Multimed Tools Appl (2015) 74:655–669
DOI 10.1007/s11042-014-2008-0

Mining near duplicate image groups

Jing Li · Xueming Qian · Qing Li · Yisi Zhao · Liejun Wang · Yuan Yan Tang

Received: 29 October 2013 / Revised: 16 March 2014 / Accepted: 4 April 2014 / Published online: 2 May 2014
© Springer Science+Business Media New York 2014

Abstract Most recently the social media sharing websites such as Flickr, Facebook, and Picasa have allowed users to share their personal photos with friends. Moreover, people like to follow, forward their favorite images, which is one of the main source of near duplicate images. And also, the worldwide place of interests such as Roma, Statue of Liberty and London Tower Bridge etc., attract world-wide visitors. For these places, travelers take photos, write travelogues and share them with their social friends. The photos taken from the same place with or without viewpoint variations are near duplicate images. How to detect them is an ad-hoc problem in the area of image analysis and multimedia processing. The existing near duplicate image processing approaches mainly focused on finding the near duplicate images for a given input image, where a query image is needed. However, how to find the near duplicate image groups (NDIG) automatically from the web-scale social images is very challenging. So, in this paper, instead of searching near duplicates image for certain input image, we proposed an automatic NDIG mining approach by utilizing adaptive global feature clustering and local feature refinement. The proposed NDIG mining approach is achieved by utilizing a hierarchical model. It is a two-layer hierarchical structure by first utilizing adaptive global feature clustering based candidate NDIG detection and then using local feature refinement based NDIG verification. The global clustering is mainly for reducing computational cost for processing the large scale image set. The local refinement is for improving NDIG detection performances. Experiments on four datasets show the effectiveness of our approach.

Keywords Near duplicate image group · Social media · Image retrieval

This work is supported in part by National Natural Science Foundation of China (NSFC) Project No.61173109 and No.60903121, No.61261036, and Microsoft Research Asia.

J. Li · X. Qian (✉) · Q. Li · Y. Zhao
SMILES LAB at School of Electronics and Information Engineering, Xi'an Jiaotong University,
Xi'an 710049, China
e-mail: qianxm@mail.xjtu.edu.cn

L. Wang (✉)
Xinjiang University, Urumqi, China
e-mail: wljxju@xju.edu.cn

Y. Y. Tang
Macau University, Macau, China

1 Introduction

Social networks are very popular for young people to acquire information, especially with the development of internet and smartphone. People tend to share, forward and follow what they are interested in. Flickr is one of the favorite image sharing websites. The total number of images is more than 6 billion. Facebook has gathered about one billion users, and about 0.25 billion images are uploaded per day. How to manage the big image data is very challenge for effective indexing and retrieval. We find that most of the places of interests have large amount of images shared by world-wide users. Meanwhile, some of images are forwarded, modified and copied by other users before being shared in social communities. Accordingly, there exists large number of near duplicate images in the website. Correspondingly, mining near duplicate image groups automatically from shared social media in sharing websites is very useful. For example, if images from the same places can be grouped together and remove near duplicated images, then we can make the image retrieval results more diversified. In the other hand, intelligent protection can also be achieved, since illegal copy could be detected during the process.

To our knowledge, few attentions have been paid on automatically mining NDIG in social media websites [14, 22]. Most of recent near duplicated image detection approaches need a query image, and then image retrieval are carried out [7, 16, 1, 24, 25, 20, 17, 26, 23, 19, 9, 10]. Some approaches are aiming at detecting NDIG directly by utilizing graph theory [14]. Philbin model images as nodes of the graph and image-to-image similarity as edge between the corresponding nodes. With the modeled graph, clustering based approach is adopted to divide the graph into smaller groups containing near duplicate image groups. However, in graph based NDIG detection [14], the weights of edges are merely measured by the concurrence of visual words which neglects the context information between images. Moreover, the image-to-image similarity computing is too time-consuming for a large scale dataset.

There are two main challenges in near duplicate image group mining from a web-scale image set: 1) computational cost of choosing near duplicated image groups from a very large scale image set, and 2) near duplicate image, different from exact duplicate images, is with various time, viewpoint, illumination and resolution are belong to one same near duplicated group. It is much more complicate than the duplicated image detection.

In this paper, we propose a NDIG mining approach by utilizing hierarchical model, which is both time efficient and effective. It is a two-layer hierarchical structure. by first utilizing adaptive global feature clustering based candidate NDIG detection and then using local feature refinement based NDIG verification. The global clustering is mainly for reducing computational cost for processing the large scale image set. The local refinement is for improving NDIG detection performances.

The main contributions of this paper are as follows: 1) we propose an effective and efficient two layer hierarchical system utilizing both global and local feature to mine NDIG, 2) we propose an adaptive clustering method to automatically find out the number of NDIG, and 3) we use local feature refinement to guarantee that the NDIG really contains near duplicate images.

The rest of this paper is organized as follows. The related works on near duplicate image detection is reviewed in Section 2. In Section 3, the whole system of finding out near duplicate image groups is given. The result of experiment is provided in Section 4. We conclude the paper and discuss future work on Section 5.

2 Related work

A direct ways for NDIG detection is using each image in the dataset as a query to accomplish image retrieval and then detecting near duplicated image groups utilizing the image retrieval results. Actually, this kind of NDIG consists of two steps: near duplicated image retrieval and NDIG detection. Near duplicate image retrieval is to find near duplicate images for a given query image [7, 16, 1, 24, 25, 20, 17, 26, 23, 19, 9, 10]. However compared with retrieval, NDIG detection is more complicated as it not only has to compare similarity for all the potential image pair in the dataset and has to judge whether they construct near duplicate [14, 22]. Hereinafter we give a brief overview of existing works on near duplicated image retrieval and NDIG detection.

2.1 Near duplicate image retrieval

Recently, many researches have paid their attention on near duplicated image retrieval, BoW and Hash are the tools usually utilized. Even though, BoW has shown its efficiency in both image retrieval and image classification, the negligence of spatial context between visual words makes it less reliable. There are many researches aiming at improving the BoW by considering spatial context or building Bag of Phrases. Hu et al. proposed a coherent phrase model for image near-duplicate retrieval [7]. Different from the standard BoW, their model represents every local region using multiple descriptors and enforces the coherency across multiple descriptors for every local region. Feature coherent phrase and spatial coherent phrase are designed to represent feature and spatial coherency. They mentioned that near duplicate image retrieval approach was hard to achieve the task of near duplicate image groups detection [7].

Gao et.al simultaneously utilize both visual and textual information to estimate images' relevance which is determined with a hypergraph learning approach [4]. In addition, they propose an interactive 3-D object retrieval scheme [3]. They incrementally select query views in each round of relevance feedback. They learn a distance metric for the newly selected query view and the weights for combining all of the selected query views. Wang et.al obtaine relevant and diverse images by exploring image content and the associated tags [21]. They utilize a greedy ordering algorithm which optimizes average diverse precision as the ranking method.

Han et al. proposed a framework of image retrieval with manifold learning [6]. The method of Local Regression and Global Alignment has been adopted to learn a robust Laplacian matrix for data ranking for the sake of image classification. In addition, considering that visual attributes can be considered as a middle-level semantic cue, they developed a method in which a well-defined set of attributes from auxiliary images to a target image is utilized, thus assisting in predicting appropriate attributes for the target image [5]. Sayad et al. proposed a higher-level image representation [16], which is a semantically significant visual glossary (SSVG). They introduce a two layer model in order to select Semantically Significant Visual Words (SSVWs) from the classical visual words and then exploit the spatial co-occurrence information of the SSVWs and their semantic coherency to generate Semantically Significant Visual Phrases and at last combine the two representation methods to form a SSVG representation. However, the process of model to build SSVG needs the knowledge about which images construct near duplicate beforehand.

Battiato et al. also aimed at improving coherent phrase model (Bags of Phrases) for near duplicate image retrieval [1]. They augment the original paradigm exploiting coherence between different feature spaces during the codebook generation step. This is achieved through alignment of the feature space partitions which are obtained from independent clustering.

There are also many works on utilizing local features in the near duplicate image retrieval. In [24], a multilevel spatial matching framework with two stage matching is proposed by Xu et al. to deal with spatial shifts and scale variations for image-based near duplicate identification. Although multi spatial matching is effective, when it comes to the large scale image dataset, the computing cost seems too high. They also utilized the multi-level matching algorithms on the near duplicate video retrieval [25].

There are also many researches aiming at decreasing the time cost of near duplicate image retrieval by using hash code [20, 17]. In [20], Wang et al. first calculate a K-bit ($K<32$) hash code for each image and conduct the duplicate image detection with only the hash codes. Because the hash codes are very compact representation of the image content, the detection process is very fast. However, for the situation of near duplicate images, the hash code changes a lot even there is only a slight change between the images. Song et al. proposed a new multiple features hashing (MFH) method for large scale near duplicate video retrieval [17]. Instead of using single feature, MFH employs a machine learning approach to exploit the local structure of each individual feature and fuse multiple features in a joint framework. However, it needs to train the videos and to have some near duplicate videos beforehand.

In [26], Zhou et al. proposed a scheme of spatial coding for large scale partial-duplicate image search. The spatial coding encodes the relative spatial locations among features in an image and discovers false feature matches between images. As for partial-duplicate image retrieval, spatial coding achieves even better performance. However, it works well on partial-duplicate image retrieval where the partial duplicate images usually contain some part of duplicate image in some part of the image. The method seems weak in the situation of near duplicate in actual world.

Nowadays, there are many applications of duplicate or near duplicate image retrieval [23, 19, 9]. Wu et al. [23] and Wang et al. [19] concentrate on automated image tagging and annotation applications. Lee et al. [9] detected taboo image in Tattoo Image Database. Near duplicate image retrieval is also utilized in social media processing [10].

Wang et al. proposed a novel near duplicated image based image annotation [19]. Their motivations are from the fact that tags for the duplicated images are similar. Thus finding near duplicated images for an image needed to be annotated is the key step of their approach. They utilize image signature and compacted global features in their duplicated image searching system. The method shows its time efficiency and robustness in duplicates detection in a very large scale image sets with billions of photos. In [19], PCA and hashing based near duplicated image detection approach is proposed. The approach aims at reducing the dimensions of the feature for saving computational cost.

2.2 Near duplicate image group detection

Due to the complexity of near duplicate detection compared with retrieval, few works has been done on the detection. In [14], a system is provided which mines near duplicate image groups by building a graph with images as nodes and local feature similarity as edges. Clustering is then done in order to find groups of near duplicate images. The local feature similarity between two images is measured by the percentage of number of SIFT in the sum of the SIFT number in the two images. Although it shows its efficiency, the need of computing local feature similarity between every image pair in a large dataset seems too time consuming. In another work, Wang et al. utilizes a combination of local and global features for duplicated image group detection [22]. However the method confines to judge whether two images are near duplicate or not instead of mining potential near duplicate image groups.

In our work, instead of detecting near duplicate image pairs, our system is to find the potential near duplicate image groups automatically only through utilizing the visual information of images without any other clues such as tags or GPS information.

3 Near duplicate image groups mining

Figure 1 shows the illustrations of our NDIG detection approach. Each image is denoted as a circle dot, and the line between them indicates the visual similarity. Intuitively, the images in the same group have high similarities than the images in different groups. Thus, we use solid line to link images with large visual similarity and virtual line to represent images in different groups with weak similarity. As shown in Fig. 1, images from *Oxford* (group A), *Tiantan of Beijing* (group B), *Eiffel Tower* (group C) and *Elysee Palace* (group D) are saliently different from the images taken from each other. Images in the same group have same content while different groups with significant global appearance variations.

Thus, we propose a global feature clustering and local feature refinement based NDIG detection approach. The global feature clustering is to find out the near duplicate image groups as completely as possible. This might make some near duplicated image groups divided into sub-groups. The local feature refinement can accomplish to merge the sub-groups. And at the same time it can be served as a verification step to make sure the images intra a candidate group actually consists of near duplicate images.

The block diagram of our NDIG mining approach is shown in Fig. 2, which consists the following three parts: feature extraction, adaptive clustering and the local refinement. The adaptive clustering has three advantages:1) divide images into candidate groups which have similar global appearances, 2) reduce computation cost by confining the time consuming procedure of local feature match into a smaller scope, and 3) automatically find out how many near duplicate groups in the dataset and also each groups contain how many images.

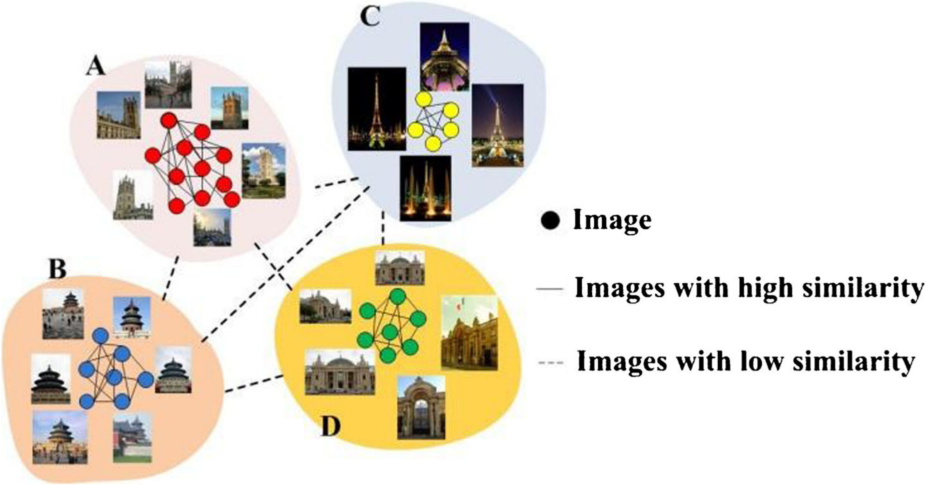


Fig. 1 Illustration of near duplicate image groups detection with four different groups *A*, *B*, *C*, and *D*, which are from *Oxford*, *Tiantan of Beijing*, *Eiffel Tower* and *Elysee Palace*

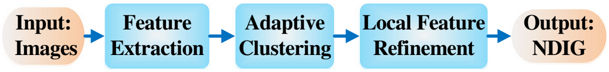


Fig. 2 Block diagram of near duplicate image groups mining. A system contains adaptive clustering to find out the potential near duplicate image groups, and local feature refinement to make sure the selected groups are true near duplicates

3.1 Feature extraction

Considering their relatively low dimension and well descriptive ability, global features are utilized first in our method. In this paper, 45-D color moment (CM) [18] and 170-D hierarchical wavelet packet descriptor (HWVP) [15] are utilized as the global features. And we also extract Scale Invariant Feature Transform (SIFT) [13] for all the images. Here we utilize the same features as in [11, 12] both the global feature and the local feature.

3.2 Adaptive clustering

As for a given unlabeled image dataset, it is difficult to estimate whether it contains NDIG and how many NDIG it contains. It is also hard to set a definite number to accomplish the clustering. Instead of setting the cluster number, we propose an adaptive clustering based method. The goal of adaptive clustering is two-fold: decreasing computational cost and refining performance. The detailed flowchart of adaptive clustering is shown in Fig. 3. It consists of the following two steps: initial K-means clustering and further K-means clustering.

Fig. 3 Flowchart of adaptive clustering. K is the number in the K-means clustering; H is the total number of clusters obtained after adaptive clustering

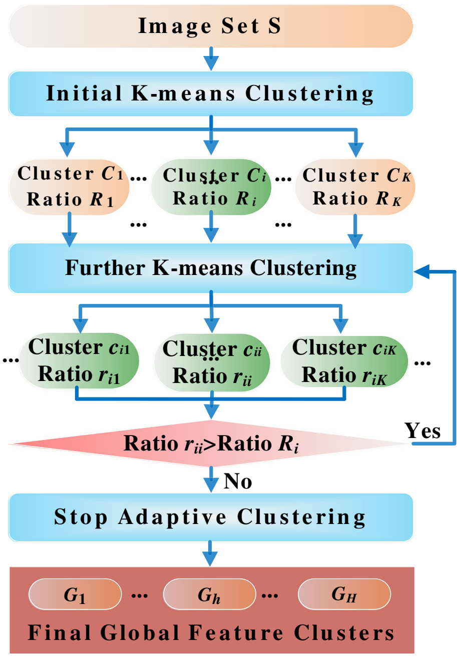


Image Set S
Initial K-means Clustering
Cluster C_1 Ratio R_1
...
Cluster C_i Ratio R_i
...
Cluster C_K Ratio R_K
Further K-means Clustering
Cluster c_{i1} Ratio r_{i1}
Cluster c_{ii} Ratio r_{ii}
Cluster c_{iK} Ratio r_{iK}
Ratio r_{ii}>Ratio R_i
Yes
No
Stop Adaptive Clustering
G_1
G_h
G_H
Final Global Feature Clusters

- Initial K-means clustering

  The global feature clustering is carried out on the 215 dimensional vector including 45d color moment and 170d hierarchical wavelet packet. The global features of all the images in the dataset are grouped into K centroids (denoted as $C_1,\cdots C_K$) using k-means. Each centroid $C_i(i=1,\ldots,K)$ is a 215d vector. Let N_i denotes the total images in the cluster C_i. The total number of images in image set is $N=\sum_{i=1}^{K}N_i$. Let L_x denote the 215d global features of the x-th image in the center C_i. In this step, the distance d_x between the x-th image's global feature and the center C_i is computed as follows

$$d_x = \|L_x - C_i\|, (x = 1, \ldots, N_i) \tag{1}$$

  where ||*|| denotes the norm of *. For each cluster C_i, we can obtain the corresponding maximum distance $\max\{d_x\}$ and minimum distance $\min\{d_x\}$. Then the ratio for C_i can be computed as follows

$$R_i = \max\{d_x\}/\min\{d_x\} \tag{2}$$

  The ratio R_i can be utilized as an indicator of the coherency of the cluster. From which we can determine whether carry out further clustering.

  Obviously, the lower the ratio, the relatively closer the cluster gets. So the ratio is utilized as the basis to judge whether to stop the clustering. If the ratio increases after clustering, the C_i is refreshed by the newly gained cluster and further k-means clustering is needed, nevertheless if the ratio decreases, there will be no further clustering.

- Further K-means clustering

  After the global feature clustering, we get K centroids. Each centroid is a 215d vector. We carry out an adaptive hierarchical clustering for each cluster C_i. So the centroid C_i is departed into K sub-centroids $c_{i1},\ldots,c_{iK}$ after hierarchical clustering. Each centroid $c_{ij}(j=1,\ldots,\mathrm{K})$ is also a 215d vector. Let r_{ij} $(j=1,\ldots,\mathrm{K})$ denote the coherency of c_{ij} which is computed by Eq. (2). Whether the centroid is undergoing a next layer clustering is determined adaptive by comparing the average coherencies of the K sub-centroids as follows

$$B = \begin{cases} 1, & \text{if } r_{ij} > R_i \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

  If the ratio $r_{ij} > R_i$ after further clustering, this shows that coherency of images in this centroid is not large content, the further k-means clustering is carried out, meanwhile if the ratio decreases, we end the further clustering step.

  After the adaptive clustering, H candidate groups $\{G_1, \ldots, G_H\}$ are obtained for the whole image dataset.

3.3 Local feature refinement

To guarantee that the candidate group G_h $(h=1,..,H)$ actually contains near duplicate image, local feature refinement is utilized here. In this process, the image content overlapping is considered to judge whether two images in each group to be near duplicates.

In this step, SIFT match is utilized to determine NDIG for each candidate group G_h after adaptive clustering. The method in [8] is adopted here to decide whether two images match or not. If two images have sufficient matched SIFT point pairs [8], they are considered a match. Otherwise they are not match. However, the feature matching is comparatively computationally intensive. So, here we utilize an inverted file structure based approach [13]. First all the SIFT features are quantized into Q centroids. Instead of pairwise matching between each SIFT point from two images, all the SIFT points are quantized into one of the centroids. Then the number of matched SIFT points between to image is computed by counting the corresponding indexed centroids [13]. Based on this, the final NDIG $\{\Omega_1,\ldots,\Omega_l\}$ for the images in G_h $(h=1,\ldots,H)$ is determined iteratively. The details are shown in **Algorithm1**. Finally the total NDIG for the H centroids are obtained and denoted as $\{\Omega_1,\ldots,\Omega_L\}$, $L>l$.

Algorithm 1: Finding Near Duplicate Image Group

```
Input:
    All the images in $G_h$ denoted $D$
Initial:
    One-to-one matching for images in $D$;
    Remove images with none matched image from $D$;
    Determine matched image number for each image in D by matching SIFT point number.
    $A \leftarrow$ the image with most matched images in $D$;
    Update: $l \leftarrow 1$, $\Omega_l \leftarrow A$, $D \leftarrow D-A$
    Label image A as representative image of this NDIG.
    Determine NDIG for image $A$ iteratively as follows.
while $D$ is not null, Do
    $A \leftarrow$ the image with most matched images in $D$;
    $P_A \leftarrow$ SIFT point number of image $A$;
    for $k=1:l$
        Count the average number of matched SIFT point $n_k$ between $A$ and all images in $\Omega_k$;
    end
    $P_* \leftarrow \max\{n_1,\cdots,n_l\}$, $* \leftarrow \arg\max_k\{n_k\}$
    if $P_* > P_A/2$
        image $A$ be a near duplicate images of $\Omega_*$,
        update: $\Omega_* \leftarrow \Omega_* + A$; $D \leftarrow D-A$
    Else
        assign a new NDIG for image A
        Label image A as representative image of this NDIG.
        update: $l \leftarrow l+1$; $\Omega_l \leftarrow A$; $D \leftarrow D-A$
end
Output: final NDIG $\{\Omega_1,\ldots,\Omega_l\}$ for images in $G_h$
```

By utilizing the method of clustering, it is possible that actually near duplicate images are divided into different NDIGs. Thus after NDIG detection, we merge the result $\{\Omega_1,\ldots,\Omega_L\}$. At this step, only the representative images (as labeled in **Algorithm 1**) from each group are utilized. If the two representative images A_i and A_j from two groups Ω_i and Ω_j are with high matching score, then we merge the corresponding two groups. Let the SIFT points of A_i and A_j are P_i and P_j, and we use M_{ij} denote their match SIFT point number. Correspondingly, Ω_i and Ω_j are merged or not is determined as follows:

$$Merge(\Omega_i, \Omega_j) = \begin{cases} 1, & \text{if } M_{ij} \geq P_i/2 \& M_{ij}/ \geq P_j/2 \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

We repeat the merging process until of all the NDIG (including the merged NDIG groups) is checked. The remained groups after merging are the final output of determined NDIG.

4 Experiments and discussions

In order to show the effectiveness of the proposed adaptive clustering and local feature refinement (denoted ACLR) based NDIG method, we make comparisons with hash code grouping (denoted HC) [19], and graph cutting (denoted GC) [14] based near duplicate image group detection methods. The near duplicate image groups detection in the HC is implemented by clustering the hash codes. Experiments are conducted on the four datasets: **COREL5k** [2], **OxBuild5k** [2], **GOLD** [11, 12] and **GOLDEN** [12]. All experiments are implemented on a server with 2.0 GHz CPU and 24 GB memory, and all the experiments are performed under the environment of C.

4.1 Experiment setup

- Datasets

  Near duplicate image groups of each test dataset are manually labeled, where 10 volunteers are involved. The manually labeled near duplicated image groups are utilized for testing the performances ACLR, HC and GC.

  COREL5k [2] is with 5,000 images and the near duplicate image group number is 50.
  OxBuild5k [2] is with 5,000 images and the near duplicate image group number is 51.
  GOLD is an image set containing about 230,000 images taken from 80 famous travel sites crawled from Flickr [11, 4]. The near duplicated image group number is 494.
  GOLDEN is extended dataset from **GOLD** crawled from Flickr containing about 5,200,000 images taken from 1,447 different places all over the world [4]. The places are selected by referring to the landmark and landscape list from WIKI.com. There are **2,184** near duplicated image groups.

- Evaluation Criteria

  In this paper, we use the following three criteria to measure the performances of the near duplicated image group detection. They are precision (PR), recall (RC) and F-Measure, which are expressed as follows

$$PR = AS/AC \times 100\% \tag{5}$$

$$RC = AS/TC \times 100\% \tag{6}$$

$$F-measure = \frac{2 \times PR \times RC}{PR + RC} \tag{7}$$

where AC is the number of detected NDIG, AS is the number of correctly detected NDIG, and TC is the number of NDIG in ground truth.

4.2 Comparisons

The values of PR, RC, and F-measure of the HC, GC, and ACLR, and their computation costs on the four test datasets are evaluated. For HC, we followed the parameter setting in [19] to

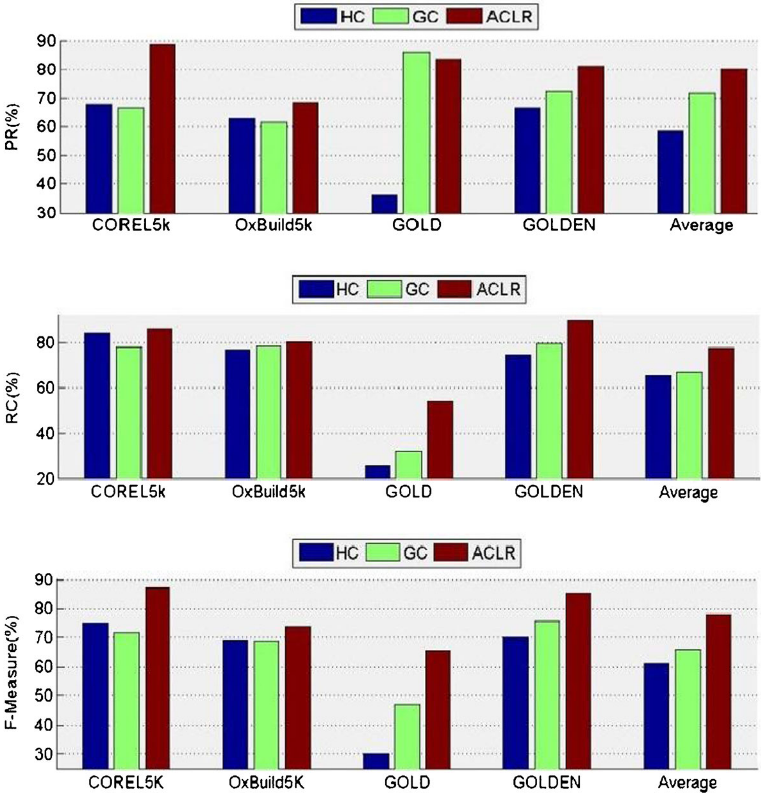


Fig. 4 Comparison of PR(%), RC(%) and F-Measure (%) for HC, GC, and ACLR on COREL5k, OxBuild5k, GOLD and GOLDEN

carry out image representation and hash code generation. For ACLR, we set the K in the k-means as 2. Further the parameter is also discussed in Section 4.3.

Figure 4 shows the performance of our proposed method ACLR, HC, and GC. From Fig. 4, the result shows that our method outperforms all the other methods in both precision and recall. As for PR, ACLR achieves 88.58 %, 68.30 %, 83.44 % and 80.97 % in COREL5k, OxBuild5k, GOLD and GOLDEDN respectively, while the HC is only 67.74 %, 62.90 %, 36.21 %, and 66.63 % on the four datasets and GC is 66.67 %, 61.54 %, 85.98 % and 72.28 %. For the precision on GOLD, the best the performance is obtained by the method of GC. Analyzing the reason, it is caused by the fact that ACLR finds out more than the number of ground truth which lead to a low precision. However, on the other datasets, ACLR outperforms

Table 1 Comparison of Time cost (s) for HC, GC, and ACLR on CORE5k, OxBuild5k, GOLD and GOLDEN

	HC	GC	ACLR
COREL5K	1,053 s (17.55 min)	39,212 s (10.89 h)	2,184 s (36.4 min)
OxBuild5K	2,276 s (37.93 min)	42,013 s (11.67 h)	3,012 s (50.2 min)
GOLD	8,904 s (2.47 h)	273,920 s (3.17 days)	12,038 s (3.34 h)
GOLDEN	101,026 s (1.17 day)	10 days	132,785 s (1.54 day)

Table 2 PR, RC and F-measure of ACLR or like using adaptive clustering vs. setting cluster number on GOLD

	K = *H*	K = T	*K* = 100	*K* = 500	ACLR
PR(%)	43.97	38.28	40.13	46.27	86.00
RC(%)	39.74	32.84	37.81	36.59	80.39
F-Measure(%)	41.75	35.35	38.94	40.86	83.10

all the other methods. The similar result is also illustrated in the recall. ACLR outperforms all the other methods in all the four datasets. It achieves 86.00 %, 80.39 %, 54.05 %, and 89.90 % in the four dataset respectively. As for the time cost, it also shows its time efficiency. It needs only 36**min** to complete NDIG detection on COREL5k, 50**min** for OxBuild5k, 3.3 h for GOLD and 1.5 days for GOLDEDN.

4.3 Adaptive clustering vs setting cluster number

Here, we discuss the performances of adaptive clustering and local refinement vs clustering the whole dataset directly by setting the final cluster number as a constant then further using local refinement to detect NDIG. This discussion is on GOLD. Several comparisons are made for the following five cases: 1) K = H, directly setting K to be *H* (number of clusters obtained in the adaptive clustering); 2) K = T, T is the ground-truth NDIG number obtained for each dataset set, for GOLD, *T*=494; 3) *K*=100; 4) *K*=500; and 5) adaptive clustering with *K*=2. The corresponding PR, RC and F-measure are given in Table 1, Table 2 and shown in Fig. 5. We can see that ACLR outperforms all the situations of K set to be a constant. When K is set to be *H*, the PR, RC and F-Measure are 43.97 %, 39.74 and 41.75 % respectively. When K is set to be the number of ground-truth, the result is the worst, which is 38.28 %, 32.84 % and 35.34 % separately for PR, RC and F-Measure. From the result we can see that, instead of setting K to be a constant, the adaptive clustering is more effective in NDIG detection.

5 Conclusion

In this paper, we propose an algorithm to automatically mine NDIG for any given unlabeled image dataset. We first cluster the dataset utilizing an adaptive clustering on global features.

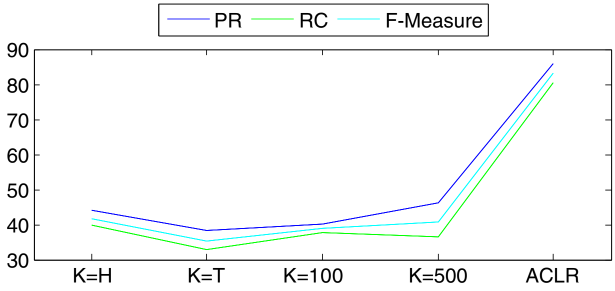


Fig. 5 PR, RC and F-Measure of ACLR or like using adaptive clustering vs. setting cluster number on GOLD

Then local feature refinement is utilized to select out the near duplicate image groups. Finally, we merge the over-partitioned groups. The adaptive clustering is to constrain the local feature utilization into a small scale. The experiments show that our proposed algorithm can find out near duplicate image groups with good performance.

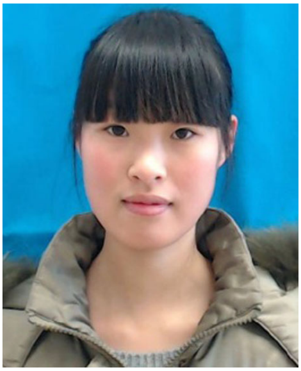
References

1. Battiato S, Farinella GM, Guarnera GC (2010) Bags of phrases with codebooks alignment for near duplicate image detection, MiFOR'10, October 29, 2010, Firenze, Italy
2. Chum J, Isard O, Sivic M, Zisserman JA (2007) Object retrieval with large vocabularies and fast spatial matching. Comput Vis Pattern Recognit, CVPR '07
3. Gao Y, Wang M, Tian Q (2011) Less is more: efficient 3-D object retrieval with query view selection. IEEE Trans Multimed 13(5)
4. Gao Y, Wang M, Shen J (2013) Visual-textual joint relevance learning for tag-based social image search. IEEE Trans Image Process 22(1)
5. Han Y, Xu Z, Ma Z, Huang Z (2013) Image classification with manifold learning forout-of-sample data. Sig Process 93(8), August
6. Han Y, Yang Y, Ma Z, Shen H, Sebe N, Zhou X (2014) Image attribute adaptation. IEEE Trans Multimed
7. Hu Y, Cheng X, Chia LT, Xie X (2009) Coherent phrase model for efficient image near-duplicate retrieval. IEEE Trans Multimed 11(8), DECEMBER, 2009
8. Kennedy L, Naaman M (2008) Generating diverse and representative image search results for landmarks. WWW
9. Lee J, Tong W, Jin R, Jain AK Image retrieval in forensics: application to tattoo image database. IEEE Multimed
10. Li X, Snoek CGM, Worring M, Smeulders AWM (2011) Social negative bootstrapping for visual categorization. ICMR'11, April 17–20, Trento, Italy
11. Li J, Qian X, Tang Y, Yang L (2013) GPS estimation for users' photos. MMM
12. Li J, Qian X, Tang Y, Yang L, Mei T (2014) GPS estimation for places of interest from social users' uploaded photos. IEEE Trans Multimed
13. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. Int J Comput Vis 60:91–110
14. Philbin J, Sivic J, Zisserman A (2010) Geometric latent Dirichlet allocation on amatching graph for large-scale image datasets. Int J Comput Vis
15. Qian X, Guo D, Hou X, Li Z, Wang H, Liu G (2012) HWVP: Hierarchical wavelet packet descriptors and their applications in scene categorization and semantic concept retrieval. Multimed Tools Appl pp 1–24
16. Sayad IE, Martinet J, Urruty T, Benabbas Y, Djeraba C (2011) A semantically significant visual representation for social image retrieval vol. 978-1-61284-350-6/11/$26.00 ©2011 IEEE
17. Song J, Yi Y, Huang Z, Shen H, Hong R (2011) Multiple feature hashing for real-time large scale near-duplicate video retrieval. MM'11, November 28–December 1, 2011, Scottsdale, Arizona, USA
18. Stricker MA, Orengo M (1995) Similarity of color images. Proc IS&T/SPIE\'s Symp Electron Imaging Sci Technol pp 381–392
19. Wang X, Zhang L, Ma W Duplicate-search-based image annotation using web-scaledata. doi:10.1109/JPROC.2012.2193109
20. Wang B, Li Z, Li M, Ma W (2006) Large-scale duplicate detection for web image search. ICME
21. Wang M, Yang K, Hua X (2010) Towards a relevant and diverse search of social images. IEEE Trans Multimed 12(8)
22. Wang Y, Hou Z, Leman K, Pham NT, Chua TW (2011) Combination of local and global features for near-duplicate detection. In: Lee K-T et al. (eds), MMM 2011, Part I, LNCS 6523, pp 328–338
23. Wu P, Hoi SCH, Zhao P, He Y (2011) Mining social images with distance metric learning for automated image tagging. WSDM'11, February 9–12, 2011, Hong Kong, China
24. Xu D, Cham TJ, Yan S, Chang S Near duplicate image identification with spatially aligned pyramid matching
25. Xu D, Cham TJ, Yan S, Duan L, Chang S (2010) Near duplicate identification with spatially aligned pyramid matching. IEEE Trans Circ Syst Video Technol
26. Zhou W, Lu Y, Li H, Song Y, Tian Q (2010) Spatial coding for large scale partial-duplicate web image search. MM'10, October 25–29, 2010, Firenze, Italy

Jing Li received the B.A. degree and MSD from Xi’an Jiaotong University in 2010 and 2013, she pursuit. From Sept. 2010 to July 2013, she was a MSD student at SMILES LAB, Xi’an Jiaotong Univeristy. Now she is a PhD student at Purdue University. Her research interests include computer vision, large scale image retrieval & recognition and data mining and knowledge discovery from social multimedia.

Xueming Qian (M’10) received the B.S. and M.S. degrees in Xi’an University of Technology, Xi’an, China, in 1999 and 2004, respectively, and the Ph.D. degree in the School of Electronics and Information Engineering, Xi’an Jiaotong University, Xi’an, China, in 2008, after that he was an assistant professor. He was an associate professor from Nov. 2011 to March 2014, and now he was a full professor. He was awarded Microsoft fellowship in 2006. He was awarded outstanding doctoral dissertations of Xi’an Jiaotong Univeristy and Shaanxi Province in 2010 and 2011 respectively. He is the director of SMILES LAB. He was a visit scholar at Microsoft research Asia from Aug. 2010 to March 2011. His research interests include social media big data mining and search. His research is supported by NSFC, Microsoft Research, and MOST.

Qing Li received the B.A. degree from Xi'an Jiaotong University in 2013. From Sept. 2012 to July 2013, he was a visiting student at SMILES LAB, Xi'an Jiaotong Univeristy. Now he is a PhD student at Wisconsin University.



Yisi Zhao From Sept. 2012 to July 2015, she was a MSD student at SMILES LAB, Xi'an Jiaotong Univeristy. Her research interests include large scale image retrieval and image content understanding.

Liejun Wang received both his B.S. and Ph.D.degrees in the area of information engineering from Xi'an Jiaotong University, Xi'an, China. He is now a Professor of the Information Science and Engineering, Xinjiang University. His research interest covers image processing and pattern recognition.

Yuan Yan Tang (F'04) received the B.S. degree in electrical and computer engineering from Chongqing University, Chongqing, China, the M.S. degree in electrical engineering from the Beijing University of Post and Telecommunications, Beijing, China, and the Ph.D. degree in computer science from Concordia University, Montreal, QC, Canada. He is currently a Professor with the Department of Computer Science, Chongqing University, a Chair Professor with the Department of Computer Science, Hong Kong Baptist University, Hong Kong, and an Adjunct Professor of computer science with Concordia University. He is an Honorary Lecturer with the University of Hong Kong and an Advisory Professor with many institutes in China. He is the Founder and Editor-in-Chief of the *International Journal on Wavelets*, *Multiresolution*, and *Information Processing* and the Associate Editor for several international journals on pattern recognition and artificial intelligence. He has published more than 250 technical papers and is the author or a coauthor of 21 books and book chapters on several subjects, e.g., electrical engineering and computer science. His research interests include wavelet theory and applications, pattern recognition, image processing, document processing, artificial intelligence, parallel processing, Chinese computing and VLSI architecture. Prof. Tang is a Fellow of the International Association for Pattern Recognition. He has been the General Chair, the Program Chair, and a Committee Member for many international conferences. He was the General Chair of the 19th International Conference on Pattern Recognition (ICPR 2006).