A Large-Scale Image Retrieval Method Based on Image Elimination Technology and Supervised Kernel Hash

Zhiming Yin(✉), Jianguo Sun, Xingjian Zhang, Liu Sun, and Hanqi Yin

College of Computer Science and Technology, Harbin Engineering University, Harbin, China
{yinzhiming,sunjianguo,yinhanqi}@hrbeu.edu.cn
riosmail@126.com, sunliuhrbeu@163.com

Abstract. The Internet develops rapidly in the era of big data, which can be shown by the widespread uses of image processing software as well as digital images skills. However, there are a large number of redundant images in the network, which not only occupy the network storage but also slow down image search speed. At the same time, the image hash algorithm has received extensive attention due to its advantages of improving the image retrieval efficiency while reducing storage space. Therefore, this paper aims to propose a large-scale image retrieval method based on image redundancy and hash algorithm for large-scale image retrieval system with a large number of redundant images. I look upon the method into two phases: The first phase is eliminating the redundancy of repetitive images. As usual, image features need to be extracted from search results. Next, I use the K-way, Min-Max algorithm to cluster and sort the returned images and filter out the image classes in the end to improve the speed and accuracy of the image retrieval. Fuzzy logic reasoning comes to the last part. It can help to select the centroid image so as to achieve redundancy. The second phase is image matching. In this stage, the supervised kernel hashing is used to supervise the deep features of high-dimensional images and the high-dimensional features are mapped into low-dimensional Hamming space to generate compact hash codes. Finally, accomplish the efficient retrieval of large-scale image data in low-dimensional Hamming of the space. After texting three common dataset, the preliminary results show that the computational time can be reduced by the search image redundancy technology when filter out the invalid images. This greatly improves the efficiency of large-scale image retrieval and its image retrieval performance is better than the current mainstream method.

Keywords: Image retrieval · Image redundancy · Fast matching · Supervised kernel hashing · Fuzzy logic inference

This work is supported by the Fundamental Research Funds for the Central Universities (HEUCFG201827, HEUCFP201839).

S. Liu and G. Yang (Eds.): ADHIP 2018, LNICST 279, pp. 233–243, 2019.
https://doi.org/10.1007/978-3-030-19086-6_26

1 Introduction

In recent years, with the rapid development of technologies such as mobile Internet and social network media, hundreds of millions of pictures, video and other multimedia data are generated every day on the Internet. In the era of mobile Internet, people can take a variety of pictures and videos anytime and anywhere, and share them with friends on the Internet. These bring the explosion of digital pictures and videos through the Internet. Therefore, we should pay more extensive attention to the research of image retrieval technology.

The core of image retrieval technology was to find images that match the user's needs from the image database based on the information which provided by the user. The research of technology has been started since the 1970s and has been a research hotspot in the computer field. In the image search engine (Google, Baidu, Bing), it retrieved and query pictures in the massive Internet pictures through the image retrieval technology. In the application of e-commerce websites, it is necessary to find out the products that meet the requirements in a large number of products and use the mobile phone photographs to input the query pictures in order to realize shopping navigation. The map search technology is widely used in multiple major applications (Amazon shopping search, Baidu map, Taobao search).

At the same time, with the rapid development of the social networks such as Facebook, Flickr, and YouTube, the copying and dissemination of multimedia data such as images, videos, and audio is more convenient and fast. This make the Internet full of redundant data. For example, the literature [1] shows that up to 40% of the content on the Internet is duplicated. The literature [2] also pointed out that the video community YouTube's transmission traffic accounts for 20% of the entire Web traffic and 10% of the entire Internet traffic, of which more than 25% of the video is repeated or approximately repeated. The emergence of this situation not only causes a waste of a large amount of storage space but also makes the data retrieval take more time, which will seriously affecting the user experience.

In order to realize effective retrieval of the large-scale high-dimensional image data. The researchers proposed the Approximate Nearest Neighbor (ANN). Among them, the hash technique is the mainstream method to solve the problem of approximate nearest neighbor retrieval. The idea is to use the hash function family to map high-dimensional image features into low-dimensional space and map the closer distances of the original space to low-dimensional at the same time. The distance is still close in this space. Early hashing methods, such as position-sensitive hashing [3] (Locality SensitiveHashing, LSH) and its improved algorithm [4,5], use the random mapping to construct a hash function. In order to ensure a higher accuracy, it is necessary to generate a longer hash code. But as the hash code grows, the probability of the similar images' hash code mapping to go to the same hash bucket will gradually decrease and it will result in a lower recall rate. The hash functions constructed by LSH and its improved algorithm [4,5] are data-independent. In recent years, researchers have constructed some effective methods for how to combine data features. The compact hash functions proposes many algorithms. Weiss et al. [6] proposed a spectral hashing method (Spectral Hashing, SH), firstly analyzes the Laplacian matrix eigenvalues and eigenvectors of similar graphs, and then transforms the image feature vector coding problem into the dimension reduction problem of the Laplacian feature graph by relaxing the constraint conditions. Relying on the data to generate an index can achieves a higher accuracy than the random hash function method. The unsupervised method does not consider the semantic information of the image but the users tend to search the semantic information of the result. For this reason, Wang et al. [7] proposed the semi-supervised hash method by using the semantic similarity of the image as the supervised information (Semi-Supervised Hashing, SSH). On the basis of the semi-supervised, the researchers also proposed some full-supervised hash methods. The full-supervised hash method can achieve higher accuracy than the unsupervised method. Rate, but there are some problems such as complicated optimization process and low training efficiency, which seriously limit its application on large-scale data sets.

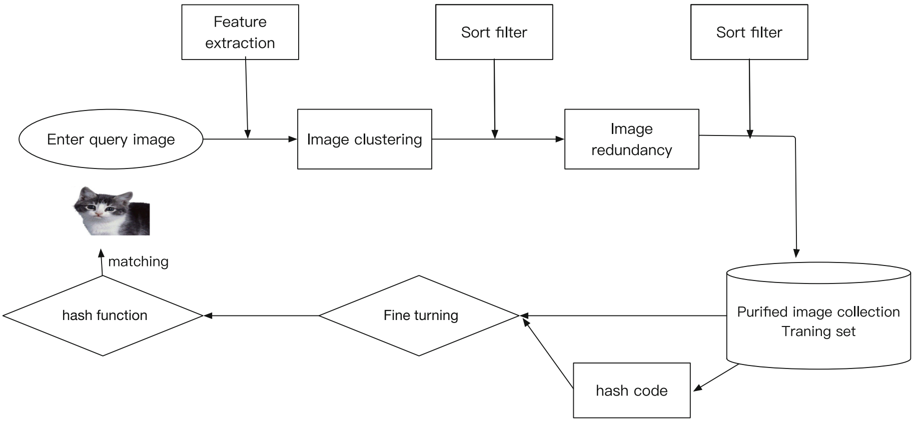

Fig. 1. Schematic diagram of image redundancy technology and large-scale image retrieval method for supervising kernel hashing

Therefore, in order to deal with the above challenges. This paper proposes a method based on large-scale image retrieval. The core idea is to reduce the computation time by filtering out invalid images through image redundancy technology, which greatly improves the efficiency of large-scale image retrieval. The whole method process is shown in Fig. 1. The method is divided into two phases: (1) The stage of eliminating the redundancy of repetitive images, at which the image feature will be performed on the retrieval result. Then use the K-way, Min-Max algorithm to cluster and sort the returned images and filter out the image classes in the end to improve the image retrieval speed and accuracy. Finally, use the fuzzy logic reasoning to simulate the human decision-making process. According to the image attribute information, an optimal solution that conforms to human perception is selected as the centroid image to achieve redundancy. (2) In the image matching stage: use the supervised kernel hash method to supervise the deep features of high-dimensional images, enhance the resolution of linear indivisible data, and use the equivalent relationship between hash code inner product and Hamming distance. A simple and effective objective function, combined with the similarity information of the training image, supervises the high-dimensional image features and generates a compact hash code. Finally, use the trained hash function to construct the image index to achieve efficient retrieval of large-scale image data.

2 Image Redundancy Based on the Clustering Algorithm and the Fuzzy Logic Inference

2.1 Clustering Strategy

Since there is a connection between most of the images returned by the search and the search terms, there should also be a connection between these images. Although the visual difference of the returned results is very large, we can always divide these images into classes with appropriate semantic interpretations. Based on this fact, we can confirm that cluster analysis is a feasible method.

The K-way min-max cut provides a more robust clustering by maximizing the cumulative intra-cluster similarity and minimizing the cumulative inter-cluster similarity simultaneously. The K-way min-max cut algorithm takes the following steps: (a) For a given image topic(query) C returned with n images, a graph G is constructed according to their visual similarity, where nodes are images and edges are characterized by the mixture-of-kernels $\kappa(\cdot,\cdot)$. (b) All the images for the topic C can be partitioned into K clusters automatically by minimizing the following objective function:

$$\min\{\psi(C, K, \hat{\beta}) = \sum_{i=1}^{K} \frac{s(G_i, \frac{G}{G_i})}{s(G_i, G_i)} \tag{1}$$

where $G = \{G_i | i = 1, \ldots, K\}$ is used to represent K images clusters, G/G_i describes the residual image clusters in G except G_i, and $\hat{\beta}$ is the set of optimal kernel weights. $s(G_i, G_i)$ and $s(G_i, \frac{G}{G_i})$ are respectively the similarity within the accumulative class and the similarity between the accumulative classes.

To solve this optimization problem by matrix form, we define $X = [X_1, \ldots, X_i, \ldots, X_K]$ where Xi is a binary indicator (0 and 1) used to indicate the appearance of images in the ith cluster G_i, i.e.:

$$x_i(u) = \begin{cases} 1, & u \in G_i \\ 0, & otherwise \end{cases} \tag{2}$$

and W is defined as a n $\times$ n symmetrical matrix with entry to be $W_{u,v} = \kappa(u, v)$. D is defined as a diagonal matrix and its diagonal components are $D_{u,v} =$

$\sum_{v=1}^{n} W_{u,v}$. Then an optimal partition of returned images can be achieved by:

$$\min\left\{\sum_{i=1}^{K}\frac{x_i^T(D-W)x_i}{X_i^TWX_i}=\sum_{i=1}^{K}\frac{x_i^TDXi}{X_i^TWX_i}-K\right\} \tag{3}$$

Then we can use $\tilde{W}=D^{-1/2}WD^{-1/2}$ and $\tilde{X}_i=\frac{D^{1/2}X_i}{\|D^{1/2}X_i\|}$ to get the objective function:

$$\min\left\{\sum_{i=1}^{K}\frac{1}{X_i^TWX_i}-K\right\} \tag{4}$$

Finally, the optimal solution for Eq. (1) can be achieved by solving the following eigen-problem: $.\tilde{W}\cdot\tilde{X}_i=\lambda_i\cdot\tilde{X}_i, i\in[1,\ldots,K]$.

2.2 Eliminate the Redundancy of Repetitive Images

Eliminating the redundancy of repetitive images means the image attribute should be consistent with human perception when select an optimal solution in the repeated image collection. Then delete other image copies and create a pointer to a centroid image, when necessary, it is transformed from the centroid image. The image has a lot of properties, different fields have different interest. There is currently no uniform standard. According to the purpose of redundancy and the visual perception characteristics of human beings, we give the image attribute definition at first.

Theorem 1. *Image attributes refer to the nature of the image. You can use a triple to represent P_i = (size, resolution, cLarity). In this triple, i represents the i-th image, P_i represents the image attribute of the user's attention with scalability, size represents the size of the image, resolution represents the image resolution, and clarity represents image clarity.*

According to the Theorem 1, we can see that the image attribute values can be easily obtained in practice and can be extended according to the situation. However, the values of these attributes represent different meanings and importance in different range. In practice, we need to select the appropriate centroid image according to experience. There is no specific quantitative standard yet, which can be the reference for the selection of centroid images. Therefore, this paper starts with the fuzzy logic reasoning. The empirical rules are used to simulate the human decision-making mode, and the quantitative information CQV (comprehensive quantitative values) which can represent the comprehensive information of the image is inferred from the image information of each dimension of the image. Select a centroid image based on CQV. The fuzzy logic reasoning process includes fuzzy, rule base, inference engine, and defuzzification. The specific process is shown in Fig. 2.

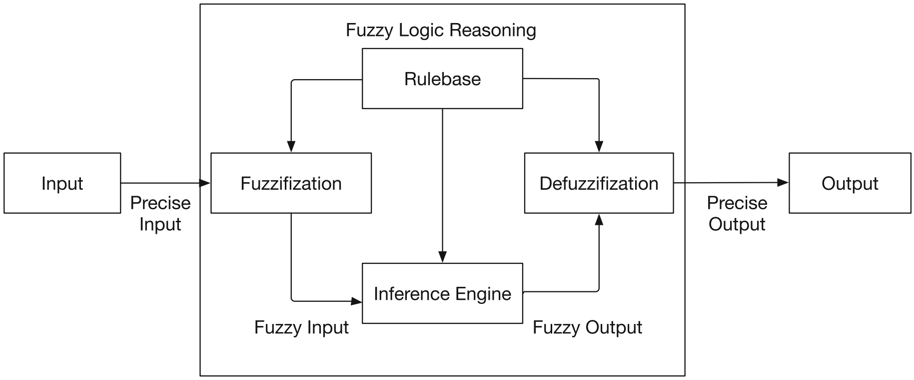

Fig. 2. Fuzzy logic reasoning

(1) Fuzzification: Converting explicit data from the outside world into appropriate linguistic ambiguous information. The ambiguous subset defined in this paper is L, ML, M, MB, B. The corresponding language variable is L = little, ML = medium little, M = medium, MB = medium big, B = big. Blurring process: First, the explicit data Pi is normalized and converted into corresponding linguistic variables. Then use the Gaussian membership function to assign a membership degree to each linguistic variable.
(2) Rule base: The rule base represents the whole thought system, it can be provided by experts or the test data sets which extracted by special training. The quality of the rule can directly affects the whole reasoning effect.
(3) Inference engine: Simulating the decision patterns of human thought by the approximate reasoning or the fuzzy reasoning. The core of the inference engine is to generate a fuzzy relation matrix according to the rules R. When the fuzzy relation matrix R is determined, the input fuzzy matrix A' can be outputted via R to B'.
(4) Unambiguous: The conclusion B' generated by fuzzy logic inference is a fuzzy output, and we need to convert it to a clear value. Here, the center of gravity method is used to deblur. As shown in Eq. (5), the end result is CQV.

$$CQV = \frac{\sum_{i=1}^{n} u_c(y_i) \times y_i}{\sum_{i=1}^{n} u_c(yi)} \tag{5}$$

3 Image Retrieval Based on Supervised Kernel Hash

To enhance the resolution of the hash function to the linearly inseparable high-dimensional data $X = \{x_1, \ldots, x_n\} \subset R^d$. We use the kernel function $k : R^d \cdot R^d \rightarrow R$ to construct the hash function $k : R^d \rightarrow \{1, -1\}$ and map the high-dimensional data to generate a hash code. The specific form of the hash function is:

$$f(x) = \sum_{j=1}^{m} k(x_{(j)}, x)a_j - b \tag{6}$$

$$h(f(x)) = sgn(f(x)) = \begin{cases} 1, & f(x) > 0 \\ -1, & f(x) \leq 0 \end{cases} \tag{7}$$

among them, $a_i \in R, b \in R, x_{(1)}, \ldots, x_{(n)}$ is randomly selected samples from X and its total number is m. In order to implement the fast hash mapping, m is a constant much smaller than n. In addition to satisfying the similarity between the low-dimensional Hamming space and the original high-dimensional space, the hash function h(x) should also ensure that the generated hash code is balanced, that is, the hash function h(x) should be ensure that $\sum_{i=1}^{n} h(x_i) = 0$, then biased $b = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m} k(x_{(i)}, x)a_j$, substituting the value of b into the Eq. (6), we can know that:

$$f(x) = \sum_{j=1}^{m} \left(k(x_{(j)}, x) - \frac{1}{n}\sum_{i=1}^{n} k(x_{(j)}, x)a_j\right) = \mathbf{a}^T\overline{\mathbf{k}}(x) \tag{8}$$

In this equation, $\mathbf{a} = [a_1, \ldots, a_n]^T$, $\overline{\mathbf{k}} : R^d \to R^m$ is the mapping vector: $\overline{k} = \left[k(x_{(1)}, x) - \mu_1, \ldots, k(x_{(m)}, x) - \mu_m\right]^T$, $\mu_j = \frac{1}{n}\sum_{i=1}^{n} k(x_{(j)}, x_i)$, we can get the μ_j from beforehand calculate and the correlation information of training data is used for the supervised learning to get the vector $\mathbf{a}$.

If we ensure the dimension of the hash code is r, we need r vectors $\mathbf{a_1}, \ldots, \mathbf{a_r}$ to construct a hash function $\mathscr{H} = \{h_k(x) = sgn(\mathbf{a}^T\overline{\mathbf{k}}(x))|k \in [1, r]\}$.

We can get the training image's tag information from the image's semantic relevance and spatial distance, $lable(x_i, x_j) = 1$ express that the image x_i and x_j is similar, $lable(x_i, x_j) = -1$ express that the image x_i and x_j has large difference. To describe the relationship between the image xi, xj in the label image set $\chi = x_1, \ldots, x_l$, we define the supervised matrix $\mathbf{S}$:

$$s_{ij} = \begin{cases} 1, & lable(x_i, x_j) = 1 \\ -1, & lable(x_i, x_j) = -1 \\ 0, & otherwise \end{cases} \tag{9}$$

among them, $lable(x_i, x_j) = 1, S_{ii} = 1, S_{ij} = 0$ represent that the similarity between image x_i and x_j is uncertainty. In order to enhance the distinguishing ability of hash codes and let the similarity between images can be judged efficiently in the bright space, the Hamming distance $D_h(x_i, x_j)$ of the images x_i, x_j should be made as much as possible meets:

$$D_h(x_i, x_j) = \begin{cases} 0, & S_{ij} = 1 \\ r, & S_{ij} = -1 \end{cases} \tag{10}$$

Due to the complex form of the Hamming distance calculation formula, it is difficult to directly optimizat, so this paper uses the vector inner product operation to calculate the distance between the hash code. Remember that the hash code for image x is code, $(x) = [h_1(x), \ldots, h_r(x)] \in \{1, -1\}^{1\times r}$, the distance $D(x_i, x_j)$ of the image x_i, x_j is:

$$\begin{aligned} D(x_i, x_j) &= code_r(x_i) \cdot code_r(x_j) \\ &= |\{k|h_k(x_i) = h_k(x_j), 1 \le k \le r\}| \\ &\quad - |\{k|h_k(x_i) \ne h_k(x_j), 1 \le k \le r\}| \\ &= r - 2|\{k|h_k(x_i) \ne h_k(x_j), 1 \le k \le r\}| \\ &= r - 2D_h(x_i, x_j) \end{aligned} \tag{11}$$

Equation (11) shows that the inner product of the hash code is consistent with the Hamming distance operation and $D(x_i, x_j) \in [-r, r]$. Normalized the $D(x_i, x_j)$ we can get that $S_{ij} = \frac{D(x_i,x_j)}{r} \in [-1, 1]$. In order to let the distance between the similarity matrix $S' = \frac{1}{r} H_l H_l^T$ supervising matrix S is the smallest, defining the objective function:

$$min\Gamma = \left\| \frac{1}{r} H_l H_l^T - S \right\|_F^2 \tag{12}$$

Of which, $\|\cdot\|_F^2$ represent to get the norm of the matrix Frobenius, $H_l = \begin{bmatrix} code_r(x_l) \\ \cdots \\ code_r(x_l) \end{bmatrix} \in \{1, -1\}^{l \times r}$ is the hash code matrix $\mathbf{x}_l$ of label image set. Extend $sgn(\cdot)$ to the matrix form. According to formula (8), H_l can be expressed as:

$$H_l = \begin{bmatrix} h_l(x_l) \dots h_r(x_l) & \\ \dots \quad \dots \dot{h}_l(x_l) & \dots h_r(x_l) \end{bmatrix} = sgn(\bar{K}_l A) \tag{13}$$

Then sorting H_l into formula (12):

$$min\Gamma(A) = \left\| \sum_{k=1}^{r} sgn(\bar{K}_l \alpha_k)(sgn(\bar{k}_l \alpha_k))^T - rS \right\|_F^2 \tag{14}$$

Compared with the BRE, the objective function $\Gamma(A)$ calculates the similarity by the inner product, and the parameter A is more intuitive. Suppose that at the time of t = k, the vector $a_1^*, \dots, a_{k-1}^*$ is known, and it is necessary to estimate a_k and define the matrix $R_{k-1} = rS - \sum_{i=l}^{k-1} sgn(\bar{k}_l a_l^*)(sgn(\bar{k}_l a_l^*))^T$, specialize $R_0 = rS$, the a_k can be estimated step by step through the greedy algorithm minimum (14):

$$\begin{aligned} &\|sgn(\bar{K}_l a_k)(sgn(\bar{K}_l a_k))^T - R_{k-1}\|_F^2 \\ &= ((sgn(\bar{K}_l a_k))^T sgn(\bar{K}_l a_k))^2 - 2(sgn(\bar{K}_l a_k))^T R_{k-1} sgn(\bar{K}_l a_k) + tr(R_{k-1}^2) \\ &= -2(sgn(\bar{K}_l a_k))^T R_{k-1} sgn(\bar{K}_l a_k) + l^2 + tr(R_{k-1}^2) \\ &= -2(sgn(\bar{K}_l a_k))^T R_{k-1} sgn(\bar{K}_l a_k) + const \end{aligned} \tag{15}$$

By removing the constant term, we can get a more concise objective function: $\vartheta(a_k) = -(sgn(\bar{K}_l a_k))^T R_{k-1} sgn(\bar{K}_l a_k)$

Due to the sgn(x) function in the objective function $\vartheta(a_k)$ is not continuous, and $\vartheta(a_k)$ is not a convex function, it is difficult to directly minimize the

$\vartheta(a_k)$. The literature [8] research show that when $|x| > 6$, the continuous function $\varphi(x) = 2/(1 + exp(-x)) - 1$ can have a good approximation to sgn(x). Therefore, this paper replaces sgn(x) to $\varphi(x)$ to get the approximate objective function $\vartheta(a_k)$: $\vartheta(a_k) = -(\varphi(\bar{K}_l a_k))^T T_{k-1} \varphi(\bar{K}_l a)k)$. We can minimize the $\vartheta(a_k)$ by gradient descent, and get the gradient of the $\vartheta(a_k)$ related a_k:

$$\nabla\vartheta(a_k) = -\bar{K}_l^T((R_{k-1}b) \bigodot (\mathbf{1} - b \bigodot b)) \tag{16}$$

$\mathbf{b} = \varphi(\bar{k}_l, a_k) \in R^l$, $\mathbf{1} = [1, \ldots, 1]$ and $\bigodot$ represent the operation of inner product. The smoothed ϑ is not a convex function, and the global optimal solution cannot be obtained. In order to accelerate ϑ convergence, this paper uses the spectral analysis method in spectral hash [6] to generate the initial value of a_k^0, then use the method [9] to accelerate the gradient optimization process. After the vector coefficient $\mathbf{a}$ is obtained by supervised learning, the hash function $\mathscr{H}$ and the hash table H can be generated. Hash mapping the deep features of the query image can get the $code_r(x_q)$. Calculate the distance between $code_r(x_q)$ and the hash code in the hash table H, and put the return image with a closer distance as a result of the search.

4 Experiment and Analysis

4.1 Experimental Setup and Performance Evaluation

This article is based on the Image Net-1000 [10], and MNIST [11] to assess the methods in thus paper. The Image Net-1000 image set is a subset of the ImageNet and it is a evaluation data set of the Large Scale Visual Recognition Challenge (LSVRC). The evaluation dataset contains 1000 categories and a total of 1.2 million images. The dataset contains 70,000 sheets of 28×28 size handwritten digital grayscale pictures, numbers from 0 to 9, and in each category thousand pictures have 7 numbers.

The experimental hardware is configured as a 6G GPU which device is GTX Titan and Intel Xeon CPU, 16G server. The image retrieval performance indicators use Mean Average Precision (MAP), which is defined as follow:

$$MAP = \frac{\text{Average precision of multiple image retrieval}}{Number\,of\,searches} \times 100\% \tag{17}$$

4.2 Experimental Results and Analysis

In order to verify the performance of the Supervised Kernel Hash (KSH) search whether has an excellent performance. The more savvy, there are some current mainstream hashing methods in Image Net-1000. Experimental comparisons were made on the image set, including Locally Sensitive Hash Algorithm (LSH), Spectral hash algorithm (SH), Unsupervised iterative quantization hash algorithm (ITQ), The supervised hash algorithm that minimizes losses (MLH),

Table 1. Hamming sorting using different length hash codes on ImageNet-1000

	12bits	24bits	32bits	48bits
KSH	0.303	0.334	0.344	0.356
ITQ-CCA	0.259	0.278	0.281	0.286
MLH	0.168	0.189	0.204	0.208
BRE	0.151	0.180	0.193	0.195
SH	0.122	0.122	0.124	0.125
ITQ	0.160	0.164	0.170	0.173
LSH	0.120	0.122	0.117	0.117

Table 2. Hamming sorting using different length hash codes on MNIST

	12bits	24bits	32bits	48bits
KSH	0.857	0.877	0.884	0.892
ITQ-CCA	0.644	0.687	0.710	0.718
MLH	0.477	0.598	0.649	0.650
BRE	0.511	0.574	0.602	0.619
SH	0.270	0.274	0.275	0.277
ITQ	0.397	0.399	0.410	0.414
LSH	0.185	0.197	0.213	0.132

Binary Reconstruction Embedded Supervised Hash Algorithm (BRE), Supervised Iterative Quantization Hash Algorithm (ITQ-CCA) and other methods.

It can be seen from the Tables 1 and 2 that as the number of the bits of hash code r increased, the MAP value of each method increased. Comparing the image retrieval of each hash method MAP, the value shows that the search use this method (KSH) has a better performance than other mainstream methods. It is because that the unsupervised hash methods (such as LSH, SH, DSH, PCA-ITQ, etc.) and the supervised hash method BRE do not have a good use of the semantic information of the image to construct the hash function, and result a lower retrieval performance. But the KSH introduces the kernel function to construct hash, the function enhances the resolving power of linear indivisible data, and combines the similarity information of the image to train the hash function to generate a more compact hash code, this will improve the image retrieval performance.

5 Conclusion

In this paper, a large-scale image retrieval method based on image redundancy and hash algorithm is proposed for the large-scale image retrieval system with a large number of redundant images. The core idea of this method is to use the image redundancy technology to reduce the computation time by filtering out invalid images, and its image retrieval performance is better than the current mainstream methods. Beside, it has greatly improved the efficiency of large-scale image retrieval.

References

1. Yang, X., Zhu, Q., Cheng, K.T.: Near-duplicate detection for images and videos. In: ACM Workshop on Large-Scale Multimedia Retrieval and Mining, pp. 73–80. ACM (2009)
2. Pedro, J.S., Siersdorfer, S., Sanderson, M.: Content redundancy in YouTube and its application to video tagging. ACM Trans. Inf. Syst. **29**(3), 13–43 (2011)
3. Indyk, P., Levi, R., Rubinfeld, R.: Erratum for: approximating and testing k-histogram distributions in sub-linear time. In: ACM Symposium on Principles of Database Systems, pp. 343–343. ACM (2015)
4. Kulis, B., Grauman, K.: Kernelized locality-sensitive hashing. IEEE Comput. Soc. **34**(6), 1092–1104 (2012)
5. Datar, M., Immorlica, N., Indyk, P., et al.: Locality-sensitive hashing scheme based on p-stable distributions. In: Twentieth Symposium on Computational Geometry, pp. 253–262. ACM (2004)
6. Weiss, Y., Torralba, A., Fergus, R.: Spectral hashing. In: International Conference on Neural Information Processing Systems, pp. 1753–1760. Curran Associates Inc. (2008)
7. Wang, J., Kumar, S., Chang, S.F.: Semi-supervised hashing for large-scale search. IEEE Trans. Pattern Anal. Mach. Intell. **34**(12), 2393–2406 (2012)
8. Chang, S.F.: Supervised hashing with kernels. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2074–2081. IEEE Computer Society (2012)
9. Nesterov, Y.: Introductory lectures on convex optimization. Appl. Optim. **87**(5), xviii, 236 (2004)
10. Deng, J., Dong, W., Socher, R., ImageNet: a large-scale hierarchical image database. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009, pp. 248–255. IEEE (2009)
11. Li, F.F., Fergus, R., Perona, P.: Learning generative visual models from few training examples: an incremental bayesian approach tested on 101 object categories. In: Conference on Computer Vision and Pattern Recognition Workshop, p. 178. IEEE Computer Society (2004)