Acta Technica *62* (2017), No. 7A, 1001–1012

Distributed image retrieval based on hadoop technology and sparse representation model

Nian Fu[1,2]

Abstract. In the process of image search and matching, we use only one image as a query image. The content-based image retrieval technology using color, shape and texture features of the image, as in the vector space, by some distance between two points is calculated to measure the similarity between images, however, when the image feature extraction of image features of the same type will appear inconsistent, greatly affect the retrieval accuracy, so, distributed image retrieval based on hadoop technology and sparse representation model is proposed. The sparse distributed image retrieval based on model, and then use Map to match the features and characteristics of library example, based on the image set sparse low rank description, maintain the global structure of the same category features, but also reduce the noise for the local sensitivity, robustness the retrieval algorithm, the calculation results of Reduce receives the Map task, and sorted according to the similarity of the query image and the size of the local deposit Some of the content related graphs are combined, and the most representative feature points are screened from which to calculate and sort the weights of these features. This method greatly reduces the number of feature points and enhances the robustness of image retrieval. The experimental results on the Corel image set show that this method has a better retrieval effect than the existing content based image retrieval method.

Key words. Large Data, Image, Retrieval, Feature, Hadoop, Sort

.

1. Introduction

As the digital image technology and internet technology have been boosted by leaps and bounds, considerable amount of images are being produced on a day-to-day basis. How to rapidly and accurately seek out the image as required from the

[1]JIANGXI INSTITUTE OF ECONOMIC ADMINISTRATORS, Jiangxi Nanchang, 330088, China

[2]JIANGXI Electronic commerce and Industry upgrade 2011 Collaborative innovation center, Jiangxi Nanchang, 330088, China

database of image has been primarily discussed and studied in the field of computer vision in recent years. The text-based image retrieval technique is to retrieve the keywords of image. Yet labelling the keywords shall consume manpower and materials, and the keywords cannot fully describe the entire contents of the image, so they fail to adapt to the retrieval for a wide range of images. The visual features commonly used in computer vision are colors, textures, shapes, etc. The color is the simplest image feature, which is less dependent on the size, direction and rotation of objects in the image. Texture refers to a visual form that has homologous properties, whereas this form is not represented by a single color or brightness. Shape can distinguish between targets in different images. The technique of content-based image retrieval is to describe the image content based on the visual characteristic information of the image, and transform the feature information into the feature vector. Subsequently, relevant methods are adopted to retrieve the qualified images in the image data.

2. Hadoop theory

Hadoop counts as the distributed computation under open-source framework using the Java, through which the large-scale data can be developed and parallel-processed. It is primarily encompassed by the parallel computation model of HDFS and MapReduce. When developing data based on Hadoop, distributed parallel programs can be operated on a large-scale cluster system consisting of a large number of nodes to complete the calculation of massive data. And it is not necessary to factor in work scheduling, distributed storage, fault tolerance processing, network communication, load balancing and other problems existing in parallel programming. HDFS adopts master/slave framework, and an HDFS cluster consists of a NameNode and a group of DataNode. NameNode serves as the central node, charged with managing the namespace of file system and accessing to the file on client side.

Normally, one DataNode shall be operated on one node in a cluster, charged with the data storage on the node where it is pertained to, and responsible for processing the writing requests on the client side of the file system. The creation, deletion and replication of data blocks are conducted under the unified scheduling of NameNode [11]. HDFS splits files into blocks, which are stored separately on different DataNodes. Each block can be copied several times for storage in different DataNode. For these reasons, HDFS takes on high tolerance and high throughput of data reading and writing.

3. Feature extraction and classification

3.1. Color feature

Most digital images use RGB tricolor storage. Yet, RGB color space can be less distinguishable, which is not consistent with human's understanding of color. The human eye is sensitive to the H component (tone) in the color space of HSV. The

three components of H, S and V are mutually independent, and correspond to the color feature that the human eye can perceive. HSV color space is more suitable to express the difference between different colors [6]. Image color feature extraction first converts images from RGB format to HSV format, and then the image HSV color space is quantized by means of non-equal interval quantization method. The hue h is divided into 16 parts, the saturation s and the brightness v are divided into 4 parts. After the quantification, three components are incorporated as one $I = 16H + 4S + V$, and I is ranged from [0,255]. Eventually, following the function: $V_{color}(i) = n_i/N$, $i = 0, 1, \ldots, L-1$. The histogram statistics is conducted on color of image, and a 256 D feature vector is attained, where i denotes gray level, L denotes type of gray level, n_i indicates total amount of pixels in i gral level in images, and N denotes the overall pixel amount in image.

3.2. Texture feature

The basic principle of LBP [7] is to compare the gray level of pixels in the image with the pixel of 8 neighborhoods surrounding around them. If the center pixel has a gray value less than the surrounding pixel, the value is 1, otherwise it is 0. Through a sorting in clockwise direction, an 8-bit binary number can be attained. Convert the binary code into decimal number, and the LBP value of the center pixel is obtained. As shown in figure 1, the binary code in figure 1 is 10101100 and the LBP value is 172. The specific calculation method of LBP can be expressed in the following formula:

$$LBP(x, y) = \sum_{i=0}^{7} s(p_i - p_c) \times 2^i . \tag{1}$$

$$s(x) = \begin{cases} 1 & x \geq 0 \\ 0 & otherwise \end{cases} \tag{2}$$

Where (x, y) denotes the coordinate of pixel, P_c represents the gray level of central pixel, and P_i indicates the gray level of the neighborhoods. The LBP value of each pixel in the image shows the grayscale distribution of its neighborhood, By collecting the probability of different LBP values; occurrence in the image, a texture histogram is obtained to describe the texture structure of the area (i.e. LBP descriptor).

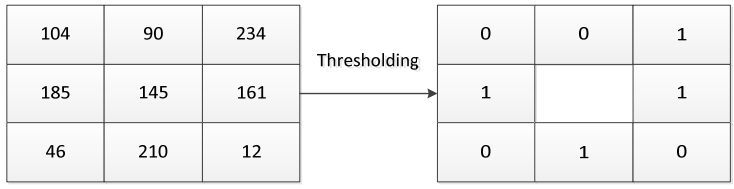

Fig. 1. Basic examples of LBP operator

3.3. shape texture

Shape is a more advanced feature than color or texture, usually associated with the target in the image. As shape feature is adopted, the accuracy of retrieval can further improved. The Zernike moment is based on the orthogonalization function of Zernike polynomials, advantaged by rotation, proportional peace shift, and small redundancy, etc. It is a widely used shape feature descriptor [8]. When calculating the moment of Zernike, it is not necessary to calculate the edge of the image. It can readily establish any high order moment of the image.

Zernike polynomial can be defined in polar coordinates of unit circle ($x^2+y^2<1$) as:

$$V_{nm}(x,y) = R_{nm}(r) \cdot \exp(jm\theta) \,. \tag{3}$$

Where $R_{nm}(r)$ is a set of orthogonal polynomials defined in the unit circle, i.e. the polynomial of Zemike moment in the radial direction. The specific definition is presented below:

$$R_{nm}(r) = \sum_{s}^{(n-|m|)/2} \frac{(-1)^s (n-s)!}{s!\left(\frac{n+|m|}{2}-s\right)!\left(\frac{n-|m|}{2}-s\right)!} r^{n-2s} \,. \tag{4}$$

Where n and m indicate the order of Zernike polynomial, n is the nonnegative integer, $n-|m|$ is the even number, and $n \geq |m|$. For an image $f(x,y)$ with order as n and repetitive rate as m, its Zemike moment shall be calculated as:

$$Z_{nm} = \frac{n+1}{\pi} \sum_{x} \sum_{y} f(x,y) \cdot V_{nm}(x,y) \,. \tag{5}$$

In this paper, the 10th-order Zernike moment [9] is calculated for each image, and the 36 vector is obtained as the shape feature descriptor of the image.

4. Low-rank representation of feature

The color, texture and shape feature extracted from the image are represented in vector form, and each image can be regarded as a point in vector space. The image set can be described using a subspace composed by all the image features. Although there are common features between images in the same image set, the differences are still relatively large. The data do not overall pertain to the same feature subspace, but are generated by multiple subspaces [10]. Using a subspace to describe all the feature data can lead to inaccurate data descriptions, and data is often interfered with noise. The data is required to be extracted from multiple sub-spatial structures and the contaminated data is to be repaired. Although the differences between images are usually large, each image still has the common features representing the type. In this regard, the literature [10] used low-rank representation, LRR[11] to extract the common features of the image concentration, and to remove the unstable

characteristics:

$$\begin{array}{ll} \min\limits_{Z,E} & \|Z\|_* + \lambda \|E\|_{2,1} \\ s.t. & D = AZ + E \end{array} \tag{6}$$

where D is the set of feature vectors of the image set, and A is the over-complete feature dictionary that can be spanned by D, and λ refers to the index of regular terms, adopted to balance the low-rank and noisy parts. The nuclear norm $\|Z\|_* = \sum_{k=1}^{m} \sigma_k(Z)$ of Z is the sum of its singular values. The norm $l_{2,1}$ is the sum of the norm l_2 of each column in the matrix E: $\|E\|_{2,1} = \sum\limits_i \sqrt{\sum\limits_j (E(j,i))^2}$ ($E(j,i)$ is the(j,i)-th element in the matric E)

The main purpose of LRR is to capture the global subspace minimum structure of the data vector and to mitigate the influence of overall noise. Yet LRR is susceptible to the local noise. In fact, the features of all the images in the image set should have a local linear structure while all images in the image set take on low-rank property. This can ensure the minimum number of neighbors in the same subspace and robustness to local noise. Sparse low-rank Representation is adopted in this paper to establish the data structure of the image feature.

$$\begin{array}{ll} \min\limits_{Z,E} & \|Z\|_* + \eta \|Z\|_1 + \lambda \|E\|_{2,1} \\ s.t. & D = AZ + E\,. \end{array} \tag{7}$$

Where $\|Z\|_1$ is adopted to describe the partial linear structure corresponding to the characteristics of images in identical types, l_1 norm refers to the sum of absolute values of all elements in computing matrix Z: $\|Z\|_1 = \sum |Z(j,i)|$, parameter η is adopted to indicate the influence exerted by the sparse area on the result. The data D is adopted as the dictionary of feature, and question (7) can be converted to the following convex optimization problem:

$$\begin{array}{ll} \min\limits_{J,L,E} & \|J\|_* + \eta \|L\|_1 + \lambda \|E\|_{2,1} \\ s.t. & \begin{array}{c} D = DZ + E \\ Z = J, Z = L \end{array} \end{array} \tag{8}$$

Given that (8) is virtually a convex optimization problem, which can be resolved through adopting the method of augmented lagrangian multiplier [13]. Additionally, as literature [13] demonstrates, if μ is valued properly, the solution of question (8) shall be identical to the solution of the minimization problem shown below:

$$\begin{array}{l} \min\limits_{J,L,E,Z,Y_1,Y_2,Y_3} \|J\|_* + \eta \|L\|_1 + \lambda \|E\|_{2,1} \\ + tr\left(Y_1^T (D - DZ - E)\right) + tr\left(Y_2^T (Z - J)\right) + tr\left(Y_3^T (Z - L)\right) \\ + \dfrac{\mu}{2}\left(\|D - DZ - E\|_F^2 + \|Z - J\|_F^2 + \|Z - L\|_F^2\right). \end{array} \tag{9}$$

Where Y_1, Y_2 and Y_3 are lagrangian multipliers, μ refers to the parameter for regulation and control. The calculation process of the whole augmented Lagrangian

multiplier method can be generalized to algorithm 1. Although the steps 1, 2 and 4 of algorithm 1 are convex optimizations, they have different solutions. Step is solved through adopting singular value thresholding operator, step 2 is solved through adopting the soft-thresholding operator in literature [10], and step 4 is solved abiding by the theorem 3.2 mentioned in literature [11]. As the optimal solution of function (8) is attained as (Z^*, E^*), the feature of the image set under the sparse low-rank constraints are described as: $V = D \times Z^*$. The noise in the unstable characteristic is eliminated, which is conducive to the accurate measurement of the similarity in the image retrieval.

5. Hadoop-based image retrieval

Assume that the image to be retrieved is Q, taking on the color feature V_Q^C, texture feature V_Q^T and shape feature V_Q^S; the image in the image library is P, taking on the color feature V_P^C, texture feature V_P^T and shape feature V_P^S. The image feature is represented by the vector form, so that the image can be deemed as a point in the vector space. The similarity between images is measured by calculating some distance between two points [16]. This paper adopts Euclidean[16] to measure the distance between feature vectors, and sets different weights according to different vectors. The Euclidean distance between image Q and the j image in the library is defined as:

$$d\left(P_j, Q\right) = \sqrt{\sum_{i=1}^{n} \left(V_{pi} - V_{qi}\right)^2}. \tag{10}$$

n refers to the dimensionality of feature vector. Through adopting function (10), the similarity of color, texture and shape feature of two images are calculated as d_c, d_t and d_s, respectively. Thus, the similarity of any two images is defined as:

$$S\left(P, Q\right) = W_c d_c + W_t d_t + W_S d_S\,. \tag{11}$$

W_c, W_t and W_S are the weighted value of the distance of color, texture, and shape, $W_c + W_t + W_S = 1$. After retrieval, the three weights can be adjusted by the feedback link to conduct best retrieval optimally. Eventually, the most similar images are regulated as output result according to the similarity.

6. Experimental results and analysis

All the experiments in this article were performed on the computer with CPU for P4 Dual-Core 2.5ghz, memory for 2G, operating system for Windows7, and the experimental environment was Matlab2009b. The experimental data set is the standard image set of the SIMPLIcity system downloaded from Internet. These images are in JPG format of 384*256 and 256* 384, with 10 categories inclusive of cars, dinosaurs, horses, flowers and buildings. Each category contains 100 images, with a total of 1,000 images.

Fig. 2 is the first 20 images of African indigenous peoples retrieved by different methods. The image in the upper left corner of each image is the query image. The rest of the image is the query result. The similarity decreases in turn from left to right and from top to bottom. Fig. 2 (a) is result attained merely retrieved by color feature. Fig. 2 (b) is a search result attained through adopting color, texture and shape, whereas the noise in the feature is not processed. Fig. 2 (c) is the retrieval result with the noise eliminated in the feature vector through adopting LRR. Fig. 2 (d) is the retrieval result through adopting the text-based method.

Fig. 2. The result of different image retrieval methods on indigenous peoples

As can be seen from the Fig. 2 (d), the 20 images retrieved by this method are indigenous and the accuracy is 100%. 10 images of the search results of color histogram were not correct, and the retrieval accuracy was 50%, as shown in Fig.2 (a). Three images of the results attained through adopting three-feature retrieval are incorrect and the retrieval accuracy was 85%, as shown in figure 2(b). One image is not correct in the results retrieved as LRR eliminates the noise, and the retrieval accuracy was 95%, as shown in figure 2(c). It is proved that this algorithm can effectively combine the color histogram, local binary model and Zernike matrix, and effectively remove the noise in the feature vector to achieve higher accuracy.

In order to further explain the performance of this method, the algorithm in this paper is compared with that in the literature [4] and the literature [5]. The retrieval results are evaluated through adopting mean precision ratio. The specific ratio is that: From the image library, 10 images were randomly selected and each image was queried, and the mean precision of each category of image was obtained. The precision ratio is defined as:

$$precision = \frac{s}{s+u}. \tag{12}$$

Where S is the number of related images retrieved in a query, and u is the number of unrelated images retrieved in a retrieval course. FIG. 3 shows the histogram of

average precision, which indicates the method proposed in this paper is better in retrieval than that in literature [4] and literature [5]. Another way to measure the results of a retrieval is to check the accuracy rate – the recall ratio curve. The recall ratio can be defined as:

$$recall = \frac{s}{s+v} \tag{13}$$

V is the number of images associated with the retrieved images but not retrieved in the image library. The higher the precision ratio and recall ratio, the better the retrieval performance. The higher the precision ratio, the lower the recall ratio, and the higher the recall ratio, the lower the precision ratio. Under the identical condition of recall ratio, the higher the precision ratio, the better the performance. Fig. 4 indicates the curves of mean precision ratio and recall ratio of all retrieved images. As indicated from the Fig., the algorithm is this paper outperforms the other two methods.

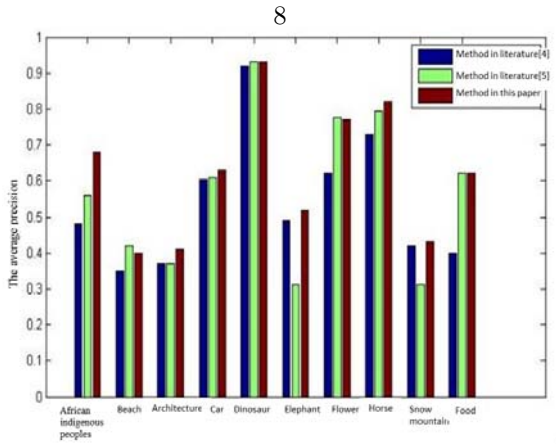

Fig. 3. The comparative average precision of different methods

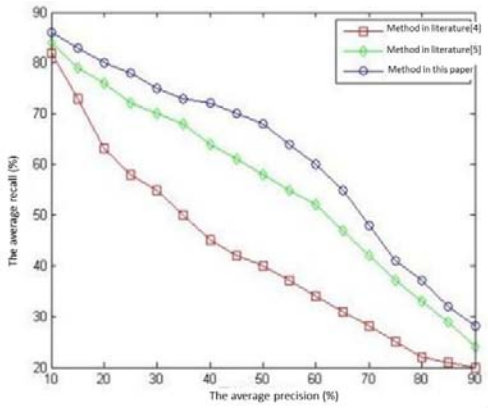

Fig. 4. The average precision—recall cures

As the amount of data retrieved is 200 thousand, 400 thousand, 600 thousand, 800 thousand, 1 million and 1.2 million, the image retrieval time of Hadoop distributed image retrieval system and B/S single-node system is shown in figure 10. When 200,000 images are collected, the retrieval speed of Hadoop distributed system is slower than that of the B/S single node. This is because the amount of image data is small at this point. Hadoop takes the data set as a Map task and processes it on one node. Only one node performs a Map task. The retrieval time is extended due to the time consumed in initialization, assignment and empty operation.

Two Map tasks shall be offered when the data of image contains 400 thousand images. The speed of image retrieval of Hadoop distributed system was improved greatly, compared with the B/S single node system, the retrieval speed of Hadoop distributed system is increased by 41.15%. Three Map tasks shall be offered when the data of image contains 600 thousand images. Compared with the B/S single node system, the retrieval speed of Hadoop distributed system is increased by 33.89%. as the data size continues to be enlarged (over 600 thousand images), the retrieval time of Hadoop distributed system shall be linearly increased. This is because the Map task has exceeded the number of nodes in the distributed system. Some nodes will assign multiple Map tasks, whereas one node can only handle one Map task at the same time. Thus, the retrieval time of Hadoop distributed system shall be linearly increased. Therefore, it is necessary to increase the number of nodes of Hadoop distributed system to enhance the ability of distributed system to process Map tasks in parallel and improve image retrieval speed.

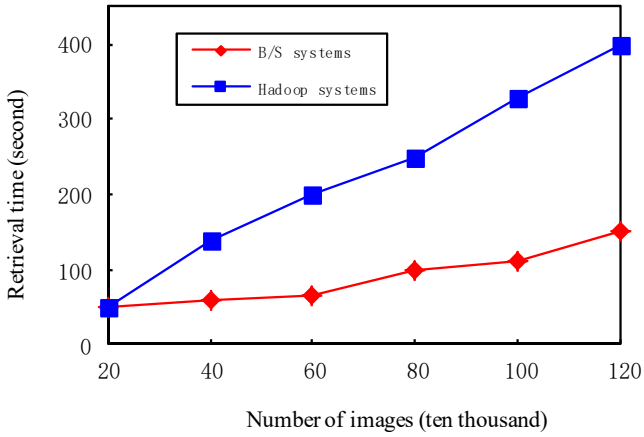

Fig. 5. Comparison between image retrieval efficiency of two systems

7. Conclusion

In terms of the instability of feature vectors extracted from content-based image retrieval methods and the sensitivity to the local noise, this paper presents an image retrieval method with sparse and low rank description. When the image set is

retrieved, the color, texture and shape of the image are extracted first, and then the stable feature is obtained through the sparse and low-rank description method. The image shall be retrieved on the basis of image description. The experiment shows that this method has better retrieval performance, which can transform unstable features into stable features. Furthermore, the influence of local noise on the retrieval results is reduced while the minimized structure of the global subspace is maintained.

Acknowledgement

2011 collaborative innovation center of electronic commerce and industry upgrading in Jiangxi Province – Research on large data platform.

References

[1] N. Arunkumar, K. R. Kumar, V. Venkataraman: *Automatic detection of epileptic seizures using new entropy measures*, Journal of Medical Imaging and Health Informatics, *6* (2016), No. 3, 724–730.

[2] R. Hamza, K. Muhammad, N. Arunkumar, G. R. González: *Hash based Encryption for Keyframes of Diagnostic Hysteroscopy*, IEEE Access (2017), https://doi.org/10.1109/ACCESS.2017.2762405

[3] D. S. Abdelhamid, Y. Y. Zhang, D. R. Lewis, P. V. Moghe, W. J. Welsh, and K. E. Uhrich: *Tartaric Acid-based Amphiphilic Macromolecules with Ether Linkages Exhibit Enhanced Repression of Oxidized Low Density Lipoprotein Uptake*, Biomaterials, *53* (2015), 32–39.

[4] N. Arunkumar, K. Ramkumar, S. Hema, A. Nithya, P. Prakash, V. Kirthika: *Fuzzy Lyapunov exponent based onset detection of the epileptic seizures*, 2013 IEEE Conference on Information and Communication Technologies, ICT 2013, (2013), art. No. 6558185, 701–706.

[5] J. J. Faig, A. Moretti, L. B. Joseph, Y. Y. Zhang, M. J. Nova, K. Smith, and K. E. Uhrich: *Biodegradable Kojic Acid-Based Polymers: Controlled Delivery of Bioactives for Melanogenesis Inhibition*, Biomacromolecules, *18* (2017), No. 2, 363-373.

[6] Z. Lv, A. Halawani, S. Feng, H. Li, S. U. Réhman: *Multimodal hand and foot gesture interaction for handheld devices.* ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), *11* (2014), No. 1s, 10.

[7] Y. Z. Chen, F. J. Tang, Y. Bao, Y. Tang, G. D. Chen: A Fe-C coated long period fiber grating sensor for corrosion induced mass loss measurement. Optics letters, *41* (2016), 2306–2309.

[8] N. Arunkumar, S. Jayalalitha, S. Dinesh, A. Venugopal, D. Sekar: *Sample entropy based ayurvedic pulse diagnosis for diabetics*, IEEE-International Conference on Advances in Engineering, Science and Management, ICAESM-2012, (2012), art. No. 6215973, 61–62.

[9] J. W. Chan, Y. Y. Zhang, and K. E. Uhrich: *Amphiphilic Macromolecule Self-Assembled Monolayers Suppress Smooth Muscle Cell Proliferation*, Bioconjugate Chemistry, *26* (2015), No. 7, 1359–1369.

[10] M. P. Malarkodi, N. Arunkumar, V. Venkataraman: *Gabor wavelet based approach for face recognition*, International Journal of Applied Engineering Research, *8* (2013), No. 15, 1831–1840.

[11] L. R. Stephygraph, N. Arunkumar: *Brain-actuated wireless mobile robot control through an adaptive human-machine interface*, Advances in Intelligent Systems and Computing, *397* (2016), 537–549.

[12] N. Arunkumar, V. Venkataraman, T. Lavanya: *A moving window approximate entropy based neural network for detecting the onset of epileptic seizures*, International Journal of Applied Engineering Research, *8* (2013), No. 15, 1841–1847.

[13] J. W. Chan, Y. Y. Zhang, and K. E. Uhrich: *Amphiphilic Macromolecule Self-Assembled Monolayers Suppress Smooth Muscle Cell Proliferation*, Bioconjugate Chemistry, *26* (2015), No. 7, 1359–1369.

Received May 7, 2017