Journal of Engineering Science and Technology
Vol. 15, No. 1 (2020) 524 - 540

RETRIEVING MOBILE BASED SCALABLE IMAGES USING POSITION SCALE ORIENTATION-SCALE INVARIANT FEATURE TRANSFORM ALGORITHM

K. S. AISWARYA[1,*], N. SANTHI[2], K. RAMAR[3]

[1]Research Scholar, Department of Electronics and Communication Engineering,
Noorul Islam Centre for Higher Education,
Kumaracoil, Kanyakumari District- 629180, Tamilnadu, India
[2]Associate Professor, Department of Electronics and Communication Engineering,
Noorul Islam Centre for Higher Education,
Kumaracoil, Kanyakumari District- 629180, Tamilnadu, India
[3]Professor, Department of Computer Science Engineering, Einstein College of Engineering,
Tirunelveli-627012, Tamilnadu, India
*Corresponding Author: aiswaryachandrachood2000@gmail.com

Abstract

Smartphones are extensively used for capturing real-time images. These images can be shared very easily through any network. Multiple images can be captured to make sure the information is properly photographed. Scale Invariant Feature Transform and its variants are extensively used in image retrieval applications. Still, image retrieval can be done effectively in certain circumstances involving intensity, position and orientation mismatch with the help of position, scale and orientation - Scale Invariant Feature Transform, a powerful variant of Scale Invariant Feature Transform. In this paper, an efficient retrieval approach for mobile images using position, scale and orientation - Scale Invariant Feature Transform algorithm is proposed. The main aim is to find relevant information by exploring the keypoints from a group of images than using one input image. The proposed method then mines similar images based on the saliency from a group of images and then calculates the key features for image retrieval. The precision achieved by the above algorithm is better in terms of retrieved images has been confirmed from the test results.

Keywords: Feature extractions, Position scale orientation, Scale-invariant feature transform.

1. Introduction

In this modern era, mobile phones are extensively used. Statistics say there are billions of smartphones used in this world in recent years. Young people cannot avoid mobile phones. They become part of their life as it can be used for basic applications like the internet, booking tickets and sharing images with the beloved ones. Hence, retrieving of mobile images are very essential, especially for searching for a particular landmark or specific scenic spots. Unknown objects can be found out using this technique. When a person comes across an unknown object and he wants to explore it then he can take pictures of it and finally investigate the internet and can gather all information about it by seeing the similar objects.

The performance of retrieved images from mobile depends upon the channel, which is used for communication. Bandwidth occupancy and channel stability play an important role. Retrieving images from mobile is done using some text. Google-based searches bring a lot of matched images depending upon a textual based query from the user. Image retrieving process using text is proportional to the tags in images, subsequent labels and matched text. In certain cases, where images come without tags may face inaccuracy while retrieving images through text. Retrieving images based on contents is possible with cameras attached to mobile. Mobile-based search engine with visuals is achieved by researchers recently. Descriptors with more compactness [1, 2] or BoW compression-based histogram [3-6] are the recent state of the technique. To transmit in a limited bandwidth criterion, the histograms based on BoW is compressed. Some deficiencies are noted with BoW histograms, loss due to quantization is its first deficiency. The second deficiency is, during the comparison of visual words, the important relationship in spatial form of these words is neglected. The descriptive power of BoW representation is also poor. As there is a loss due to quantization, the selected visual word cannot exactly explain the region of an image locally.

The algorithm used for retrieving images normally has a single image as a query. The abovementioned algorithm provides poor performance if the query image cannot be clearly viewed. But if we use a SIFT algorithm there are lot many salient points (feature points) developed for a query image. Out of the above salient points, most of them are either noisy or very less relevant to the important object mentioned in the image. This leads to the complexities in the computation load. Also, the salient points, which are noisy will affect the retrieval performance of the algorithm. Latest algorithm using query expansion [7] techniques makes the noise weaker by simply combining the retrieved results with the query. Its main aim is to identify similar images to overcome the side effects of single query based retrieving techniques.

In General, we used to take different snaps of a scene to get minimum one photo, which satisfies our requirements. Retrieving techniques using mobiles, an individual takes a lot many photos before choosing one photo for the server to process the retrieving technique. From the multiple photos available, the salient points (context-based) are mined to get good retrieving performances and to use the bandwidth effectively. Thus, an algorithm is proposed based on visual search and scalable via finding context-based salient features using a lot many identical images.

The algorithm followed in this paper has 3 important steps:

- Mining of multiple images. The mobile phone of a user will have different sets of images or photos corresponding to the image, which is given for retrieval. The mobile also has certain other images, which are not relevant to the image submitted

for retrieval. Hence it is very essential to sort out or select different images that contain common characteristics as that of the query image. The extraction of Salient points from different relevant images is carried out by efficient and high-performance PSO SIFT algorithm.

- To find out the salient features or points from the group of images, which are very relevant. With the above-determined images, we again mine the salient points for retrieval as the major information is commonly repeated in all the above relevant images. The salient points calculated from the above relevant images are more significant, rigid and usually stable. Further to this, we can enforce geometric methods on the above salient points for getting good retrieval performance.
- Setting ranks for the salient points for improving retrieval performance. The contribution of salient features in this retrieval is found out. This is useful when the network usable bandwidth is varying with respect to time. A major contribution of the above salient points is very useful in determining the priority, in which, these features are to be transmitted. This gives us added advantage, when the network uses low bandwidth, as the high priority features can be transmitted.

The major work done in this paper is summarized below:

- Retrieving mobile images using a novel technique by extracting salient points in the relevant multiple images using PSO SIFT algorithm.
- As this algorithm employs salient feature extraction from many images, the extracted points are unique, stable as well as very robust compared to the extraction using a single image.
- Ranking of Salient features is also done as a part of extraction as this may help to finalize the priority in limited bandwidth transmissions.

The organization of the paper is as follows. Section 2 includes a review of the related work. Section 3 contains an overview of the proposed system in mobile image retrieval. Sections 4 and 5 depict the mining of appropriate images and exploration of contextual keypoints. Section 6 explains the comparing experiments and discussion about those parameters and conclusions are detailed in Section 7.

2. Related Work

Image Retrieving techniques using contextual saliency is gaining more popularity as it uses BoW formatting and extraction of its local features using powerful tools such as SIFT [8], SURF [9] and PCA SIFT [10] (a variant of SIFT). Image retrieval in large scales is also possible using vocabulary tree with hierarchical form.

If we consider the different visual words, the spatial information is not considered often but it has a lot of significance in the image retrieval process, which includes reranking and weighting salient features. The spatial verification process is usually done in the image retrieval process. Visual words in images usually contain a lot of spatial information [11-15]. Chen et al. [11] proposed a geometric model, in which, the orientation and scaling among visual words are described. Zhang et al. [12] introduced a feature group extraction from different images and among these groups, the geometrical contextual similarity is measured thereby calculating the better-matched value to find out group distance and ranking. Zhou et al. [16] coordinate based matrix, which is introduced to find the geometric relationship between the features. Zhang et al. [13] introduced the construction of image phrases and embedding it in the geometric

constraints for retrieving the image. Zhang et al. [17] generated descriptive visual phrases and words [17].

All related works mentioned above makes use of geometric spaces based on a query image. During the image retrieval process, a lot of features are extracted. Among these features, some are non-stable with noise background. Based on studies by Xue and Qian [18], Qian et al. [19] and Yang et al. [20], salient features or points, which are identical are used from the same spots in the images.

Identical Key Points (IKPs) are extracted from identical SIFT based points, which means these points repeatedly occurs in images. Basically, speaking these points determines the unique aspects of landmark images. Xue and Qian [18] described the Key Visual Words (KVWs) from an image group are extracted. KVWs are generated from IKPs, in addition to that IKP features are also correlated in the form of visual image words.

Chum et al. [7, 21] proposed multiple query-based techniques, which filter the queries by mixing the results obtained to form advanced query-based approach. Different results retrieved are summed up to yield better results. The precision can be improved by adopting ranking philosophy to the retrieved candidates [22]. Fernando and Tuytelaars [23] proposed the image retrieving techniques by using patterns derived from multiple query images.

These patterns are nothing but nearby words that repeatedly occur in these queries. The geometric characteristics are also explored in addition to similarity-based approaches. Fischler and Bolles [24] mentioned that a spatial approach is employed in contrast to Key Visual Pair (KVP) and Geometric Layout Descriptor (GLD).

Smartphones are widely used nowadays and retrieving mobile-based images calls everyone's attention. Work done in these areas can be categorized as the following:

Extraction of features using powerful descriptors such as PCA SIFT [10]. Chen et al. [3], Ji et al. [6, 25, 26] described compression of Bag of Word histograms. Lin et al. [5], Lowe [8] and Jie et al. [25-27] described the above technique, in which, is usually achieved by: (a) Correct Bag of Word is transmitted by avoiding the repeated points, (b) Compressing the vocabulary tree and (c) low dimensional matrix is employed by converting the high dimensional into a low by using the appropriate matrix for transformation. Latest techniques called sparse coding is employed in Lasso [28], which operates on BOW and making it in a form represented by certain elements in the dictionary.

3. Overview of Proposed System

Contextual keypoints based image retrieval techniques consists of the following steps:

- Mining of multiple images, exploring keypoints contextually and ranking of the obtained keypoints. In our approach, when a user produces a query image, the algorithm automatically searches for multiple images, which are more relevant in place of the user providing them.
- The next step is with the obtained multiple relevant images, Keypoints are calculated mainly to discriminate noisy one and to improve computation speed as well as precision. There is a lot of redundant information as multiple images are used. With the help of information based on contextuality, stable keypoints can be filtered out.

- Finally, these keypoints are ranked and sent to the wireless channel based on the priority. As the bandwidth of the channel is limited, only a few keypoints based on the priority, limited keypoints are transmitted at the server end for image retrieval.

4. Mining of Appropriate Images

It is understood that there may be many appropriate images corresponding to the input image, which was submitted by the user for retrieval purpose. Figure 1 represents the Input Query Image. The concept followed here is the mining of different images from the mobile of the user and then to understand keypoints contextually for performing image retrieval. Mining of such appropriate images straight away gives advantages corresponding to the performances of retrieved images in addition to this, the transmission bandwidth is also reduced as it limits the total number of features to be transmitted. The important steps followed are: (1) Extraction of features and processing the keypoint descriptors and (2) Mining of appropriate images.

Fig. 1. Input query image.

4.1. Extraction of features and processing the keypoint descriptors

Feature descriptors are extracted by using PSO-SIFT (Position Scale Orientation-Scale Invariant Feature Transform) [29] algorithm, which is the advanced and modified form of scale-invariant feature transform. SIFT [8] features are more robust against rotation and scale changes. Even though the extracted features using the SIFT algorithm are very robust in terms of changes in rotation and scale, their efficiency is not as expected when we experience an angle change of illumination change. In the modified SIFT (PSO SIFT) uses an advanced detection of gradient in the existing SIFT [30] algorithm, which increases the feature descriptors robustness when a variation in position occurs due to angle changes.

In general, the extractions of keypoints using SIFT is done based on the following steps:

- Detection of Scale-space extremes: Gaussian filtering using multiple standard deviations at various scale points is applied for generating a scale space. The next step is to compute the extrema locations with the help of gaussian differences among scale-spaces and to detect the appropriate key-points.

- Localization of keypoints and irrelevant points removal: The extrema points that were detected are mapped to appropriate pixel space. The next step is to identify the stable points, which is done by selecting the keypoints at the corners having sufficient contrast values for the feature vector.
- Assignment of Orientation: Evaluation of a reference orientation is carried out by calculating orientations and magnitude by comparison of neighbourhood patches among every keypoints.
- Keypoint descriptors: Image gradients are calculated at a selected scale around keypoints. Histogram from surrounding pixels (16 cells each cell has 4×4 pixels) among keypoint is generated depending on the magnitude and orientation angle. Histogram bins (8 bins) were updated using a weighted Gaussian orientation magnitude and this histogram with normalization is the keypoint feature descriptor.

In PSO-SIFT, instead of Gaussian weighted gradient, which is used as the last stage of the SIFT algorithm for updating the histogram bins a circular neighbourhood gradient with radius 12σ and 17 log-polar bins were employed for representing the feature. An orientation histogram of 8-bins is applied in PSO SIFT, and hence, the feature vector of 136-dimensional points will be generated.

Figure 2 represents the extracted PSO SIFT keypoints. Further, the hierarchically based quantization is applied to the output descriptors from the image. The quantization is done with an '*N*' code words based visual vocabulary.

Fig. 2. PSO SIFT keypoints.

4.2. Mining of appropriate images

The aim of mining of appropriate images is to group identical images in the mobile of a user. The photos available with the user may contain visual-based information, place and time information. For a time-range user may click different images of a particular place and even the geographical terrain-based information of above images may also available.

The above two information is very much useful in the process of image mining. In certain circumstances, if the GPS installed in mobiles face signal strength issues, it may not open and which, leads to non-availability of geographical information. In the above case, key point extraction based on the contextual property, which is explored.

In order to find the appropriate images, the basic approach followed is feature-based technique. Mining of appropriate images for the query-based image is done by the

following steps: (1) Identifying appropriate candidate images and (2) Eliminating the unwanted noises from the images.

4.2.1. Identifying appropriate candidate images

To obtain the relevant images we usually find the common features (similarity) among the input query image and the images available in the users mobile. The similarity is calculated based on a histogram approach. A retrieval technique or scheme using Bag of words is selected here and the query image is expanded. Instead of using an entire full image a portion of the query image is used in this approach as an overall search model. False positives in the search model are reduced by limiting the Bag of Words. The BoW histograms corresponding to the input query image denoted by h_{iq}, the histogram of a mobile image is $h_{mi}(n)$ and $\| \|$ represents the $L1$ norm. Here $L1$ normalization is used because it is the efficient and stable normalisation technique and it has only one solution. The common features (similarity) of the n^{th} image in users mobile to input query $S(n)$ is obtained using the distance estimation technique as given below.

$$S(n) = \exp\left(-\left\|h_{iq} - h_{mi}(n)\right\|\right) \tag{1}$$

where $\|dv\|$ denotes the normalization of vector dv and $\| \|$ gives the $L1$ norm, where $n = 1,\ldots\ldots T$, where T denotes the total images available in the users mobile.

4.2.2. Eliminating the unwanted noises from the images

Among the distance obtained by using the above formula, the grouping is done from the increasing order. The images in the mobile end are arranged in the sorted order of similarity score and the relevant image with high scores of similarities are used to create the multi-image query space. From the results, M-1 (M is the number of multiple relevant images) top-ranked are chosen as the candidates mined images.

Along with this ranked candidate images, there are certain noisy characteristics exists within them. Hence, these noises are to be removed as they affect the performance. We set a threshold of 0.95 to eliminate the candidates if its common features (similarity) are too small (to avoid noise) or high (to avoid duplication). Now the remaining candidates are the final appropriate relevant images, which are selected for finding out the salient features.

5. Exploration of Contextual Key Points

The stable and robust keypoints in the multiple images are explored after finding the appropriate relevant images. The keypoints are explored based on the following: (1) Contextuality based information is extracted from the appropriate multiple images and (2) Relationship among the keypoints geometrically.

5.1. Contextuality based information is extracted from the appropriate multiple images

The multiple relevant images have identical crucial information, which is relevant contextually. These contents occur many times in these images. In these multiple images, we have some crucial information, which often occurs than the others. Our aim is to identify the most frequently occurring visuals for the purpose of image retrieval. The keypoints are mined based on the following two steps:

- Identical key point detection (IKPs)
- Ranking of key visual words.

5.1.1. Identical key point detection

The similarity content is captured from the multiple images based on the matching pair [14]. The correct PSO SIFT points (v, r) and its corresponding matched values $MV_{(v,r)}$ are captured. The score of similarity is obtained by using the matched values given as follows:

$$MV_{(v,r)} = (v,r)/(|v| \cdot |r|) \tag{2}$$

where, v and r represent the PSO SIFT vector descriptor and $|q|$ represents the normalization vector q and "." means dot product.

Identical keypoints are calculated as per the matched values. Figure 3 represents the Identical Keypoints (IKPs). An IKP is a group of PSO SIFT matched points. It is denoted by:

$$IKP_n = \{Z_n{}^1, Z_n{}^2, \ldots, Z_n{}^i, \ldots, Z_n{}^M\} \tag{3}$$

where, IKP_nrepresents the nth IKP, M represents the total no. of relevant images. Zn^1 denotes the PSO SIFT descriptor of n^{th} IKP in i^{th} image.

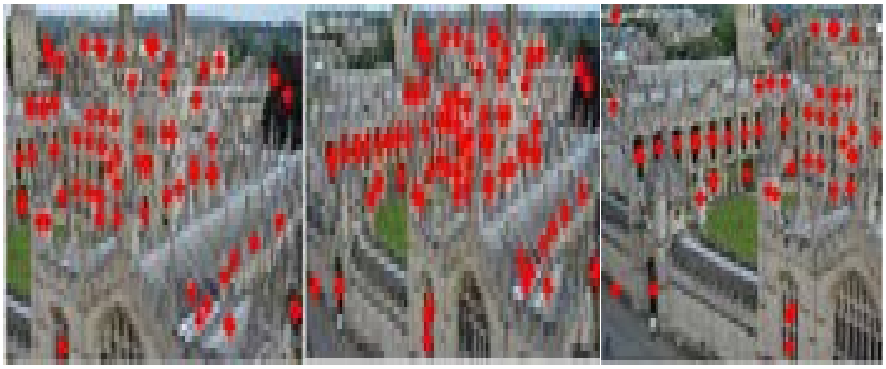

Fig. 3. Identical keypoints (IKPs).

5.1.2. Ranking of key visual words

The Identical Key Points (IKPs) are formed by finding identical elements in the spatial descriptors. Our concept is mainly based on the visuals in the images and it is well known that IKP will be different for different photos. The PSO-SIFT features corresponding to IKP should satisfy the consistency property among the visual words. IKP should be aligned to the corresponding visual word. The consistency of IKP is mentioned as:

$$Cs_l = \{Cs_l{}^1, \ldots\ldots\ldots Cs_l{}^i, \ldots\ldots\ldots Cs_l{}^M\} \tag{4}$$

where $Cs_l{}^i$ denotes the consistency corresponding to l^{th} IKP of i^{th} image. $Cs_l{}^i$= 1, if the n^{th} IKP is visible in the i^{th} image and Zn^i is counted to the same visual word as Zn^i, else $Cs_l{}^i$= 0. The importance of l^{th} IKP is computed with the score of consistency (SC) in different images. The score of consistency is represented as:

$$SC_l = \sum_{j=1}^{M} CS_l{}^j \tag{5}$$

Ranking of the score of consistency is done for the identical keypoints and based on the ranking top IKPs are selected. IKPs corresponding to the same frequency is given equal weights. Again, they are ranked according to their consistency or stability. Key Visual Words (KVW) are generated from the visual words of IKP, which are more consistent. Such IKPs are very few compared with the original IKPs. By the above steps, we understood that by finding the contextual saliency the effective bandwidth is properly utilized and very effective in the total computation cost. The Bandwidth is effectively saved because only certain features in the images are to be transmitted. In addition, the KVWs are consistent and most stable and it contains the crucial and very important information and it explores the major salient features, which can effectively reduce the noises from the images.

5.2. Relationship among the keypoints geometrically

The exploration of information based on contextual saliency in different images we can extract the key visual words. But these KVWs cannot occur for a region because of the unwanted noise and loss due to quantization. As a general concept, these KVWs formed from identical image points should remain the same. From this, if a corresponding visual word maps same content visually in two different images its nearby features can be given to identical Visual words. One can identify the similarity and relation among nearby neighbourhood points (geometric relation). This may help in distinguishing some features, which are mapped to identical visual words. We can find out the above geometrical relation by the following:

- Formation of Key Visual Pairs (KVPs) from Key Visual Words (KVWs), which are neighbours. This helps in strengthening the discriminative power among key visual words.
- The above formed KVPs can be described as a Descriptive Spatial Layout (DSL) to put the geometric relation on KVPs.
- Mixing the different KVPs that contain same KVWs in different images for utilizing the bandwidth effectively by reducing the data to be transmitted and by calculating the consistency of mixed KVPs.

5.2.1. Formation of KVPs

According to Qamra and Chang [31] and Luo et al. [32], the nearby features are used to form the KVPs. If we consider any IKP it is merged with the closest and the second closest neighbourhood IKPs thereby forming two KVPs. If we take any point (*Pn*), it is surrounded by *Pne*1 and *Pne*2 respectively. *Pne*1 and *Pne*2 correspond to the first neighbourhood and second neighbourhood *Pn*. If the nearest IKP, surrounding an IKP is variable in associated multiple images then the next nearest IKP surrounding the IKP is considered.

The above-mentioned condition is well illustrated in a certain image where the nearest IKP *Pne*1 is not visible when it is hidden by some other objects, which forces to take *Pne*2. With the help of spatial relationships, in which, the key visual word is formed and the generated KVWs may vary in multiple images. Figure 4 represents the Key Visual Words (KVWs). Similarly, KVPs formed in these images may also vary. To effectively utilize the bandwidth thereby reducing the size of the transmitted data we select the constant and stable KVPs from the identical KVPs:

$$\begin{cases} KVW_{m1}{}^{j} = KVW_{m1}{}^{i} \\ KVW_{m2}{}^{j} = KVW_{m2}{}^{i} \end{cases} \quad (6)$$

where, $KVW_{m1}j$ and $KVW_{m2}j$ correspond to the KVWs of IKPs.

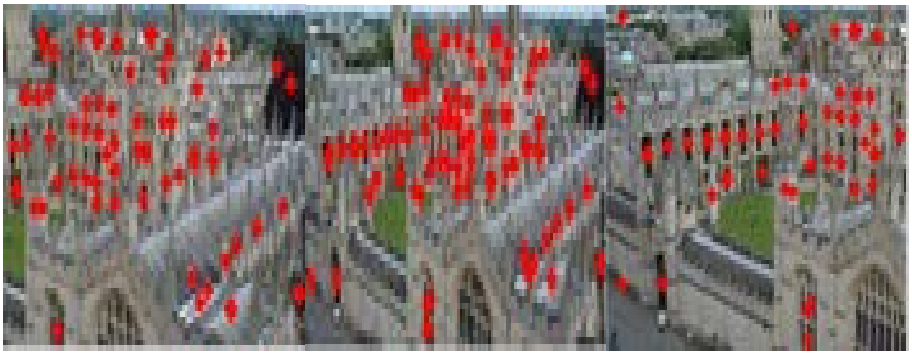

Fig. 4. Key visual words (KVWs).

5.2.2. Description of KVP

A pair of Key Visual Words forms a KVP. The similarities among the KVPs in the query image and the matched image is very difficult to identify if only Key Visual Pairs are used for image retrieval. KVPs in the two regions even though appear to be same but in actual they are different because of the spatial layout difference among two words visually for that region. Even though KVPs are generated from the identical words, again consider the similarity score based on the spatial layouts among the words for every image. Luo et al. [32] proposed that for every KVP calculate a descriptor *D* based on geometric layout, which measures the Euclidean distance and scaling known as (SED) content among the visual words.

$$SED_l{}^{i} = ED\left(KVW_{l1}{}^{i}, KVW_{l2}{}^{i}\right)/\left(s\left(KVW_{l1}{}^{i}\right) + s\left(KVW_{l2}{}^{i}\right)\right) \quad (7)$$

where, $SED_l{}^{i}$ indicates the SED of l^{th} KVP in the i^{th} image. $s\left(KVW_{l1}{}^{i}\right)$ and $s\left(KVW_{l2}{}^{i}\right)$ are the scale of $KVW_{l1}{}^{i}$ and $KVW_{l2}{}^{i}$. $ED\left(KVW_{l1}{}^{i}, KVW_{l2}{}^{i}\right)$ is the Euclidean distance among the key features assigned to $KVW_{l1}{}^{i}$ and $KVW_{l2}{}^{i}$ in the i^{th} a portion of a visual image.

$$d\left(KVW_{l1}{}^{i}, KVW_{l2}{}^{i}\right) = \sqrt{(U_{l1}{}^{i} - U_{l2}{}^{i})^2 + (V_{l1}{}^{i} - V_{l2}{}^{i})^2} \quad (8)$$

where, $U_{l1}{}^{i}$, $V_{l1}{}^{i}$ are the coordinate values of $KVW_{l1}{}^{i}$ in $\underline{i}^{th}$ visual image. KVP and descriptor *D* enhances the Key visual words altogether. KVP restricts the repetition on visual words and descriptor *D* indicates the spatial geometric similarity and relations among features locally.

5.2.3. Merging of different key visual pairs

Spatial geometric relationships exploration helps every visual image to be grouped as a combined series of KVPs and its paired descriptor as shown below:

$$I_m{}^{i} = \left\{\left(KVP_1{}^{i}, D_1{}^{i}\right), \ldots, \left(KVP_l{}^{i}, D_l{}^{i}\right), \ldots \left(KVP_L{}^{i}, D_L{}^{i}\right)\right\} \quad (9)$$

where, ${I_m}^i$ indicates the ith image. ${KVP_l}^i$ is the *l*th KVP in i^{th} image. ${D_l}^i$ is the *D* for ${KVP_l}^i$, *L* denotes the number of SVPs in the multiple-images.

The next step is to find the average Scaled Euclidean Distance (ASED). ASED is obtained by grouping the different KVPs given as follows:

$$ASED_l = \sum_{j=1}^{N} {SED^j}_l / N \tag{10}$$

where, *N* indicates the total number of multiple images.

There may be a chance of small difference among the geometric layout of KVPs. By considering this into account, the standard deviation of descriptors is calculated above in different images. This standard deviation is called weighted stability.

$$w_{SED_l} = \exp\left(-\sqrt{\sum_{j=1}^{N} ({SED^j}_l - AED_l)^2 / N}\right) \tag{11}$$

where, w_{SED_l} indicates the weighted stability in the SED of l^{th} KVP. Key Feature Group (KFG) is defined as a set of KVP, ASED and WSED, which is given as below:

$$KFG_l = \left(KVP_l,\ ASED_l,\ w_{SED_l}\right) \tag{12}$$

By considering different visual images it can be grouped as a series of KFGs as shown below.

$$FSG = \{KFG_1, \ldots, KFG_l, \ldots, KFG_n\} \tag{13}$$

where, FSG denotes the group finally sorted in the mobile. Based on the available bandwidth and the condition of the channel number of KFGs to be sent is finalized.

6. Key Features Ranking for Scalable Retrieval of Images

Interference is a common phenomenon occurring in the wireless communication channel. Also, the signals become very weak. It is advantageous to opt for scalable image retrieval depending upon the condition of the wireless channel. Ranking can be done on the KFGs based on their significance in retrieval. This helps us to transmit a few KFGs based on their significances, hence, large data volumes can be avoided. Image retrieval process employs the following steps:

- Ranking of KVGs.
- Searching for the similarities.

6.1. Ranking of KVGs

As the channel capacity is very small, it is preferred to transmit the KVGs based on their ranking. By sending the ranked KVGs, the latency is avoided. KVG is transmitted based on its significant contribution to the image retrieval process. Ranking of KVGs can be done by two methods: (1) frequency of its occurrence and (2) Consistency in multiple images.

6.1.1. Ranking depending on its frequency of occurrence

It is common to rank KFGs based on the number of occurrences (frequency). If KVP_i occurs in *n* photos and KVP_j occurs in *n*-1 photos, then KFG_i should get more priority than KFG_j for transmission over the channel.

6.1.2. Ranking depending on consistency in multiple images

Other than ranking by frequency KFGs can be further ranked by consistency factor called weighted stability. The KVPs whose stability weight is higher will be ranked higher. If $WSED_i > WSED_j$ the KFG_i gets priority than KFG_j for transmission through the channel. The final ranking of KFGs depends on the stability factor, hence, mobile can send information to the channel effectively based on the channel condition. When the bandwidth is limited, some KFGs, which are ranked higher are sending for image retrieval. Subsequently, during better channel condition, the next lower ranked KFGs can be transmitted.

6.2. Searching for similarities

Images represented in the dataset generally are a group of words. These images also undergo quantization and extraction of SIFT features. If there are *M* KFGs that are formed from different images in the mobile some of them will be chosen and transmitted from mobile set to the server for calculating the similarity. The similarity between images is calculated based on the following steps:

- Verification of KVP stability.
- Descriptor similarity.

The similarity score for the database image is given as follows:

$$\text{Similar}(Query, I) = \sum_{i=1}^{N} exp(-|SEDqueryi - SEDIi|) * WSEDqueryi \quad (14)$$

where *SEDqueryi* and *SEDIi* indicate the scaled Euclidean distances.

Figure 5 represents the retrieved images.

Fig. 5. Retrieved images.

7. Experiments

The evaluation of the proposed algorithm is done using different samples from the Oxford buildings dataset. Around 5000 flicker images of various landmarks were used in the dataset. 55 number of query images are used in our image retrieval algorithm. The proposed approach is compared with the other state of art techniques and its results are compared and shown below.

The results are simulated using MATLAB on workstation Intel® Core (TM) i7-7700HQ CPU @ 2.80 GHz and 16 GB memory. Different perspectives such as the number of relevant images in the database (*M*), the number of key feature groups used (*L*), feature point reduction. etc., are used for evaluating the algorithm. The evaluation metrics is done by calculating the precision and recall among the different algorithms as shown in Table 1.

Table 1. Comparison of precision of SIFT and PSO SIFT algorithm.

Parameters	Values	Precision in SIFT algorithm	Precision in PSO SIFT algorithm
Impact of *M*	2	0.4	0.42
KFG = 20	3	0.45	0.49
	4	0.48	0.48
Impact of *L*	10	0.45	0.51
(KFGs)	20	0.47	0.52
	40	0.5	0.54
	60	0.49	0.5
Single query	6 (visual pairs)	0.3	0.32
KFGs = 5	8	0.34	0.38
	10	0.35	0.4
Multiple	6 (visual pairs)	0.45	0.48
photo	8	0.46	0.51
KFGs = 5	10	0.46	0.46
Scale		0.4	0.46
Distance		0.34	0.31
SD		0.4	0.39
Orient		0.4	0.47
SD + orient		0.41	0.42

- **Precision**

The average value of precision at top *Ni* (Avg.Pr@Ni) is to calculate the mean percentage of the appropriate images in the *Ni* retrieved results. Here, *Ni* is taken as 5:

$$Avg.Pr@Ni = (1/T) \sum_{i=1}^{T} (Ai\ /Ni) \quad (15)$$

where *T* is the size of the test set.

Ai is the number of retrieved appropriate images up to *Ni* for the i^{th} input query image.

- **Recall**

Correct retrieval percentage from the overall results detected termed as recall.

$$Recall@Mi = \left(\frac{1}{T}\right) \sum_{i=1}^{T} \left(Ai \frac{}{Mi}\right) \quad (16)$$

where *Ai* is the number of retrieved appropriate images from the *Mi* results of the i^{th} input query image.

The experimental results of Input Query Image, PSO SIFT keypoints, Identical Keypoints (IKPs), Key Visual Words (KVWs) and retrieved images are shown in Figs. 1 to 5 respectively. The variation of recall for different values of a number of relevant multiple images are shown in Fig. 6.

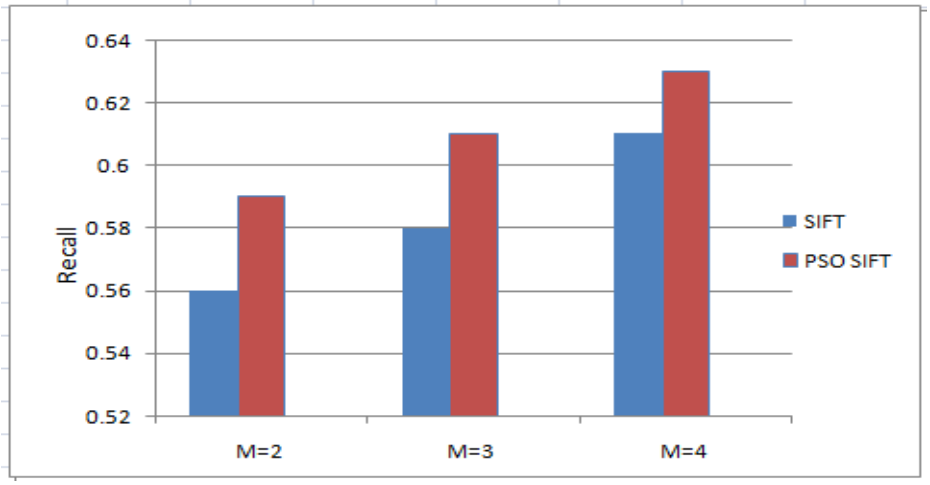

Fig. 6. Variation of recall for different values of *M* (number of relevant multiple images).

8. Conclusions

In this paper, the keypoints are explored by using a novel scalable mobile image retrieval technique using PSO SIFT algorithm. In this technique, image retrieval is rotation invariant and bandwidth-efficient. The precision in terms of retrieved images is also improved. This algorithm is not susceptible to the intensity and position variations. Experimental values confirm the improvement in precision with respect to retrieved images, robustness against orientation and position mismatch. The work can be further extended by reducing the complexities in calculating the key visual words and further improving the retrieval of images.

Nomenclatures

Ai	Number of retrieved appropriate images
ASED	Average SED of l^{th} KVP
Csl^i	Consistency of l^{th} IKP of i^{th} image
hiq	BoW histogram of input query image
hmi (n)	histogram of mobile image
IKP_n	n^{th} identical keypoint
KFG_l	Key feature group of l^{th} KVP
MV (*v*,*r*)	Matched values of the PSO SIFT vector descriptors *v* and *r*
S(n)	Similarity of n^{th} image
SC_l	Score of consistency of l^{th} IKP

SED_{li}	Scaled Euclidean distance of l^{th} KVP in i^{th} image
$WSED_l$	Weighted stability in the SED of l^{th} KVP
Zni	PSO SIFT descriptor of n^{th} IKP in i^{th} image

Abbreviations

BoW	Bag of Words
IKP	Identical Key Points
KVP	Key Visual Pair
KVW	Key Visual Words
PCA	Principal Component Analysis
PSO	Position Scale Orientation
SIFT	Scale Invariant Feature Transform
SURF	Speeded-Up Robust Features

References

1. Chandrasekhar, V.; Takacs, G.; Chen, D.; Tsai, S.; Grzeszczuk, R.; and Girod, B. (2009). CHoG: Compressed histogram of gradients - A low bit-rate feature descriptor. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* Miami, Florida, United States of America, 2504-2511.
2. Girod, B.; Chandrasekhar, V.; Chen, D.M.; Cheung, N.-M.; Grzeszczuk, R.; Rezni, Y.; Takacs, G.; Tsai, S.S.; and Vedhantham, R. (2011). Mobile visual search. *IEEE Signal Processing Magazine*, 28(4), 61-76.
3. Chen, D.M.; Tsai, S.S.; Chandrasekhar, V.; Takacs, G.; Singh, J.; and Girod, B.; (2009). Tree histogram coding for mobile image matching. *Proceedings of the Data Compression Conference.* Snowbird, Utah, 143-152.
4. Chen, J.; Duan, L.-Y.; Ji, R.; Luo, S.; and Gao, W. (2012). Pruning tree-structured vector quantizer towards low bit rate mobile visual search. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).* Kyoto, Japan, 965-968.
5. Lin, J.; Duan, L.-Y.; Chen, J.; Ji, R.; Luo, S.; and Gao, W. (2012). Learning multiple codebooks for low bit rate mobile visual search. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing.* Kyoto, Japan, 933-936.
6. Ji, R.; Duan, L.-Y.; Chen, J.; Yao, H.; Yuan, J.; Rui, Y.; and Gao, W. (2012). Location discriminative vocabulary coding for mobile landmark search. *International Journal Computer Vision*, 96(3), 290-314.
7. Chum, O.; Philbin, J.; Sivic, J.; Isard, M.; and Zisserman, A. (2007). Total recall: Automatic query expansion with a generative feature model for object retrieval. *Proceedings of the IEEE 11th International Conference on Computer Vision.* Rio de Janeiro, Brazil, 1-8.
8. Lowe, D.G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2), 91-110.
9. Bay, H.; Tuytelaars, T.; and Gool, L.V. (2006). SURF: Speeded up robust features. *Proceedings on 9th European Conference on Computer Vision*. Graz, Austria, 404-417.
10. Ke, Y.; and Sukthankar, R. (2004). PCA-SIFT: A more distinctive representation for local image descriptors. *Proceedings of the IEEE Computer Society Conference*

on Computer Vision and Pattern Recognition. Washington, D.C., United States of America, 506-513.

11. Chen, J.; Feng, B.; Zhu, L.; Ding, P.; and Xu, B. (2012). Effective near-duplicate image retrieval with image-specific visual phrase selection. *Proceedings of the 9th IEEE International Conference on Image Processing (ICIP).* Orlando, Florida, United States of America, 1909-1912.
12. Zhang, S.; Huang, Q.; Hua, G.; Jiang, S.; Gao, W.; and Tian, Q. (2010). Building contextual visual vocabulary for large-scale image applications. *Proceedings of the 18th ACM International Conference on Multimedia*. Firenze, Italy, 501-510.
13. Zhang, Y.; Jia, Z.; and Chen, T. (2011). Image retrieval with geometry-preserving visual phrases. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* Providence, Rhode Island, United States of America, 809-816.
14. Wu, Z.; Ke, Q.; Isard, M.; and Sun, J. (2009). Bundling features for large scale partial-duplicate web image search. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Miami, Florida, United States of America, 25-32.
15. Yuan, J.; Wu, Y.; and Yang, M. (2007). Discovery of collocation patterns: From visual words to visual phrases. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Minneapolis, Minnesota, United States of America, 1-8.
16. Zhou, W.; Lu, Y.; Li, H.; Song, Y.; and Tian, Q. (2010). Spatial coding for large scale partial-duplicate web image search. *Proceedings of the 18th International Conference on MultiMedia.* Firenzi, Italy, 511-520.
17. Zhang, S.; Tian, Q.; Hua, G.; Huang, Q.; and Li, S. (2009). Descriptive visual words and visual phrases for image applications. *Proceedings on 17th ACM International Conference on Multimedia*. Beijing, China, 75-84.
18. Xue, Y.; and Qian, X. (2012). Visual summarization of landmarks via viewpoint modelling. *Proceedings of the 19th IEEE International Conference on Image Processing.* Orlando, Florida, United States of America, 2873-2876.
19. Qian, X.; Xue, Y.; Yang, X.; Tang, Y.Y.; Hou, X.; and Mei, T. (2014). Landmark summarization with diverse viewpoints. *IEEE Transactions on Circuits and Systems for Video Technology*, 25(11).
20. Yang, X.; Qian, X.; and Mei, T. (2014). Learning salient visual word for scalable mobile image retrieval. *Pattern Recognition*, 48(10), 3093-3101.
21. Chum, O.; Mikulík, A.; Perdoch, M.; and Matas, J. (2011). Total recall II: Query expansion revisited. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Providence, Rhode Island, United States of America, 889-896.
22. Zhang, S.; Yang, M.; Cour, T.; Yu, K.; and Metaxas, D.N. (2015). Query specific rank fusion for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(4), 803-815.
23. Fernando, B.; and Tuytelaars, T. (2013). Mining multiple queries for image retrieval: On-the-fly learning of an object-specific mid-level representation. *Proceedings of the IEEE International Conference on Computer Vision* (ICCV). Sydney, New South Wales, Australia, 2544-2551.

24. Fischler, M.A.; and Bolles, R.C. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6), 381-395.
25. Ji, R.; Duan, L.-Y.; Chen, J.; and Gao, W. (2012). Towards compact topical descriptors. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Providence, Rhode Island, United States of America, 2925-2932.
26. Ji, R.; Duan, L.-Y.; Chen, J.; Yao, H.; and Gao, W. (2011). A low bit rate vocabulary coding scheme for mobile landmark search. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*. Prague, Czech Republic, 4 pages.
27. Ji, R.; Yao, H.; Liu, W.; Sun, X.; and Tian, Q. (2012). Task-dependent visual codebook compression. *IEEE Transactions on Image Processing*, 21(4), 2282-2293.Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of Royal Statistical Society, Series. B (Methodological)*, 58(1), 267-288.
28. Ma, W.; Wen, Z.; Wu, Y.; Jiao, L.; Gong, M.; Zheng, Y.; and Liu, L. (2017). Remote sensing image registration with modified SIFT and enhanced feature matching. *IEEE Geoscience and Remote Sensing Letters*, 14(1), 3-7.Yang, X.; Qian, X.; and Xue, Y. (2015). Scalable mobile image retrieval by exploring contextual saliency. *IEEE Transactions on Image Processing*, 24(6), 1709-1721.Qamra, A.; and Chang, E.Y. (2008). Scalable landmark recognition using EXTENT. *Multimedia Tools and Applications*, 38(2), 187-208.
29. Luo, Q.; Zhang, S.; Huang, T.; Gao, W.; and Tian, Q. (2013). Scalable mobile search with binary phrase. *Proceedings on 5th International Conference on Internet Multimedia Computing and Service (ICIMCS)*. Huangshan, China, 66-70.