journal of King Saud University – Computer and Information Sciences (2018), https://doi.org/10.1016/j.jksuci.2018.10.002

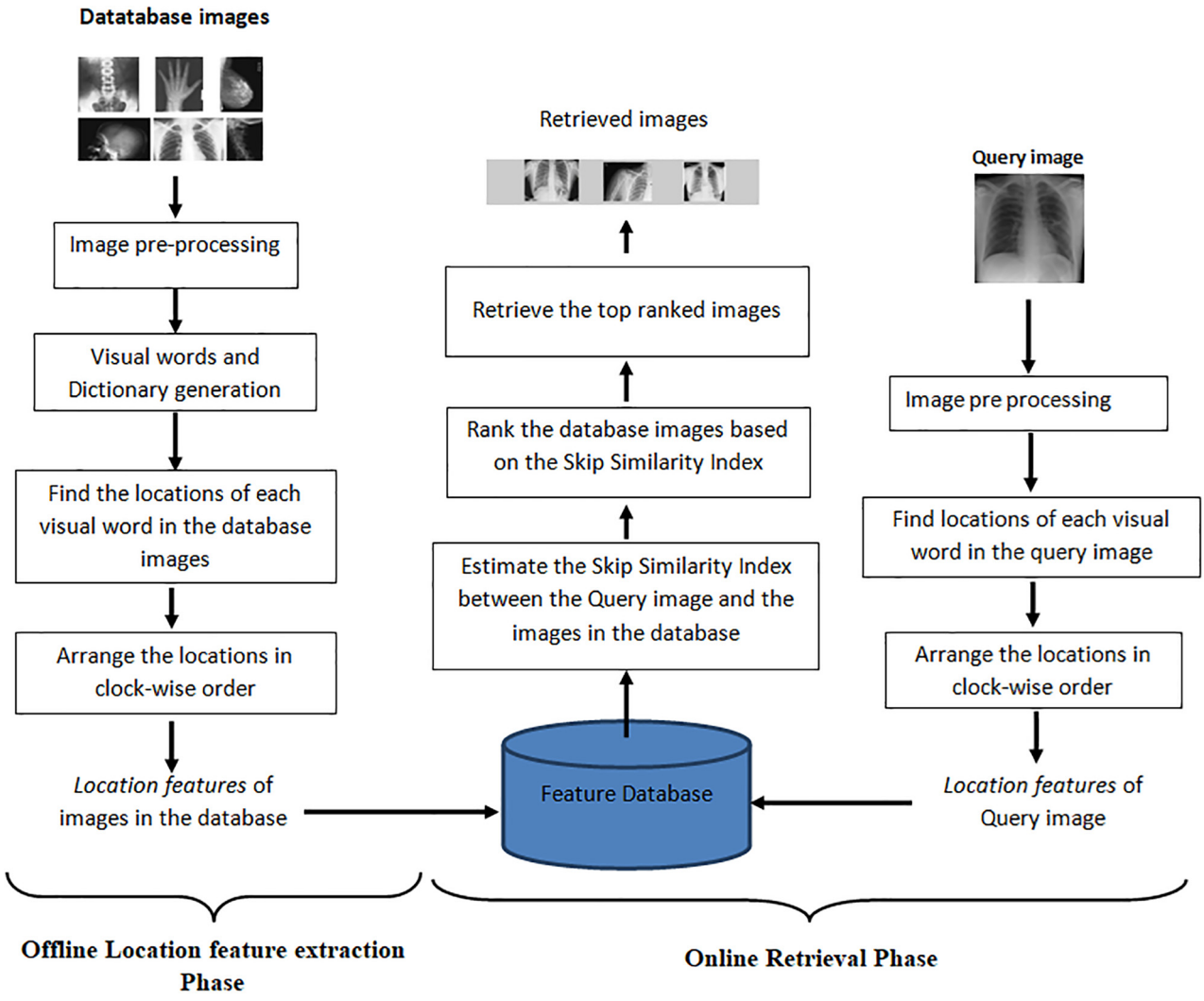


**Fig. 1.** Workflow of the proposed image retrieval system.

**Fig. 2.** Salient image regions (key-points) detected by the Difference of Gaussian key-point detector in a chest X-ray. The 'x' symbol indicates the pixel coordinate position of the key-points and circles indicate the patches around the key-points.

### 2.2. Proposed method

#### 2.2.1. Location feature

The location information of visual words is a significant feature in the image based applications. BoVW method fails to incorporate spatial details of visual words in the images. We propose a location-based correlation technique of the visual words using Skip Similarity Index. The proposed method efficiently captures the proximity in the locations of visual words, and retrieves anatomically similar images more accurately. The locations of a visual word 'v' in an image are shown in Fig. 4. The location features of 'v' are recorded as the distance and angle of each occurrence of 'v' from the centre of the image and the reference axis. These location features are arranged in the clockwise order, starting from the first quadrant to preserve the sequential ordering of the locations of visual words in the image plane.

#### 2.2.2. Skip Similarity Index

Visual words in medical images are the grey coloured patches extracted from the significant image regions. The locations of these visual words will help to discriminate the anatomical structure of the medical images. The proposed method estimates the location-

Please cite this article in press as: Shamna, P., et al. Content-based medical image retrieval by spatial matching of visual words. Journal of King Saud University – Computer and Information Sciences (2018), https://doi.org/10.1016/j.jksuci.2018.10.002

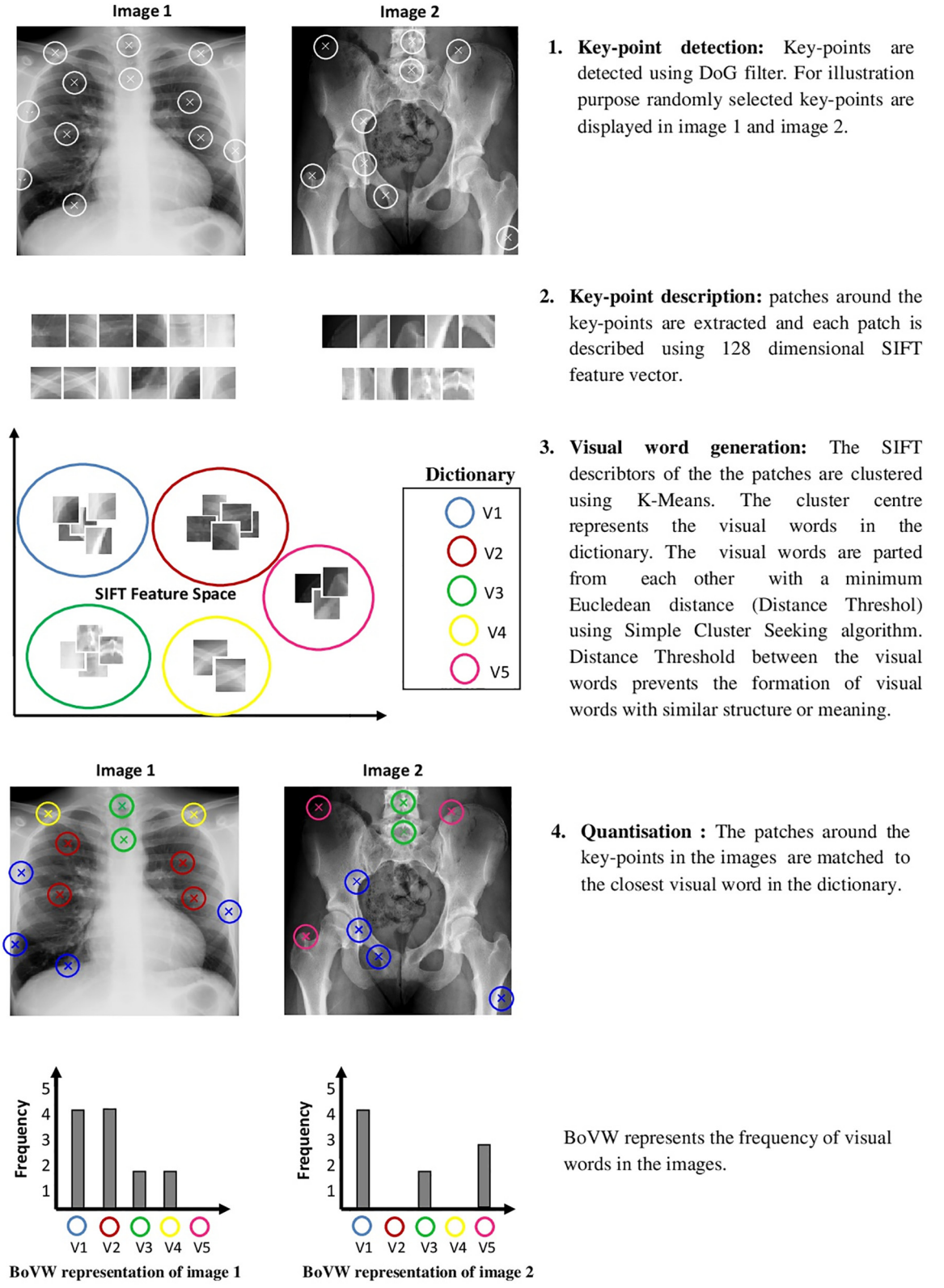


**Fig. 3.** Different steps in visual word generation and Bag of Visual Words (BoVW) representation process. The colour of the circle represent the visual words in the dictionary and the 'x' symbol represents the position of the visual words in the images.

based match of visual words, to improve the retrieval performance of medical images from anatomically diverse medical datasets. The length of the location feature vector for a visual word depends on the occurrences of the visual word in the images. We propose a novel Skip Similarity Index to compute the location similarity of these variable-length location feature vector, by skipping the

Please cite this article in press as: Shamna, P., et al. Content-based medical image retrieval by spatial matching of visual words. Journal of King Saud University – Computer and Information Sciences (2018), https://doi.org/10.1016/j.jksuci.2018.10.002

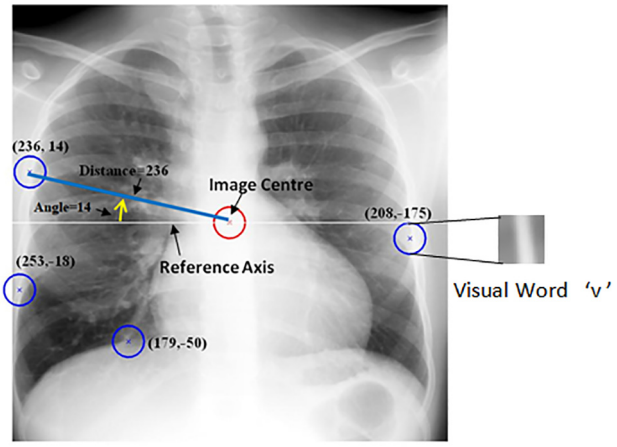


**Fig. 4.** Locations of visual word 'v' in an image recorded as the distance and angle.

unmatched features. Fig. 5 shows the location feature of a visual word 'v' in three images. The count of the visual word 'v' in the Query image and image A is four, whereas in image B visual word 'v' occurs in five locations. The Query image is matched with image B, when we compute the similarity of images based on the count of the visual word 'v'. The count-based method is unable to analyse the structural similarity between the Query image and image B. The location feature of the visual word 'v' in the Query image matches with location features in image B than that in image A. As demonstrated in the images shown in Fig. 5(b) and (c), count of a visual word 'v' can vary in the images with similar anatomy due to the physical structure of persons or due to the imaging conditions. To encode such variable length location features more accurately, we propose a new similarity index called Skip Similarity Index. Skip Similarity Index correlates the similarity between location features by skipping the extra or missing visual words.

The proposed location based correlation of visual words using Skip Similarity Index is detailed in Algorithm 1. A match between the locations of a visual word in two images is determined based on a Deviation Threshold. Deviation Threshold decides the deviation or difference allowed in locations while estimating the match between two location features. The locations are skipped if the normalised absolute difference between two location features is higher than the Deviation Threshold. The clock-wise ordering of the location features preserves the sequential arrangement of visual words in the image plane. The experimental results show that skip factor (the number of consecutive skips) limited to one location feature gives better results by considering the sequential order of locations in similar images. Circular correlation of the location features are performed to ensure the matching of location features to compute the Skip Similarity Index.

Fig. 6 shows an example of the proposed location correlation of the Query image with image A and B, given in Fig. 5. If the normalised absolute difference in location is higher than the Deviation Threshold (20%), we skipped the location (indicated by red shaded regions). If the difference in location is within the Deviation Threshold, we matched the locations (shown by grey shaded regions). The Skip Similarity Index between Query image and image A is 25% (as two locations are matched out of eight location features). The Skip Similarity Index between Query image and image B is 89% (as eight locations are matched out of nine location features). The proposed Skip Similarity Index efficiently computes the location-wise match of visual words in two images as shown in Fig. 6.

**Algorithm 1:** Skip Similarity Index calculation of visual word v in Query and Test image

*Input:*

| | |
|---|---|
| *qLocationv:* $[ql_1^v, ql_2^v, ql_3^v, \ldots, ql_M^v]$ | •Location features of visual word v in Query image, where *M* is the number of occurences of v in Query image |
| *tLocationv:* $[tl_1^v, tl_2^v, tl_3^v \ldots, tl_n^v]$ | •Location features of visual word v in Test image, where *N* is the number of occurences of v in Test image. |
| *DT:* Deviation Threshold | •The absolute difference allowed between the locations |

*Output:*

| | |
|---|---|
| $SSI^v$(Query,Test) | •Skip Similarity Index of visual word v in the Query and Test image. |

1. Let m ← 1,n ← 1, •Start from the first location feature of Query image and Test image.
2. *LocationMatch = 0* • Initialize *LocationMatch:* Location-wise match to zero.
3. Compute the *LocationDif f erence* as the normalized absolute difference in the locations $ql_m^v$ and $tl_n^v$.
4. **if** *(LocationDif ference <= DT)* **then**
5. *LocationMatch = LocationMatch + 1* • Match the location features $ql_m^v$ and $tl_n^v$,
6. **else**
7. *SkipQueryDi f f erence* = Normalized absolute difference of $ql_{(m+1)}^v$ and $tl_n^v$
8. *SkipTestDi f f erence* = Normalized absolute difference of $ql_{(m)}^v$ and $tl_{(n+1)}^v$
9. *MinLocationDif ference = Min(SkipQueryDif ference, SkipTestDif ference)*
10. **if** *(MinLocationDif ference <= DT)* **then**
11. **if (** *MinLocationDif ference = SkipQueryDif f erence)* **then**
12. Skip the $ql_{(m)}^v$ and Update m = m + 1
13. **else**
14. Skip the $tl_{(n)}^v$ and Update n = n + 1
15. **end if**
16. *LocationMatch = LocationMatch + 1* •Match the location features $ql_m^v$ and $tl_n^v$ skip the other location features
17. **end if**
18. **end if**
19. m = m + *1, n = n + 1* •Move to next location feature
20. Repeat the steps 3 to 19 until the location features of Test or Query image get exhausted
21. $SSI^v$ *(Query, T est)* = (2 * *LocationMatch)/(M + N)* •*Two location features* ($ql_m^v$ *and* $tl_n^v$) matched in each *Location Match*

The Skip Similarity Index of a visual word 'v' in ($SSI^v$)the Query and Test image is calculated based on the count of location-wise match of 'v' in the Query and Test image (Eq. (2)). The method estimates the Total Skip Similarity Index *(TotalSSI)* between the Query and the Test image by summing the Skip Similarity Index of all the visual words in the dictionary as in Eq. (3). The proposed CBMIR system ranks and retrieves images from the database based on the *TotalSSI* score. Top-ranked images have higher *TotalSSI* score, indicating stronger location based correlation with visual words in the query image.

Please cite this article in press as: Shamna, P., et al. Content-based medical image retrieval by spatial matching of visual words. Journal of King Saud University – Computer and Information Sciences (2018), https://doi.org/10.1016/j.jksuci.2018.10.002

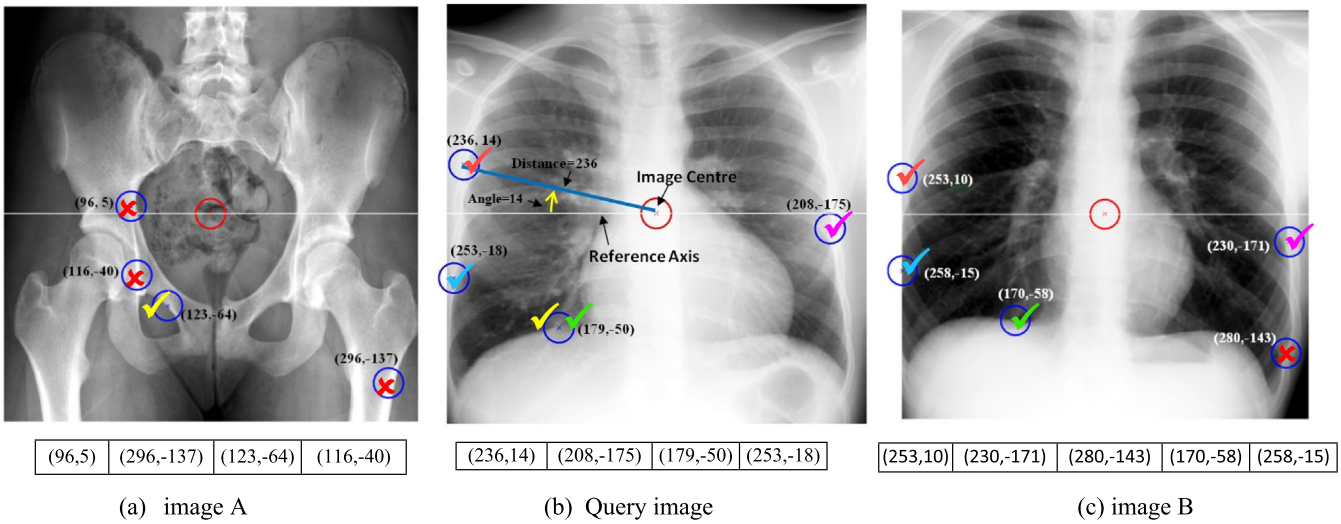


**Fig. 5.** Location features of visual word 'v' in three images (a) image A (b) Query image (c) image B. The ✗ symbol represents the mismatched location and ✓ symbol represents the matched locations of image A and B with the Query image, the colour of ✓ symbol represents the corresponding matches in locations.

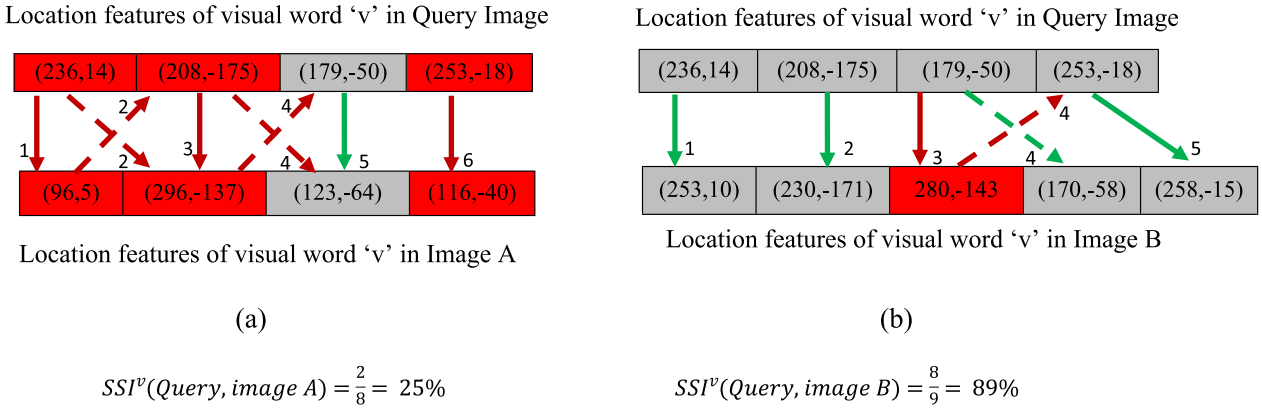


**Fig. 6.** The number near the arrows indicates the order in which matching was executed. The Red arrows indicate the mismatch and green arrows indicate match in locations. Solid arrow indicates match or mismatch without skipping the location feature. Dotted arrow indicates the match or mismatch by skipping a location feature. Red shaded regions are skipped locations, and grey shaded regions indicate the matched locations. (a) Skip Similarity Index calculation of Query image and image A (b) Skip Similarity Index calculation of Query image and image B.

$$SSI^v(Query, Test) = \frac{\text{Count of locations - wise matched visual word 'v' in the Query image and Test image}}{\text{Total number of visual word 'v' in Query image and Test image}} \quad (2)$$

$$TotalSSI(Query, Test) = \sum_{v=1}^{K} (SSI^v(Query, Test)) \quad (3)$$

where $K$ is total number of visual words in the dictionary.

## 3. Experimentation and results

### 3.1. Experimental tools and design

The proposed method was implemented using Matlab 2015b software. Thousand images comprising of ten classes were selected from the dataset to optimise the system parameters - Distance Threshold and Deviation Threshold. The Distance Threshold is used to initialise the cluster centres (K) in K-means and Deviation Threshold is used to determine the match between two location features. We used Simple Cluster Seeking (SCS) method (Tou and Gonzalez, 1974) to initialise fixed cluster centres of K-means to get a consistent result in each run. SCS will ensure that cluster centres or visual words in the dictionary are parted from each other with a Distance Threshold to avoid synonymous visual words in the dictionary. The clusters are initialised by randomly selected Distance Thresholds and evaluated by Calinski Harabasz variance ratio criterion (Calinski and Harabasz, 1974). Based on the evaluation, a Distance Threshold of 550 Euclidian distance measure between the clusters was fixed for the SCS. The SCS with the Distance Threshold 550 initialised the number of clusters (K) for Multimodal and Kvasir dataset as 55 and 50 in the experiments. Based

Please cite this article in press as: Shamna, P., et al. Content-based medical image retrieval by spatial matching of visual words. Journal of King Saud University – Computer and Information Sciences (2018), https://doi.org/10.1016/j.jksuci.2018.10.002

on heuristics the Deviation Threshold was set as 20% for the experiments. We matched the locations of a visual word if the normalised absolute difference between the locations were less than or equal to 20%. More details about the optimisation procedure are given in the Appendix A.

### 3.2. Dataset description

A direct comparison with many CBMIR systems was not possible since the images used for the experiments were not publicly available. We assessed the performance of the proposed system with two large publically available medical image datasets consisting of different body organs. All images were resized to 512 × 512 and colour images converted to a grey scale format for evaluation.

#### 3.2.1. Multimodal dataset

A multimodal dataset was gathered from publicly available medical databases consisting of 7200 images from different modalities (MR, CT, PT, PET, OPT, X-ray etc.). Based on body organ, the dataset was divided into 24 classes, and each class contains 300 images. Twenty two classes are selected from cancer imaging archive ("The Cancer Imaging Archive (TCIA) – A growing archive of medical images of cancer," 2014), and the 23rd class (Eye) is selected from Messidor dataset (Decencière et al., 2014). More details about the 23 classes are given in (Qayyum et al., 2017). The 24th class was taken as side chest X-ray images from IRMA database (Lehmann et al., 2004). Fig. 7 presents sample images from the Multimodal dataset.

#### 3.2.2. IRMA 2009 dataset

IRMA database (Lehmann et al., 2004) consists of scanned grey scale X-ray images of the different body parts in different modality and orientation as shown in Fig. 8. IRMA Images are characterised by variation in contrast and intensity. The images in IRMA contained dominant artefacts such as, artificial implants, X-ray boundaries etc. IRMA images also have significant variations within the class and substantial similarities between two classes. All these variations make automatic retrieval of medical images using IRMA databases, a challenging task.

#### 3.2.3. Kvasir dataset

The Kvasir dataset (Pogorelov et al., 2017) consists of 4000 coloured endoscopic images annotated by medical experts. The images have been grouped into 8 different classes (500 images in each class) based on anatomy, pathology, or polyp removal procedure. Fig. 9 shows sample images from Kvasir dataset.

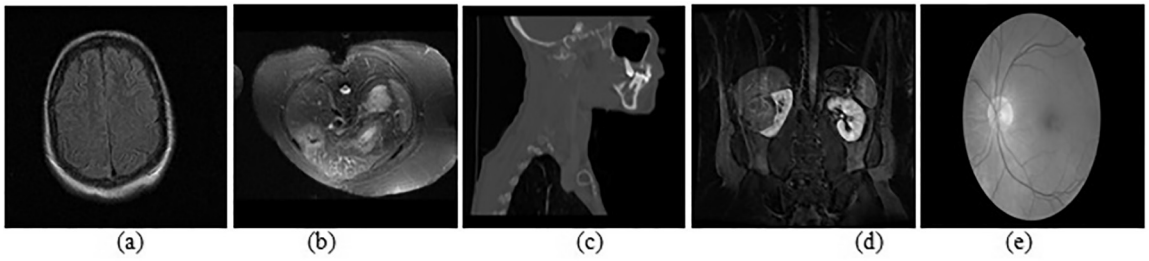


**Fig. 7.** Sample images from Multimodal dataset consisting of different body part (a) Brain (b) Liver (c) Thyroid (d) Kidney (e) Eye.

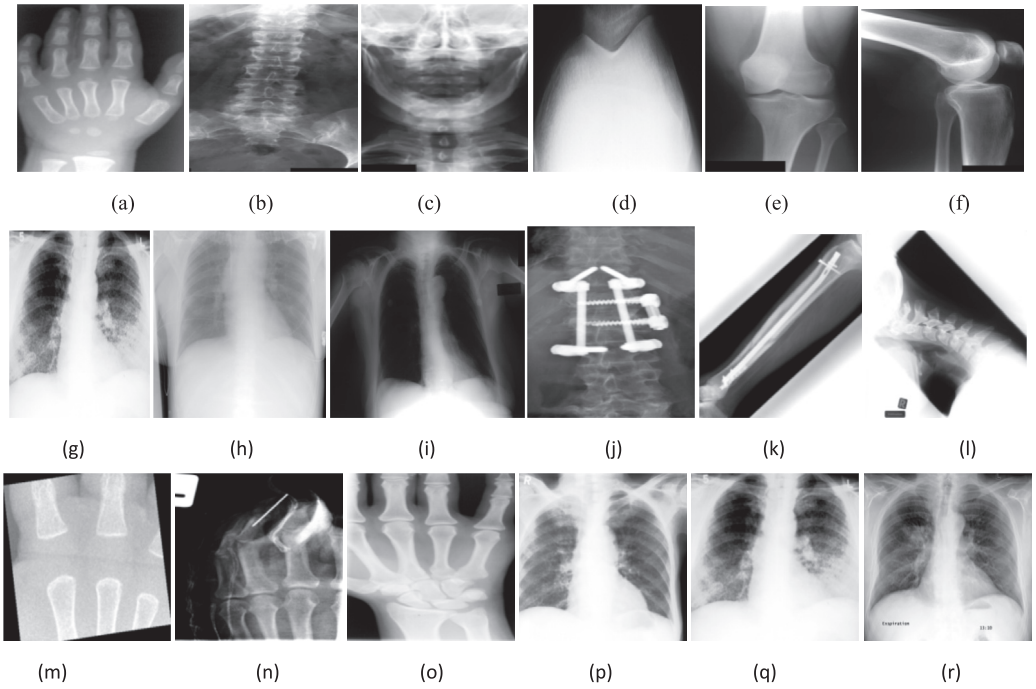


**Fig. 8.** Sample images from IRMA database:(a)-(f) Sample images from IRMA database consisting of different body parts; (g)–(i) Chest images with intensity and contrast variation; (j)–(l) Images with artefacts and image borders ; (m)–(o) Images with large intraclass variation; (p)–(r) Images with large interclass similarity.

Please cite this article in press as: Shamna, P., et al. Content-based medical image retrieval by spatial matching of visual words. Journal of King Saud University – Computer and Information Sciences (2018), https://doi.org/10.1016/j.jksuci.2018.10.002

(a) (b) (c) (d) (e)

**Fig. 9.** Sample images from Kvasir dataset consisting of endoscopic images (a) Polyp (b) Z-line (c) Dyed-lifted-polyp (d) Normal Pylorus (e) Cecum.

### 3.3. Evaluation matrices

Medical image retrieval systems are to be evaluated using rank-based measures, as doctors would probably be interested in analysing only the top results of images retrieved. We have evaluated the retrieval performance using Precision at L (P@L), MAP (Mean Average Precision) and Precision-Recall curve. For a retrieval system, precision and recall values are calculated based on the Eqs. (4) and (5). P@L represents the number of correctly retrieved images in the first L images extracted by the system.

$$Precision = \text{Number of relevant images retrieved} / \text{Total number of retrieved images} \quad (4)$$

$$Recall = \text{Number of relevant images retrieved} / \text{Total number of relevant images} \quad (5)$$

The precision and recall values can be represented graphically using the precision-recall curve, by computing them at each location in the ranked order of the images. Average precision (AverageP) was obtained by averaging the precisions at different recall levels. The MAP (Mean Average Precision) of the retrieval system was represented as the mean of the AverageP value of the total query images.

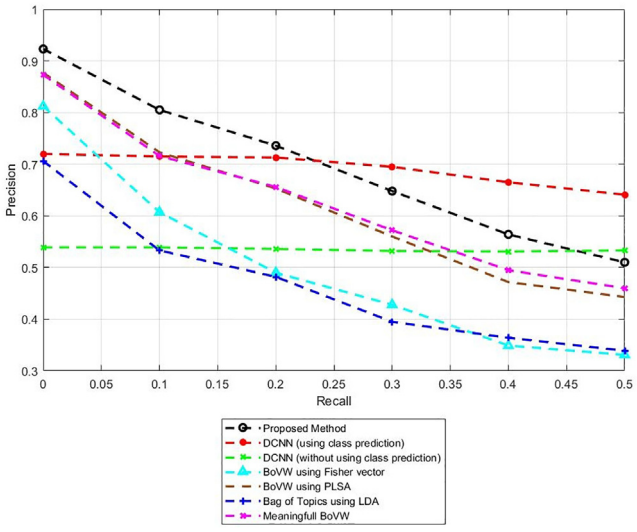


**Fig. 10.** Precision-Recall curve of Proposed method, DCNN (Qayyum et al., 2017), BoVW using Fisher vector (Suharjito and Santika, 2017), BoVW using PLSA (Cao and Cao, 2016), Bag of Topic using LDA (Li et al., 2015), Meaningful BoVW (Foncubierta-Rodríguez et al., 2013) using the Multimodal dataset.

### 3.4. Comparison with State-of-the-Art methods

We evaluated the proposed CBMIR system with supervised and unsupervised techniques in the recent literature. Since a standard medical dataset was not available for comparison, we evaluated the performance of the proposed system using the Multimodal dataset employed in Deep Convolutional Neural Network (DCNN) (Qayyum et al., 2017), IRMA 2009 dataset used by (Ahmad et al., 2018) (Srinivas et al., 2015) (Greenspan and Pinhas, 2007) and Kvasir dataset engaged in (Ahmad et al., 2018, 2017) framework.

Qayyum et al. proposed a CBMIR framework using a supervised DCNN (Qayyum et al., 2017). The DCNN was trained using randomly selected 5040 images from the Multimodal dataset and the retrieval performance was evaluated with the remaining 2160 images. Qayyum et al. evaluated the performance of retrieval system using two strategies: DCNN using class prediction, and DCNN- without using class prediction. We evaluated the proposed system with similar multimodal dataset, in which all 23 classes were same as in DCNN (Qayyum et al., 2017). Instead of the 24th class consisting of publicly inaccessible knee MRI images ("Osirix, Pixmeo SARL, Geneva," 2010) used in DCNN was replaced by the side chest X-ray images of IRMA dataset. We also evaluated the performance of the proposed method by implementing recent content-based retrieval methods using BoVW (Cao and Cao, 2016; Foncubierta-Rodríguez et al., 2013; Li et al., 2015; Suharjito and Santika, 2017). The results in Fig. 10 and Table 1 clearly reveal that the proposed method has high precision rates for the initial recall levels compared to the recent CBMIR systems in the literature.

We analysed the performance of the proposed system with recent CBMIR systems using deep learning methods (Ahmad et al., 2018, 2017; He et al., 2016; Krizhevsky et al., 2012; Qayyum et al., 2017). We evaluated the performance of the proposed system using the Multimodal dataset with fully trained (Qayyum et al., 2017) and transfer learned (He et al., 2016; Krizhevsky et al., 2012) deep learning networks. The IRMA-2009 dataset with 57 classes (Tommasi et al., 2009) is used to compare the retrieval performance of the proposed system with compact binary codes generated from convolutional features (Ahmad et al., 2018). Kvasir dataset containing coloured endoscopic images are used to evaluate the performance of the proposed system with deep convolutional features used in (Ahmad et al., 2018, 2017). The Precision@50 of Binary Codes Convolutional Feature is estimated approximately from the precision-recall curve given in (Ahmad et al., 2018). The performance of the proposed method is evaluated using randomly selected query images from the dataset. The overall results in Table 2 demonstrate that the proposed unsupervised method achieved better retrieval results than the deep learning methods used for comparison.

Srinivas et al. proposed a sparse unsupervised dictionary learning method (Srinivas et al., 2015) using 2600 images from eight classes of IRMA 2009 dataset. They incorporated the spatial information by dividing the image into small regions as shown in Fig. 11. The sparse dictionary learning method (Srinivas et al., 2015) achieved a P@10 of 97.14% using 14 randomly selected query images from the dataset. We evaluated the performance of the proposed system with the same dataset by implementing two unsupervised dictionary encoding techniques; the Saliency Coding

Please cite this article in press as: Shamna, P., et al. Content-based medical image retrieval by spatial matching of visual words. Journal of King Saud University – Computer and Information Sciences (2018), https://doi.org/10.1016/j.jksuci.2018.10.002

**Table 1**
Retrieval Results: P@L (for L = 21, 42, 63) and MAP value using the Multimodal dataset. Bold face indicates the best value in each category.

| Methods | P@21 | P@42 | P@63 | MAP |
|---|---|---|---|---|
| Proposed Method | **80.50%** | **73.60%** | 64.80% | **69.70%** |
| DeepCNN (using class prediction) Qayyum et al. (2017) | 71.50% | 71.3% | **69.50%** | 69% |
| DeepCNN (without using class Prediction) Qayyum et al. (2017) | 53.90% | 53.65 | 53.25 | 53.20% |
| BoVW using Fisher vector Suharjito and Santika (2017) | 60.70% | 49.00% | 42.80% | 50.27% |
| BoVW using PLSA Cao and Cao (2016) | 71.62% | 65.32% | 55.95% | 62.1% |
| Bag of Topics using LDA Li et al. (2015) | 53.27% | 48.28% | 39.54% | 47.00% |
| Meaningfull BoVW Foncubierta-Rodríguez et al. (2013) | 73.30% | 65.57% | 57.22% | 62.88% |

**Table 2**
Comparison of the proposed method with recent CBMIR system using deep learning methods. Bold face indicates the best value in each category (Randomly chosen query images are used for evaluation). *Not Specified: The numbers of query images or Classes are not specified.

| | #Images | #Query | #Class | Methods | Results |
|---|---|---|---|---|---|
| Dataset | Multimodal Dataset (Grey-Scale Images) | | | | |
| **I** | 7200 | 2160 | 24 | Proposed Method | **MAP = 69.70%** |
| | | | | DCNN Qayyum et al. (2017) | MAP = 69.00% |
| | | | | Residual net 18 layers He et al. (2016) | MAP = 51.03% |
| | | | | AlexNet(Krizhevsky et al., 2012) | MAP = 51.26% |
| Dataset | IRMA 2009 Dataset (Grey-Scale Images) | | | | |
| **II** | 14,410 | 100 | 57 | Proposed Method | **P@50 = 83.9%** |
| | 15,363 | *Not Specified | *Not Specified | Binary Codes- Convolutional Features Ahmad et al. (2018) | P@50 ≈ 77.00% |
| Dataset | Kvasir Dataset (Colour Images) | | | | |
| **III** | 4000 | 160 | 8 | Proposed Method | P@50 = 60.48% |
| | 4000 | *Not Specified | 8 | Binary Codes- Convolutional Features Ahmad et al. (2018) | P@50 ≈ 47% |
| | 4000 | 160 | 8 | Clustered Convolutional Features Ahmad et al. (2017) | **P@50 = 74.02%** |

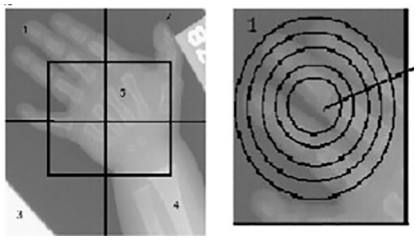

**Fig. 11.** Feature extraction techniques used in sparse dictionary learning (Srinivas et al., 2015) by dividing the image into small regions and partitioning the regions in to circles having same area.

(SC) proposed by Huang et al. (2011) and the Group Saliency Coding (GSC) proposed by Wu et al. (2012). We used two level spatial pyramids (Lazebnik and Schmid, 2006) to incorporate spatial information into the SC and GSC. Table 3 describes the results of the experiment. The proposed method achieved 100% precision in retrieving top 10 images.

To further analyse the performance of SC with pyramid (Huang et al., 2011), GSC with pyramid (Wu et al., 2012) with the proposed method, we extended the same experiment and evaluated the P@L (for L = 20,30,40) and the MAP value. The results in Fig. 12, Tables 3 and 4 illustrate that the proposed method incorporating rotation invariant spatial information without partitioning the images have better results compared to the partitioning methods used for comparison.

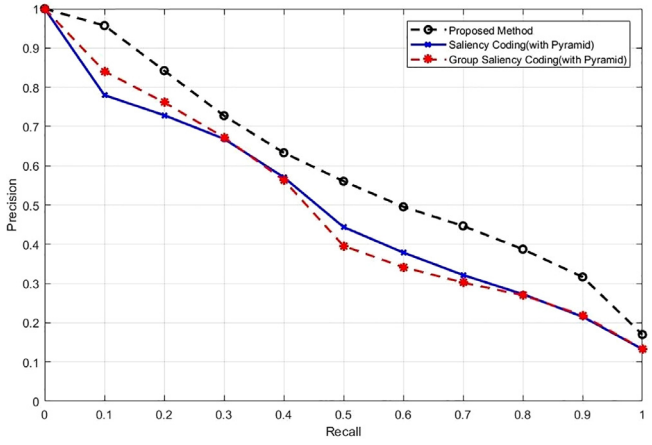


**Fig. 12.** Precision-Recall curve for the Proposed method, Group Saliency Coding with pyramid (Wu et al., 2012), Saliency Coding with pyramid (Huang et al., 2011) using using IRMA dataset. Bold face indicates the best value in each category.

**Table 3**
Precision at 10 for the proposed method, Sparse dictionary learning (Srinivas et al., 2015), Group Saliency Coding with Pyramid (Wu et al., 2012), Saliency Coding with Pyramid (Huang et al., 2011) using IRMA dataset. Bold face indicates the best value in each category.

| Method | P@10 |
|---|---|
| Proposed Method | **100%** |
| Sparse dictionary learning Srinivas et al. (2015) | 97.14% |
| GSC with Pyramid Wu et al. (2012) | 96.43% |
| SC with Pyramid Huang et al. (2011) | 90.71% |

Fig. 13 depicts the top 10 retrieved images for a query image. The retrieval results show that the top ten retrieved images by the proposed method given in Fig. 13(a) are structurally very similar to the given query image of the right knee. Whereas, SC and GSC retrieve images, given in Fig. 13(b) and (c), not only from right knee but also from left knee, skull and hand classes. The retrieval results confirm that the proposed location based encoding method efficiently captures anatomical similarity of medical images compared to the spatial pyramid saliency-based encoding techniques (Huang et al., 2011; Wu et al., 2012).

Please cite this article in press as: Shamna, P., et al. Content-based medical image retrieval by spatial matching of visual words. Journal of King Saud University – Computer and Information Sciences (2018), https://doi.org/10.1016/j.jksuci.2018.10.002

**Table 4**
P@L (for L = 20, 30, 40) and MAP of Proposed Method, Group Saliency Coding, Saliency Coding, using IRMA dataset. Bold face indicates the best value in each category.

| Method | P@20 | P@30 | P@40 | MAP |
|---|---|---|---|---|
| Proposed Method | **97.14%** | **95.43%** | **93.93%** | **59.42%** |
| GSC with Pyramid Wu et al. (2012) | 91.07% | 88.64% | 85.21% | 50% |
| SC with Pyramid Huang et al. (2011) | 84.29% | 82.36% | 79.93% | 50.12% |

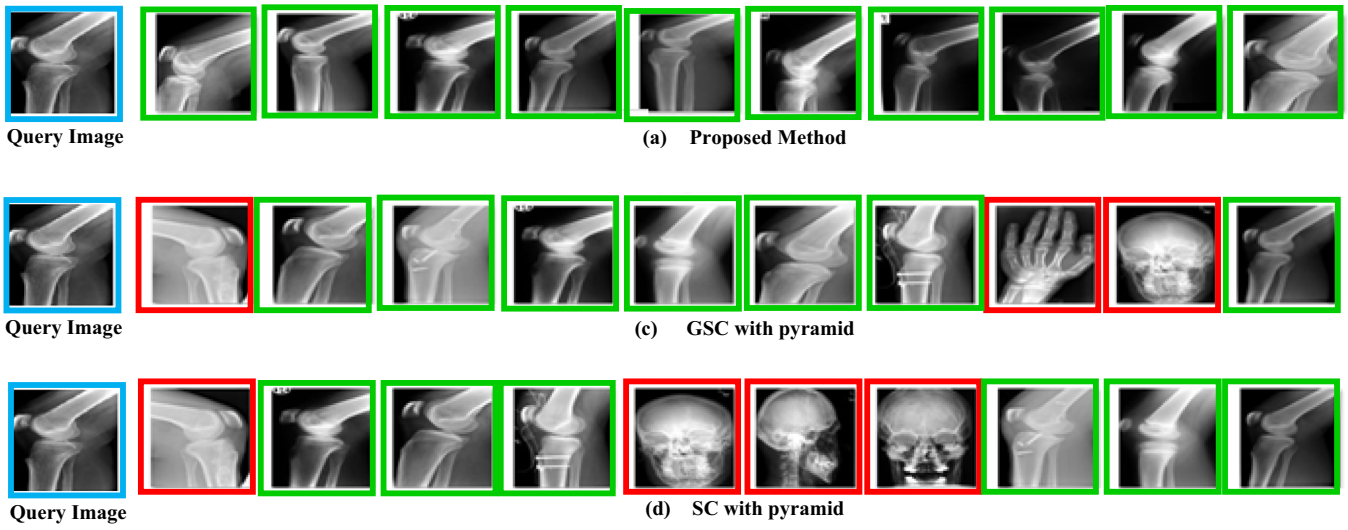


**Fig. 13.** Retrieval results of top 10 images for query image (right knee) given in blue bounday box, (a) First row represents the retrieval results of Proposed Method (b) Second row represents the retrieval results of GSC with pyramid (Wu et al., 2012) (c) Third row represents the retrieval results of SC with pyramid (Huang et al., 2011). Green boundary box represents the relevant images and red boundary box indicates the irrelevant images retrieved by the methods.

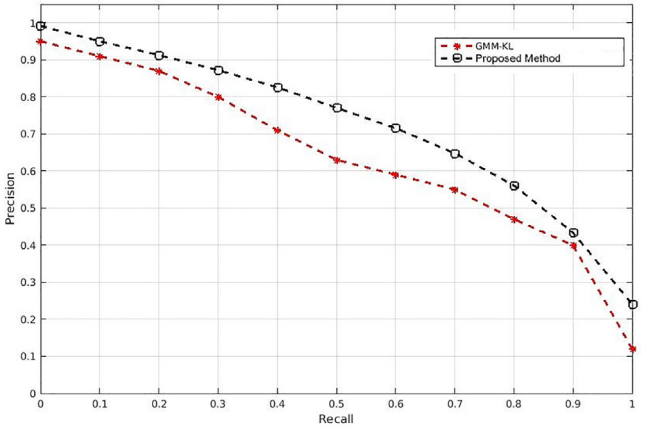


**Fig. 14.** Precision-recall curve for GMM-KL (Greenspan and Pinhas, 2007) and Proposed Method using IRMA dataset.

Gaussian Mixture Modelling with Kullback Leibler (GMM-KL) (Greenspan and Pinhas, 2007) framework used image features like intensity (I), texture contrast (C) and pixel position ((x,y) cordinates) to retrieve and classify images. GMM-KL incorporated the spatial information as spatial coordinates. We used the symmetric database used by the GMM-KL framework having 13 classes with 50 images in each class from IRMA 2009 dataset for comparison. The Precision-Recall curve in Fig. 14 shows better retrieval performance for the proposed method compared to GMM-KL. Table 5 provides a comparative analysis for the P@L (for L = 15, 30) and MAP values for the proposed and GMM-KL methods.

**Table 5**
P@L (for L = 15,30) and MAP of Proposed Method and GMM KL (Greenspan and Pinhas, 2007). Bold face indicates the best value in each category.

| Method | P @15 | P@30 | MAP |
|---|---|---|---|
| Proposed Method | **95%** | **91.2%** | **72.01%** |
| GMM KL Greenspan and Pinhas (2007) | 91% | 87% | 63% |

## 4. Discussion

We found that the location-based match of visual words improved the retrieval performance of medical images compared to the state-of-the-art CBIR systems (Ahmad et al., 2018; Cao and Cao, 2016; Foncubierta-Rodríguez et al., 2013; Greenspan and Pinhas, 2007; He et al., 2016; Huang et al., 2011; Krizhevsky et al., 2012; Li et al., 2015; Qayyum et al., 2017; Srinivas et al., 2015; Suharjito and Santika, 2017; Wu et al., 2012). The rotation invariant location features (angle and distance) efficiently recorded the spatial information of visual words compared to the spatial partitioning methods (Huang et al., 2011; Srinivas et al., 2015; Wu et al., 2012). The proposed Skip Similarity Index estimated the similarity between images by correlating the sequential ordering of the variable length location features in the images. The proposed method analysed the structure of medical images using Skip Similarity Index and retrieved images more accurately from an anatomically diverse dataset.

We compared the proposed method with the recent works in the literature using the publically available Multimodal dataset consisting of the different body parts. The proposed location based unsupervised technique performs better than the supervised DCNN (Qayyum et al., 2017) methods for initial recall level. The precisions values for the proposed method decreases for the higher recall

Please cite this article in press as: Shamna, P., et al. Content-based medical image retrieval by spatial matching of visual words. Journal of King Saud University – Computer and Information Sciences (2018), https://doi.org/10.1016/j.jksuci.2018.10.002

levels compared to the DCNN with class prediction. The retrieval performances of initial ranks are more critical in CBMIR compared to the later recall levels. The results in Table 1 and Fig. 10 indicate that the proposed method outperforms the topic modelling (Cao and Cao, 2016; Foncubierta-Rodríguez et al., 2013; Li et al., 2015) and Fisher vector based encoding (Sánchez et al., 2013; Suharjito and Santika, 2017) BoVW methods used in the recent CBIR systems. The spatial information of visual words effectively captures the anatomical structure of medical images. Moreover, the retrieval quality is also improved in terms of precision and MAP of query results, when compared to state-of-the-art methods (Cao and Cao, 2016; Foncubierta-Rodríguez et al., 2013; Li et al., 2015; Suharjito and Santika, 2017) based on the unordered BoVW.

The results in Table 2 demonstrate that the proposed method retrieved medical images more accurately compared to the state-of-the-art CBMIR systems using deep learning methodology (Ahmad et al., 2018; He et al., 2016; Krizhevsky et al., 2012; Qayyum et al., 2017). The experimental results with the Multimodal and IRMA 2009 dataset proved that the proposed method outperforms the deep learning methods in retrieving images from anatomically diverse dataset consisting of grey-coloured medical images. Fully trained DCNN based methods require an extensive collection of labelled medical images for training to achieve its full potential. Deep learning methods with transfer learning mechanism overcome this issue by fine-tuning the pre-trained DCNN (trained with natural images) with the medical images. We compared the performance of the proposed system with CBMIR methods using fully trained DCNN (Qayyum et al., 2017) and DCNN with transfer learning mechanism (Ahmad et al., 2018, 2017; He et al., 2016; Krizhevsky et al., 2012). The main focus of the proposed method is to retrieve the structurally similar medical images from repositories consisting of anatomically diverse multimodal grey-coloured medical images. Even though the Kvasir dataset consists of endoscopic coloured images of single modality, the proposed method using the location of grey-coloured visual words got a high precision value compared to the Binary Codes-Convolutional Features (Ahmad et al., 2018). The retrieval result of the proposed method is lower than Clustered Convolutional Features (Ahmad et al., 2017), which uses colour and texture features. The performance of the fully trained or transfer learned deep networks would heavily depend on the availability and quality of labelled medical data (Cho et al., 2015; Dodge and Karam, 2016; Tajbakhsh et al., 2016). The overall results in Table 2 demonstrate that the proposed method achieves better performance in medical image retrieval and does not require labelled data as in deep learning methods. As the availability of labelled data is expensive in the medical domain, the proposed unsupervised method is more suitable for CBMIR techniques than the supervised DCNN methods.

The performance of the proposed method was also evaluated with selected images from IRMA 2009 dataset. IRMA dataset have high inter-class and intra-class similarity. The images contain dominant artefacts like X-ray boundaries and artificial implants. The methods employed by (Srinivas et al., 2015) and (Greenspan and Pinhas, 2007) show that spatial information improved the retrieval results of IRMA dataset containing challenging images. We compared the performance of the proposed system with three recent unsupervised dictionary-based methods (Huang et al., 2011; Srinivas et al., 2015; Wu et al., 2012), in which the spatial information was incorporated by partitioning the images into different regions. The results in Tables 3 and 4 and Fig. 12 clearly suggest that the proposed method achieved high precision and MAP values compared to the unsupervised dictionary encoding methods using partition techniques (Huang et al., 2011; Srinivas et al., 2015; Wu et al., 2012). GMM-KL (Greenspan and Pinhas, 2007) framework included spatial coordinates to incorporate the spatial information to retrieve images from IRMA dataset. From the result shown in Table 5 and Fig. 14, we can infer that the rotation invariant location features of the proposed method performs better than the rotation variant spatial coordinate based retrieval system (Greenspan and Pinhas, 2007).

The major strength of the proposed method is the ability to correlate the spatial location of each visual word using the rotational invariant location features. Based on the observed experimental results, the proposed method demonstrates high retrieval accuracy for large, complex medical image dataset when compared to the existing CBMIR systems. The proposed system employs unsupervised clustering method to form dictionary without considering the number of classes or class label of images; which is an added advantage as the availability of labelled data is limited and expensive in the medical domain. The experimental result justify that the proposed method has better performance in retrieving multimodal medical images by body regions and with similar anatomical structures. Fig. 13 depicts the top 10 retrieved images of a sample query, which indicate that the proposed framework retrieves images efficiently with similar structure compared to the recent unsupervised dictionary based CBMIR techniques using the spatial pyramid. As the system compares the locations of visual words in the images, the computational complexity of the proposed method is high compared to the standard BoVW methods. However, this is not much of a problem, as advanced computational tools are now available. The enhanced retrieval performance of the proposed system outweighs the computational complexity limitations.

## 5. Conclusion and future work

Retrieving medical images from the anatomically diverse multimodal dataset is a challenging task. The proposed method employing spatial matching of visual words improved the retrieval performance of medical images consisting of different body parts. As medical images are mostly grey coloured, the count of visual words alone does not have the discriminative power to analyse the anatomical structure of the images. The recent retrieval systems employing BoVW methods neglect the significant spatial information of the visual words in the images. We captured the spatial information of visual words as distance and angle location features and the proposed Skip Similarity Index correlates the closeness in the location features to compute the structural similarity between images. The experimental results demonstrated that the proposed location-based method retrieved images more accurately, from anatomically diverse multimodal medical image datasets, compared to the state-of-the-art content-based image retrieval systems.

In the proposed work, we have considered all visual words with equal importance. The proposed method can be extended for complicated case-based retrieval of medical images by considering the weights of visual words according to their significance and location by incorporating the Region of Interest marked by medical experts (location of the abnormality, size of the abnormality etc). The proposed method searches images in the whole database to compute the similarity. The computation complexity can be improved by limiting the search domain and computing the location similarity with the significant visual words. In future, we anticipate developing a system that incorporates a better meaningful relationship between significant visual words for medical image retrieval and classification. Such a system will help the physician to compare case histories of patients with similar diseases and also help them to evaluate the disease stages in patients.

## 6. Declarations of interest

None.

Please cite this article in press as: Shamna, P., et al. Content-based medical image retrieval by spatial matching of visual words. Journal of King Saud University – Computer and Information Sciences (2018), https://doi.org/10.1016/j.jksuci.2018.10.002

## Appendix A

### 1. Experiments conducted for optimising system parametres

Comprehensive experiments are performed for optimizing the parameters encompassing the proposed framework. The experiments are conducted using randomly selected 1000 images, comprising of 10 classes (chest, chest side, hand, skull front, skull side, hip, spine, breast, knee, and patella) from the image database. We randomly selected 90 images from each of the classes to form the dictionary, and 10 images from each of the classes to test the retrieval performance. We focus on optimizing the following system parameters: (1) Finding the optimal Distance Threshold to initialise the cluster centres (K) in K-means, (2) Finding the optimal Deviation Threshold to determine the match between two location features.

We used Simple Cluster Seeking (SCS) method (Tou and Gonzalez, 1974) to initialise fixed cluster centres of K-means to get a consistent result in each run. SCS will ensure that cluster centres or visual words in the dictionary are parted from each other with a Distance Threshold to avoid synonymous visual words in the dictionary. The clusters are initialised by randomly selected Distance Thresholds and evaluated by Calinski Harabasz variance ratio criterion (Calinski and Harabasz, 1974). The Calinski-Harabasz index or variance ratio criterion (VRC) is determined as in Eq. (A1).

$$VRC_K = \frac{S_B}{S_W} \times \frac{(N-K)}{(K-1)} \tag{A1}$$

where, $S_B$ denotes the between-cluster variance, $S_W$ denotes the within-cluster variance, and N denotes total data points in the clusters, $K$ equals the total number of clusters. The $S_B$, the total between-cluster variance is computed as in Equation A(2).

$$S_B = \sum_{i=1}^{K} n_i EuclideanDistance(mean_i, mean) \tag{A2}$$

where K represents the number of clusters, $mean_i$ represents the centroid of clusteri, $n_i$ represents number of data points in cluster i, and *mean* is the overall mean of the sample data. The $S_W$, the overall within-cluster variance is defined as in Eq. (A3).

$$S_W = \sum_{i=1}^{K} \sum_{x \in c_i} Euclidean\ Disance(x, mean_i) \tag{A3}$$

where K represents the number of clusters, $c_i$ represents the $i^{th}$ cluster, $x$ represents a data point belongs to $c_i$, and $mean_i$ is the centroid of $c_i$. Well-partitioned clusters have a large $S_B$ and a small$S_W$. We can enhance the data partitioning by improving the $VRC_k$ ratio.

**Table A1**
Calinski Harabasz variance ratio criterion (higher is better). Bold face indicates the best value in each category.

| Distance Threshold | Number of clusters | Calinski Harabasz variance ratio criterion |
|---|---|---|
| **550** | **40** | **4.86E + 03** |
| 500 | 112 | 3.02E + 03 |
| 450 | 391 | 8.86E + 02 |

**Table A2**
Mean Average Precision based on Deviation Threshold. Bold face indicates the best value in each category.

| Deviation Threshold (%) | Mean Average Precision |
|---|---|
| 10 | 55.04% |
| **20** | **57.75%** |
| 30 | 56.72% |
| 40 | 55.60% |
| 50 | 54.41% |

**Table A3**
Result of retrieval using randomly selected 1000 images. Bold face indicates the best value in each category.

| | Bag of Visual Words Csurka et al. (2004) | Spatial Pyramid Lazebnik and Schmid (2006) | Proposed Method |
|---|---|---|---|
| P@10 | 60.20% | 63.0% | 79.8% |
| P@20 | 55.0% | 56.4% | 74.6% |
| P@30 | 51.3% | 51.9% | 69.7% |
| MAP | 43.21% | 43.66% | 57.75% |

### 2. Results

The Number of Clusters based on Distance Thresholds (Euclidean) 550, 500 and 450 are evaluated based on Calinski-Harabasz variance ratio. The optimal number of clusters, K is the result of clustering with the maximum Calinski-Harabasz variance ratio value. Based on the experimental results given in the Table A1 we fixed the Distance Threshold as 550 for the Simple Cluster Seeking Algorithm.

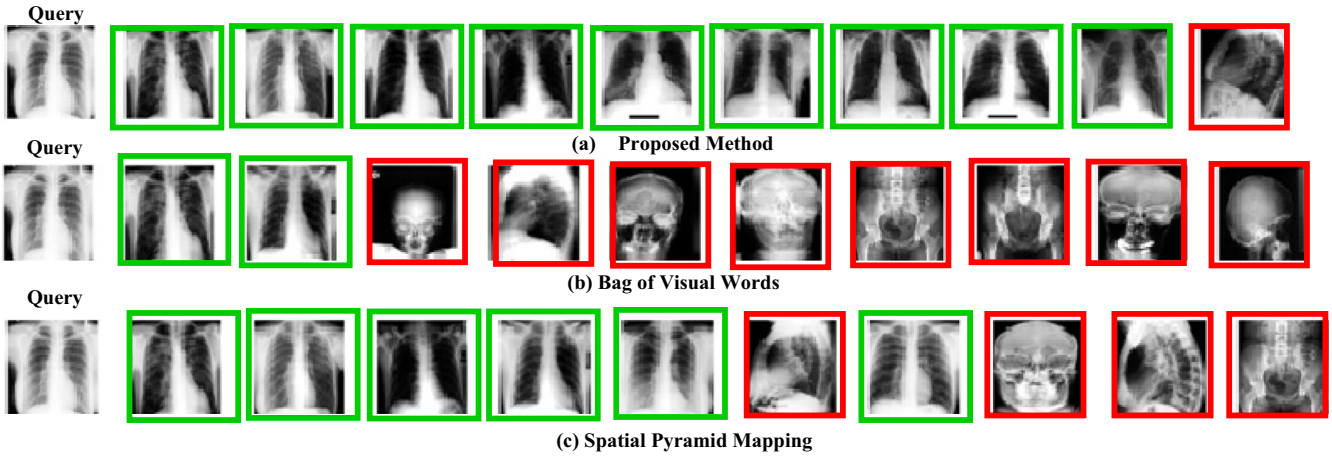


**Fig. A1.** Retrieval results of a sample query images, the first column represents query image and remaining columns represents the top ten retrieved images by (a) Proposed Method (b) Bag of Visual Words (Csurka et al., 2004) (c) Spatial Pyramid Matching (Lazebnik and Schmid, 2006) methods. Green boundary box represents the relevant images and red boundary box indicates the irrelevant images retrieved by the methods.

Please cite this article in press as: Shamna, P., et al. Content-based medical image retrieval by spatial matching of visual words. Journal of King Saud University – Computer and Information Sciences (2018), https://doi.org/10.1016/j.jksuci.2018.10.002

A match between the locations of a visual word in two images is determined based on the Deviation Threshold. Deviation Threshold decides the variance or difference allowed in the locations while estimating the match between two location features. The location feature of a visual word is represented as distance and angle of visual word in the image plane. We normalised the location features by dividing with the maximum distance from the image centre (based on the size of the image) and with max difference angle (180 degree) in the image plane. The locations are skipped if the normalised absolute difference between two location features is higher than the Deviation Threshold. We evaluated the Mean Average Precision (MAP) of the retrieval results with Deviation Threshold 10–50%. Based on the results given in Table A2 the Deviation Threshold is fixed as 20% for the experiments. The location features of visual words are matched if the normalised absolute difference between the locations is less than 20%.

After Optimising the parameters (Distance threshold as 550 and the Deviation Threshold as 20%), we evaluated the performance of the proposed system with Bag of Visual Words (Csurka et al., 2004) and Spatial Pyramid (Lazebnik and Schmid, 2006) mapping Method using 1000 images from ten classes. The results given in the Table A3 and Fig. A1 demonstrate that the proposed approach improved the retrieval performance due to the incorporation of location information.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## References

Ahmad, J., Muhammad, K., Baik, S.W., 2018. Medical image retrieval with compact binary codes generated in frequency domain using highly reactive convolutional features. J. Med. Syst. 42. https://doi.org/10.1007/s10916-017-0875-4.

Ahmad, J., Muhammad, K., Lee, M.Y., Baik, S.W., 2017. Endoscopic image classification and retrieval using clustered convolutional features. J. Med. Syst. 41. https://doi.org/10.1007/s10916-017-0836-y.

Avni, U., Greenspan, H., Konen, E., Sharon, M., Goldberger, J., 2011. X-ray categorization and retrieval on the organ and pathology level, using patch-based visual words. IEEE Trans. Med. Imag. 30, 733–746. https://doi.org/10.1109/TMI.2010.2095026.

Blei, D., Carin, L., Dunson, D., 2010. Probabilistic topic models. IEEE Signal Process. Mag. https://doi.org/10.1109/MSP.2010.938079.

Calinski, T., Harabasz, J., 1974. A dendrite method for cluster analysis. Commun. Stat. – Theory Methods 3, 1–27. https://doi.org/10.1080/03610927408827101.

Cao, C.H., Cao, H.L., 2016. The research on medical image classification algorithm based on PLSA-BOW model. Technol. Heal. Care 24, S665–S674. https://doi.org/10.3233/THC-161194.

Cao, Y., Steffey, S., Jianbiao, H., Xiao, D., Tao, C., Chen, P., Müller, H., 2015. Medical image retrieval: a multimodal approach. Cancer Inform. 13, 125–136. https://doi.org/10.4137/CIN.S14053.Received.

Cho, J., Lee, K., Shin, E., Choy, G., Do, S., 2015. How much data is needed to train a medical image deep learning system to achieve necessary high accuracy? arXiv Preprint arXiv:1511.06348, (2015).

Csurka, G., Dance, C., Fan, L., Willamowski, J., Bray, Cedric, 2004. Visual categorization with bag of keypoints. Int. Work. Stat. Learn. Comput. Vis. 1–22. https://doi.org/10.1234/12345678.

Decencière, E., Zhang, X., Cazuguel, G., Laÿ, B., Cochener, B., Trone, C., Gain, P., Ordóñez-Varela, J.R., Massin, P., Erginay, A., Charton, B., Klein, J.C., 2014. Feedback on a publicly distributed image database: The Messidor database. Image Anal. Stereol. 33, 231–234. https://doi.org/10.5566/ias.1155.

Diekhoff, T., Hermann, K.G., Pumberger, M., Hamm, B., Putzier, M., Fuchs, M., 2017. Dual-energy CT virtual non-calcium technique for detection of bone marrow edema in patients with vertebral fractures: a prospective feasibility study on a single- source volume CT scanner. Eur. J. Radiol. https://doi.org/10.1016/j.ejrad.2016.12.008.

Dodge, S., Karam, L., 2016. Understanding how Image Quality Affects Deep Neural Networks. Eighth International Conference on Quality of Multimedia Experience (QoMEX). 10.1109/QoMEX.2016.7498955.

Foncubierta-Rodríguez, A., de Herrera, A., Müller, H., 2013. Medical Image Retrieval using Bag of Meaningful Visual Words: Unsupervised Visual Vocabulary Pruning with PLSA. Proc. 1st ACM Int. Work. Multimed. Index. Inf. Retr. Healthc., pp. 75–82. 10.1145/2505323.2505336.

García Seco De Herrera, A., Markonis, D., Müller, H., 2013. Bag-of-colors for biomedical document image classification, in: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). https://doi.org/10.1007/978-3-642-36678-9_11

García Seco de Herrera, A., Müller, H., 2014. Fusion techniques in biomedical information retrieval. Inf. Fusion Comput. Vis. Concept Recognit. 209–228. https://doi.org/10.1007/978-3-319-05696-8_9.

Gherase, M.R., Feng, R., Fleming, D.E.B., 2017. Optimization of L-shell X-ray fluorescence detection of lead in bone phantoms using synchrotron radiation. X-Ray Spectrom. https://doi.org/10.1002/xrs.2792.

Greenspan, H., Pinhas, A.T., 2007. Medical image categorization and retrieval for PACS using the GMM-KL framework. IEEE Trans. Inf. Technol. Biomed. 11, 190–202. https://doi.org/10.1109/TITB.2006.874191.

Haas, S., Donner, R., Burner, A., Holzer, M., Langs, G., 2012. Superpixel-based interest points for effective bags of visual words medical image retrieval. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 7075 LNCS, 58–68. https://doi.org/10.1007/978-3-642-28460-1_6.

He, K., Zhang, X., Ren, S., Sun, J., 2016. In: in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. https://doi.org/10.1109/CVPR.2016.90.

Hofmann, T., 1999. Probabilistic Latent Semantic Analysis. Uncertainity Artifitial Intell., UAI'99

Huang, M., Yang, W., Wu, Y., Jiang, J., Gao, Y., Chen, Y., Feng, Q., Chen, W., Lu, Z., 2014. Content-based image retrieval using spatial layout information in brain tumor T1-weighted contrast-enhanced MR images. PLoS One 9, 1–13. https://doi.org/10.1371/journal.pone.0102754.

Huang, Y., Huang, K., Yu, Y., Tan, T., 2011. Salient coding for image classification. Proc. IEEE Comput. Soc. Conf. Comput. Vis Pattern Recognit., 1753–1760 https://doi.org/10.1109/CVPR.2011.5995682.

Hussain, M.A., Hodgson, A.J., Abugharbieh, R., 2017. Strain-initialized robust bone surface detection in 3-D ultrasound. Ultrasound Med. Biol. https://doi.org/10.1016/j.ultrasmedbio.2016.11.003.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks. Adv. Neural Inf. Process Syst. https://doi.org/10.1016/j.protcy.2014.09.007.

Lazebnik, S., Schmid, C., 2006. Beyond bags of features: spatial pyramid matching for recognizing. Natl. Scene Categories. https://doi.org/10.1109/CVPR.2006.68.

Lehmann, T.M., Güld, M.O., Thies, C., Fischer, B., Spitzer, K., Keysers, D., Ney, H., Kohnen, M., Schubert, H., Wein, B.B., 2004. Content-based image retrieval in medical applications. Methods Inf. Med. 43, 354–361. https://doi.org/10.1267/METH04040354.

Li, Z., Tian, W., Li, Y., Kuang, Z., Liu, Y., 2015. A more effective method for image representation: topic model based on latent dirichlet allocation. 14th Int Conf. Comput. Des. Comput. Graph, 143–148. 10.1109/CADGRAPHICS.2015.19.

Lord, M.L., McNeill, F.E., Gräfe, J.L., Galusha, A.L., Parsons, P.J., Noseworthy, M.D., Howard, L., Chettle, D.R., 2017. Confirming improved detection of gadolinium in bone using in vivo XRF. Radiat. Isot. Appl. https://doi.org/10.1016/j.apradiso.2016.12.011.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. 60, 91–110.

Markonis, D., Schaer, R., Müller, H., 2014. Multi-modal relevance feedback for medical image retrieval. Med. Inf. Retr. Work.

Miwa, S., Otsuka, T., 2017. Practical use of imaging technique for management of bone and soft tissue tumors. J. Orthop. Sci. https://doi.org/10.1016/j.jos.2017.01.006.

Muhammad, K., Sajjad, M., Lee, M.Y., Baik, S.W., 2017. Efficient visual attention driven framework for key frames extraction from hysteroscopy videos. Biomed. Signal Process. Control 33, 161–168. https://doi.org/10.1016/j.bspc.2016.11.011.

Mustra, M., Delac, K., Grgic, M., 2008. Overview of the DICOM standard. 50th Int Symp. ELMAR 1, 10–12.

Osirix,Pixmeo SARL,Geneva [WWW Document], 2010. URL http://www.osirix-viewer.com/resources/dicom-image-library/ (accessed 2.15.17).

Pogorelov, K., Randel, K.R., Griwodz, C., Eskeland, S.L., de Lange, T., Johansen, D., Spampinato, C., Dang-Nguyen, D.-T., Lux, M., Schmidt, P.T., Riegler, M., Halvorsen, P., 2017. Kvasir: A multi-class image dataset for computer aided gastrointestinal disease detection. Proceedings of the 8th ACM Multimedia Systems Conference, MMSys. 10.1145/3083187.3083212.

Qayyum, A., Muhammad, S., Awais, M., Majid, M., 2017. Neurocomputing Medical image retrieval using deep convolutional neural network. Neurocomputing 266, 8–20. https://doi.org/10.1016/j.neucom.2017.05.025.

Sánchez, J., Perronnin, F., Mensink, T., Verbeek, J., 2013. Image classification with the fisher vector: theory and practice. Int. J. Comput. Vis. 105, 222–245. https://doi.org/10.1007/s11263-013-0636-x.

Simpson, M.S., You, D., Rahman, M.M., Xue, Z., Demner-Fushman, D., Antani, S., Thoma, G., 2015. Literature-based biomedical image classification and retrieval. Comput. Med. Imaging Graph. 39, 3–13. https://doi.org/10.1016/j.compmedimag.2014.06.006.

Srinivas, M., Naidu, R.R., Sastry, C.S., Mohan, C.K., 2015. Content based medical image retrieval using dictionary learning. Neurocomputing 168, 880–895. https://doi.org/10.1016/j.neucom.2015.05.036.

Suharjito, Andy, Santika, D.D., 2017. Content based image retrieval using Bag of Visual Words and multiclass support vector machine. ICIC Express Lett, 11.

Swamy, M.S.M., Holi, M.S., 2013. Topic modeling for content based image retrieval. Lect. Notes Electr. Eng. 213, 321–332. https://doi.org/10.1007/978-81-322-1143-3.

Tajbakhsh, N., Shin, J.Y., Gurudu, S.R., Hurst, R.T., Kendall, C.B., Gotway, M.B., Liang, J., 2016. Convolutional Neural Networks for Medical Image Analysis: Full

Training or Fine Tuning? IEEE Trans. Med. Imaging. https://doi.org/10.1109/TMI.2016.2535302.

Cancer, The, 2014. Imaging Archive (TCIA) – A growing archive of medical images of cancer. [WWW Document].

Tommasi, T., Caputo, B., Welter, P., Güld, M.O., Deserno, T.M., 2009. In: Overview of the CLEF 2009 medical image annotation track. CEUR Workshop Proc., p. 1175.

Tou, J.T., Gonzalez, R.C., 1974. Pattern recognition principles. Image Rochester NY, 377. https://doi.org/10.1002/zamm.19770570626.

Villegas, M., Müller, H., Gilbert, A., Piras, L., Wang, J., Mikolajczyk, K., de Herrera, A.G. S., Bromuri, S., Amin, M.A., Mohammed, M.K., Acar, B., Uskudarli, S., Marvasti, N. B., Aldana, J.F., del Mar Roldán Garcia, M., 2015. General Overview of ImageCLEF at CLEF2015 Labs. Exp. IR Meets Multilinguality, Multimodality, Interact, 444–461.

Wang, S.H., Muhammad, K., Lv, Y., Sui, Y., Han, L., Zhang, Y.D., 2018. Identification of alcoholism based on wavelet renyi entropy and three-segment encoded jaya algorithm. Complexity 2018. https://doi.org/10.1155/2018/3198184.

Woźniak, M., Połap, D., 2018. Bio-inspired methods modeled for respiratory disease detection from medical images. Swarm Evol. Comput. https://doi.org/10.1016/j.swevo.2018.01.008.

Woźniak, M., Połap, D., Capizzi, G., Lo Sciuto, G., Kośmider, L., Frankiewicz, K., 2018a. Small lung nodules detection based on local variance analysis and probabilistic neural network. Comput. Methods Programs Biomed. https://doi.org/10.1016/j.cmpb.2018.04.025.

Woźniak, M., Połap, D., Kośmider, L., Cłapa, T., 2018b. Automated fluorescence microscopy image analysis of Pseudomonas aeruginosa bacteria in alive and dead stadium. Eng. Appl. Artif. Intell. https://doi.org/10.1016/j.engappai.2017.09.003.

Wu, Z., Huang, Y., Wang, L., Tan, T., 2012. Group encoding of local features in image classification. Proc. 21st Int Conf. Pattern Recognit, 1505–1508.

Xie, W., Wang, J., Cao, M., Hu, Z., Feng, Y., Chen, X., Jiang, N., Dai, J., Shi, Y., Babin, V., Mihóková, E., Nikl, M., Li, J., 2018. Fabrication and properties of Eu: Lu 2 O 3 transparent ceramics for X-ray radiation detectors. Opt. Mater. (Amst). https://doi.org/10.1016/j.optmat.2018.04.029.

Yang, W., Lu, Z., Yu, M., Huang, M., Feng, Q., Chen, W., 2012. Content-based retrieval of focal liver lesions using bagof-visual-words representations of single- and multiphase contrast-enhanced CT images. J. Digit. Imaging 25, 708–719. https://doi.org/10.1007/s10278-012-9495-1.

Please cite this article in press as: Shamna, P., et al. Content-based medical image retrieval by spatial matching of visual words. Journal of King Saud University – Computer and Information Sciences (2018), https://doi.org/10.1016/j.jksuci.2018.10.002