



Research on local feature indexing of multimedia video based on intelligent soft computing

Zhongli Shen¹ · Yuguang Niu¹ · Yi Zuo¹ · Qiyue Xie¹

Received: 25 December 2018 / Revised: 13 March 2019 / Accepted: 9 May 2019 /

Published online: 1 October 2019

© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

Aiming at the problems of poor precision and recall, long retrieval time and high energy consumption in current video image indexing methods, a local feature indexing method for multimedia video based on intelligent soft computing is proposed. Video image is segmented by maximum entropy threshold method. Based on the result of segmentation, features are clustered in two-dimensional space. Each video image is divided into several feature groups. Unified descriptors are generated for each feature group. The descriptors of each feature group are coded by binary coding. The similarity between index items and video images in database is calculated, and local feature indexing of media video is realized by looking up tables. The experimental results show that the method has high index precision and recall, low energy consumption and real-time performance. The proposed method has excellent performance and robustness.

Keywords Intelligent soft computing · Multimedia video · Local features · Index · Maximum entropy threshold method

1 Introduction

As a very important resource of human society, information is making mankind enter a colorful high-tech era. Efficient access to and use of information resources have a great impact on human social life and work. How to find the information that people are interested in accurately in the vast data is an effective way to solve this problem [6, 12]. With the increasing innovation of science and technology, especially the rapid development of computer and multimedia technology, the expression mode of information data is gradually enriched, in which the amount of information data carried by image and video is larger than that of voice and text, and it has the

✉ Yi Zuo
yizuocs@aliyun.com

¹ Energy and Power Engineering College, Changsha University of Science and Technology, Changsha, Hunan 410000, People's Republic of China

advantages of intuition and vividness, thus, it determines that it will become one of the most critical communication channels for the public. Video is popular among all kinds of multimedia information expression modes, such as text, image and video, because of its vivid and specific advantages. Video has a huge amount of data and abundant content, and the number and types of video information are continually expanding [1, 13]. In summary, video data information management and retrieval has become a research hotspot in this field.

Reference [15] proposed to combine orthogonal decomposition with visual word bag model, and applied the results to image retrieval to divide the image information domain into encryption domain and retrieval domain, so that encryption processing and feature extraction could be independent of each other, to prevent the interaction between encryption and feature extraction process. Among them, users in the encryption domain could choose appropriate encryption methods according to their needs; in the retrieval domain, the visual word bag model was fused, and the visual word histogram was used to represent the image, thus reducing the semantic gap between the underlying features and the high-level semantics of the image, and improving the retrieval performance. The experimental results showed that the method was secure, but the accuracy was low [15]. In reference [14], for the problems existing in the current image feature retrieval methods, a weighted quantization method was applied to image target and background fusion based on the concept of connected particles, which combined the definition of new structuring elements with the image adaptive vector model to achieve image retrieval. In the process, nine new structural elements of image features were selected according to vision, and connected granular attributes and hierarchical statistical models were designed and constructed at the same time; corresponding mapping subimages were obtained by color conversion and structural element registration, and statistical structural elements and image connectivity feature vectors were extracted from subimages; each component was merged into a set of feature vectors based on vector fusion model, and composite vectors were applied to images retrieval. The experimental results showed that the method was stable, but it had the problem of long retrieval time [14]. Reference [2] proposed an image retrieval method based on Henon mapping. Based on the principle of image imaging and multi-band features, the gray value encryption of each band was realized by improved Henon mapping. According to the “big data” characteristics of images, the image feature vectors were designed and constructed by using the interval information of calculating gray values, and the image retrieval was completed by using similarity matching algorithm. The experimental results showed that the method had low running complexity, but low recall rate [2]. Reference [3] proposed a method to apply two-dimensional shape features of salient regions to image retrieval. Firstly, the more significant shape features were extracted from the image, and the image retrieval was realized by combining the texture and color features of the image. The experimental results showed that this method had certain accuracy, but it had the problem of high energy consumption [3]. Reference [4] proposed to apply the correlation descriptor of semi-circular local binary pattern frame to image retrieval. The definition of new semi-circular local binary pattern primitives was given. According to the given definition, the hierarchical structure primitives with different quantized color were detected. The spatial distribution of new structure primitives and image contrast characteristics were extracted, and image retrieval was completed according to the extracted results. The experimental results showed that the method was efficient, but it had the problem of low precision [4].

At present, the local feature index of multimedia video does not describe the feature group of multimedia video image, so it has the problems of low precision, long retrieval time and high energy consumption. In this paper, a set of visual features corresponding to each feature group are constructed as feature group descriptors, and SIFT descriptors are constructed by using visual word-bag model. The improved VLAD method is used to enhance the above

model, which makes the model more efficient. Based on the above theory, a local feature index method for multimedia video based on intelligent soft computing is proposed. This paper not only provides a new idea for local feature indexing of multimedia video, but also lays a foundation for the further development of local feature indexing technology of multimedia video, expands the application scope of multimedia field, and provides a simpler and more accurate way for multimedia video retrieval.

The method operation framework is as follows:

- (1) Video image is segmented by maximum entropy threshold method, which lays the foundation for local feature indexing of video.
- (2) Through intelligent soft computing, K-means algorithm, VLAD method, principal component analysis method and cosine similarity calculation method are introduced to complete local feature indexing of video.
- (3) The proposed method is validated by experiments and discussions.
- (4) The full text is summarized and suggestions are put forward for further research.

2 Local feature processing of multimedia video

In order to achieve high precision, low retrieval time and low energy consumption local feature index of multimedia video, a research on local feature index of multimedia video based on intelligent soft computing is proposed. The overall framework is as follows (Fig. 1):

2.1 Video image segmentation

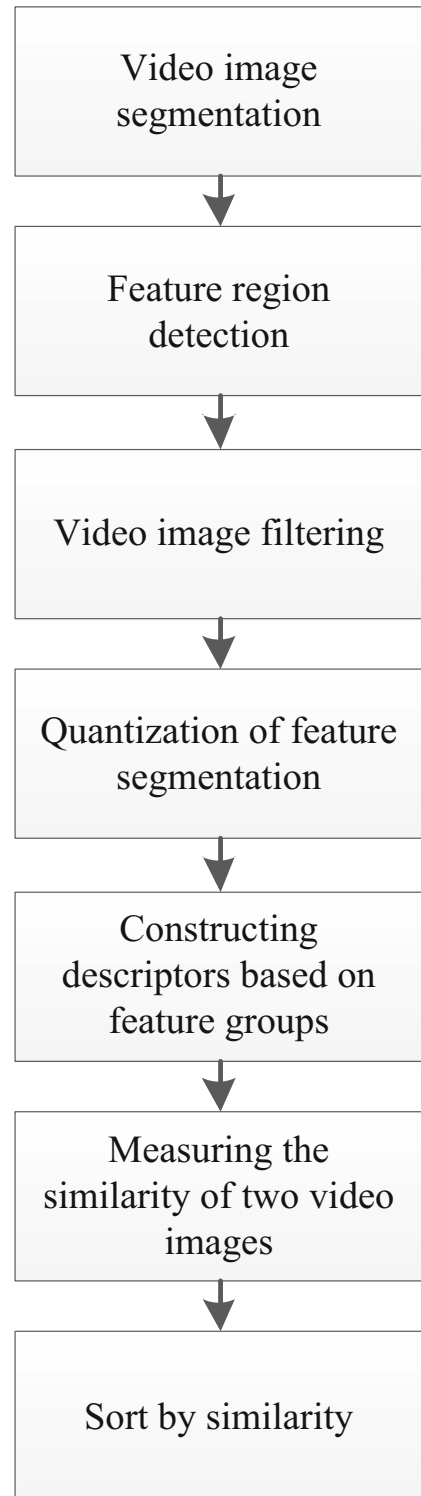
In order to improve the efficiency of video local feature indexing and enhance the accuracy of feature indexing to a certain extent, video segmentation is needed. Video is the technology of capturing, recording, transmitting and reproducing a series of static images in the form of electrical signals. In fact, continuous image changes exceed 24 frames per second. According to the principle of visual persistence, the human eye can not recognize a single static image, which shows smooth and continuous visual effects. This continuous image is called video [8]. To sum up, the video can be used as a continuous image for threshold segmentation. The average segmentation method is used to divide the whole video into multiple segments. The detailed segmentation process is as follows:

Using the definition of entropy in information theory to select segmentation threshold, one-dimensional maximum entropy threshold method can be obtained. The expression of entropy H in information theory is as follows:

$$H = -\int_{-\infty}^{+\infty} p(x) \lg p(x) dx \quad (1)$$

Where, $p(x)$ represents the probability density function that varies with the variable x . This paper applies the definition of entropy in information theory to image segmentation, uses entropy to determine a suitable threshold, divides the continuous image into two parts: target and background, and maximizes the amount of target and background information, so as to ensure the information integrity after video image segmentation [5, 11]. For gray image, the one-dimensional entropy of gray level is the largest, then a threshold t is selected to maximize the first-order gray level statistical information in the two parts of continuous image segmentation.

Fig. 1 Overall framework of local feature indexing method for multimedia video



Based on the definition of entropy, the histogram entropy expression of continuous image with K gray level is as follows:

$$H = - \sum_{i=0}^{K-1} p_i \lg p_i \tag{2}$$

In formula (2), p_i represents the probability of the occurrence of the i th gray level.

Set threshold be t , the target area O is constructed by the pixels whose gray level is higher than t in the continuous image, and the background area B is constructed by the lower part than t . The probability distribution of the two regions is: region $O:p_i/p_t, i=0, 1, \dots, t$ and region $B:p_i/(1-p_t), i=t+1, t+2, \dots, k-1$. Among them:

$$p_t = \sum_{i=0}^t p_i \tag{3}$$

According to the above calculation and analysis, the expression of entropy in the target and background region is as follows:

$$H_{O}(t) = - \sum_{i=0}^t \left(\frac{p_i}{p_t} \right) \cdot \lg \left(\frac{p_i}{p_t} \right) \tag{4}$$

$$H_{B}(t) = - \sum_{i=t+1}^{K-1} \left[\frac{p_i}{(1-p_t)} \right] \cdot \lg \left[\frac{p_i}{(1-p_t)} \right] \tag{5}$$

By using formula (4) and formula (5), the expression of entropy function of continuous image is as follows:

$$H(t) = H_{O}(t) + H_{B}(t) \tag{6}$$

The gray value corresponding to the maximum of the entropy function is the optimal segmentation threshold. Then

$$t^* = \arg \max H(t), 0 \leq t \leq K-1 \tag{7}$$

The optimal segmentation threshold obtained by formula (7) can realize the optimal segmentation of multimedia video.

2.2 Local feature indexing of multimedia video based on intelligent soft computing

The characteristic of intelligent soft computing is that it does not need to establish the exact mathematical or logical model of the problem itself, but directly processes the input data to get the result. It is more suitable for solving the problems that traditional technology can not effectively deal with, or even can not deal with.

On the basis of video image segmentation, the local feature indexing of multimedia video is realized by intelligent soft computing. Each video image is detected for its characteristic region. Assuming that the video image to be processed is $I(x', y')$ and $I(x', y')$ is filtered by Gaussian kernel function $G(x', y', \sigma)$, so that the expression $f(x', y', \sigma)$ and $f(x', y', \sigma)$ of scale space video image are obtained as follows:

$$f(x', y', \sigma) = G(x', y', \sigma) \odot I(x', y') \cdot t^* \tag{8}$$

Where, \odot represents the convolution operation. Heisen matrix operation is applied to the video image representation $f(x', y', \sigma)$ with scale features, which can detect the feature region coordinates (x', y') and scale σ information in the video image.

Video image feature detectors are constructed to detect some small video image feature regions. Based on these, a visual feature group is designed and constructed using the spatial location information of the feature regions in video images [10].

The feature areas detected in the video image I are marked as:

$$\left\{ f_i = (x'_i, y'_i) \right\}_{i=0}^{N-1} \tag{9}$$

Where, N represents the number of feature areas detected. In this paper, the standard K-means algorithm is used to classify the feature area into M' feature groups based on the spatial coordinate information of the feature area in I . The process can be expressed by formula (10):

$$D = \left\{ \min_{(\bar{x}'_j, \bar{y}'_j)} \right\} \sum_{j=0}^{M'-1} \sum_{(x', y')} \left\| (x'_i, y'_i) - (\bar{x}'_j, \bar{y}'_j) \right\|^2 \tag{10}$$

For each feature group, a radius value can be assigned to describe its scale. Its expression is as follows:

$$\bar{r}_j = \max_{(x'_i, y'_i) \in \eta_j} \sqrt{\left\| (x'_i, y'_i) - (\bar{x}'_j, \bar{y}'_j) \right\|^2} \tag{11}$$

In formula (10), η_j represents a set of feature vision. According to formula (11), image I can be represented as feature set $\left\{ (\bar{x}'_j, \bar{y}'_j, \bar{r}_j) \right\}_{j=0}^{M'-1}$, and each feature group corresponds to a set of visual features η_j . Some feature groups in video images overlap each other. The main reason is that the radius \bar{r}_j of the feature group is the feature distance which is the farthest from the center position (\bar{x}'_j, \bar{y}'_j) of the feature group.

The descriptors are generated according to the visual feature groups of the video image, and the descriptors generated by each visual feature group are analyzed. A group of visual features η_j corresponding to each feature group is designed and constructed as feature group descriptors. A unified vector description is constructed with Sift descriptor. Obviously, the visual word bag model can achieve this goal. Considering the complexity of the whole process and the energy consumption, the non-probabilistic simplified model of Fisher vector, namely the VLAD method, is used. And then the VLAD method is modified to make it more efficient.

Following is a brief analysis of the VLAD method, and then a detailed analysis of ways to make the method more efficient. Different from the quantification value in the visual word bag model and the number of visual features of visual words, the VLAD method retains the quantization residual vector between visual features and visual words, that is to say, a relatively large amount of information is recorded in the VLAD method. In each visual word, the residual vector record can be expressed by formula (12):

$$v_{k'} = \sum_{d_i \in C_{k'}} d_i - c_{k'} \tag{12}$$

In the formula, d_i represents the Sift descriptor of local feature f_i , $c_{k'}$ represents the k' th visual word in the visual codebook, and $v_{k'}$ represents all residual vectors in the visual word $c_{k'}$. The residual vectors of each visual word are connected to form the final VLAD descriptor $V = [v_{k'}]$.

In order to reduce the energy consumption of video feature indexing, a compact representation of visual feature group is obtained. The dimensionality of VLAD descriptor is reduced by PCA, as shown in Formula (13):

$$V^S = R^S \cdot (V - \bar{V}) \tag{13}$$

In the formula, R^S represents the constructed value of the S largest eigenvectors corresponding to the elements of learned principal component matrix, and \bar{V} represents the mean vector in the principal component analysis.

Generally, Sift descriptors differ in the degree of discrimination among dimensions, mainly because of the variance of each dimension. Experience shows that after balancing the variance of each dimension of the Sift descriptor, a relatively high degree of discrimination can be obtained [7, 9]. The process can be represented by formula (14):

$$v_{k'} = \sum_{d_i \in C_{k'}} \left[\frac{1}{\sqrt{\kappa_{k'}}} \right] \cdot R_{k'} \cdot (d_i - c_{k'}) \tag{14}$$

In the formula, $R_{k'}$ represents the matrix constructed by the eigenvectors in principal component analysis, and $1/\sqrt{\kappa_{k'}}$ represents the diagonal matrix formed by the square roots corresponding to all the features in $R_{k'}$. The descriptor generated by formula (14) is recorded as $BVLAD$.

After the unified descriptor is constructed by formula (14) for each feature group, formula (13) is used to reduce the descriptor to the S -dimension and encode the descriptor by piecewise quantization.

Among them, the process of feature segmentation quantization is: a descriptor $h' = V^S = [h'_0 \cdots h'_{S-1}]^T$ of S -dimension feature group is given, at the same time, it is divided into several lower dimensions, which can be recorded as l sub-vectors, $\left\{ h'_l = [h'_{l \cdot [S/l]} \cdots h'_{(l+1) \cdot [S/l]-1}]^T \right\}$. In each S/l dimension subspace, a visual codebook is constructed by using standard mean clustering algorithm, and a total of l visual codebooks $\{B_l^V\}$ are constructed. Then each descriptor can be expressed as an formula (15):

$$h' \approx [NN_0(h'_0) \cdots NN_{l-1}(h'_{l-1})]^T \tag{15}$$

Where, $NN_l(\cdot)$ represents the recent visual words in the existing visual codebook B_l^V . Assuming that the size of visual codebook B_l^V is K'_l , descriptor h' can be coded into $Z = \sum_l \lceil \log(K'_l) \rceil$ bits. To sum up, the Euclidean distance between the two descriptors is:

$$\|h' - h''\| \approx \sum_l \left\| NN_l(h'_l) - NN_l(h''_l) \right\|^2 \tag{16}$$

In the formula, h'' also represents a visual feature descriptor. Because distance $\left\| NN_l(h'_l) - NN_l(h''_l) \right\|^2$ can be calculated offline and stored in Euclidean distance table of

visual feature group descriptor, the distance between descriptor h' and h'' can be easily obtained online by looking up the table, so as to improve the recall of video feature indexing.

After designing and constructing M descriptors of visual feature group in video images, the similarity of two video images is measured based on the descriptors of visual feature group in two video images. An index video image $I_q = \{V_i^S\}_{i=0}^{M'-1}$ and a video image $I_d = \{V_j^S\}_{j=0}^{M'-1}$ in a video library are set up, and the similarity between them can be calculated by formula (17):

$$sim(I_q, I_d) = -\sum_i \min_j \left\| V_i^S - V_j^S \right\|^2 \tag{17}$$

In formula (17), the method of calculating the similarity between local features of video images can be regarded as the video image I_d to reconstruct indexing of video image I_q in video database. If the descriptor of feature group is quantized and coded by formula (16), the Euclidean distance between descriptors of two visual feature groups can be obtained by looking up tables.

Based on the above calculation and analysis, the index is implemented in large-scale databases by intelligent soft computing. Next, an index algorithm is introduced. First, the similarity between the index video image and the video image in the database is obtained according to the visual feature group. In this process, the inverted index mode is used to enhance the real-time and recall of the index. Formula (17) is used to calculate the similarity of the reconstruction errors of the returned video images, which are more advanced in partial ranking. The process can be expressed as:

$$sim^T(I_q, I_d) = \sum_i \sum_j \delta_{NN(V_i^S), NN(V_j^S)} \tag{18}$$

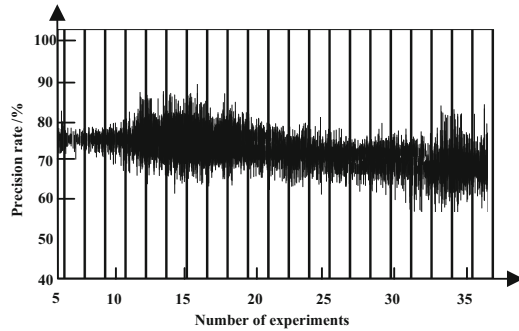
$$sim(I_q, I_d) = \begin{cases} -\sum_i \min_j \left\| V_i^S - V_j^S \right\|^2, & sim^T(I_q, I_d) \geq sim_0^T \\ -\infty, & \text{others} \end{cases} \tag{19}$$

In the formula, sim_0^T represents a threshold and δ represents a Kronecker function. According to the above expression, it is obvious that $sim^T(I_q, I_d)$ has a non-negative value. Formula (18) and Formula (19) represent the similarity between indexed video images and database video images, and $sim^T(I_q, I_d)$ can be obtained by inverted indexing. In the database, some video images whose similarity degrees $sim^T(I_q, I_d)$ is higher than the given threshold sim_0^T are selected, and then the similarity degree $sim(I_q, I_d)$ between video images is calculated accurately. In the local feature indexing of multimedia video based on intelligent soft computing, the video images and index images in the database are arranged according to the order of similarity $sim(I_q, I_d)$, and the video images satisfying the similarity $sim^T(I_q, I_d) < sim_0^T$ are sorted according to $sim^T(I_q, I_d)$, and the first one is regarded as the required video image, so as to complete the video's local feature indexing.

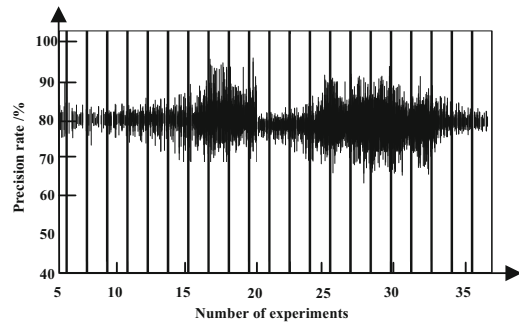
3 Experimental results and analysis

In order to verify the effectiveness of the local feature indexing method for multimedia video based on intelligent soft computing, an experimental platform is built on MATLAB. The experimental data are collected from the microblog database from October 1 to 30, 2018, and 2000 videos are selected as experimental objects. The video lasts about 2 min and the number of frames is about 3000. In order to improve the overall performance of local retrieval of multimedia video, it is necessary to denoise the video data before experiment.

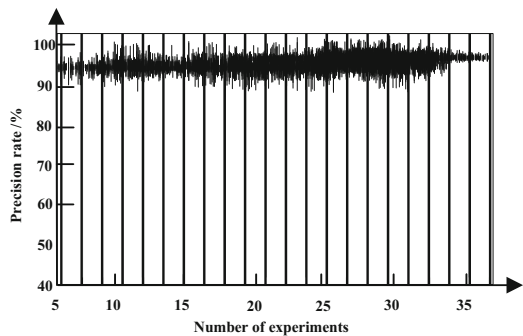
Document [15] method and document [14] method are compared with the local feature index method of multimedia video based on intelligent soft computing. Firstly, the index precision of the three methods is compared, and the higher the index precision is, the higher the accuracy of the local feature index of multimedia video is. Then the retrieval recall rate of the three methods is compared. The recall rate is the ratio of the local feature of the retrieved phase multimedia video to all the relevant local features in the multimedia video, and the recall rate of the retrieval method is measured. The higher the recall rate is, the more comprehensive the retrieval results are. Then the time consumption of the three methods is compared, which shows that the retrieval speed of



(a) Accuracy of indexing of the method in reference [14]

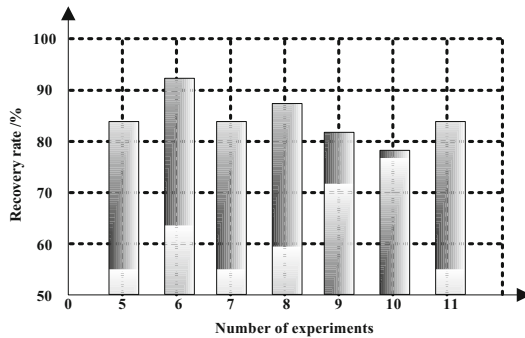


(b) Accuracy of indexing of the method in reference [15]

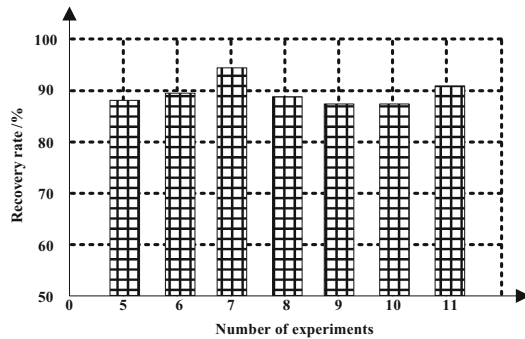


(c) Accuracy of Local Feature Indexing of Multimedia Video based on Intelligent Soft Computing

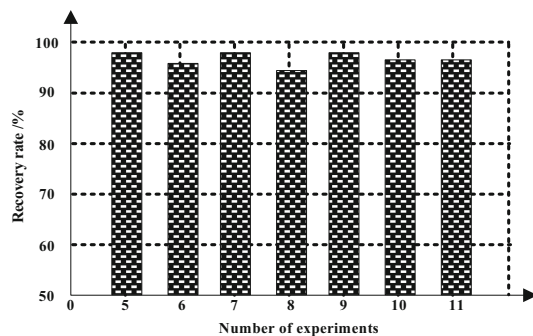
Fig. 2 Comparison of accuracy with different methods



(a) Recall rate of indexing of the method in reference [14]



(b) Recall rate of indexing of the method in reference [15]



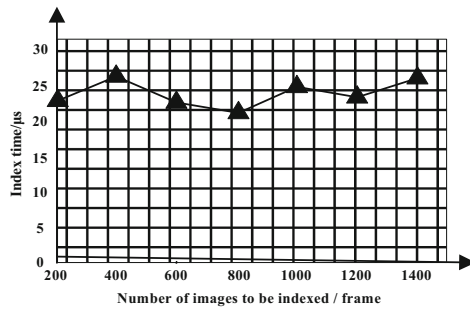
(c) Recall rate of local feature indexing of multimedia video based on intelligent soft computing

Fig. 3 Comparisons of recall rates of different methods

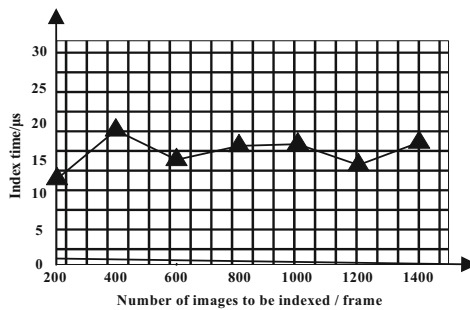
the three methods is faster. Finally, the energy consumption of the three methods is compared, and the lower the energy consumption, the smaller the fluctuation, which shows that the comprehensive performance of the method is better and more stable.

The experimental results are as follows:

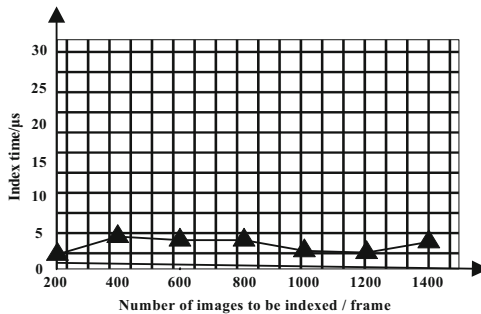
Analysis of Fig. 2 shows that although the average index precision of the method in reference [15] is about 75%, and the index precision curve fluctuates little, the precision curve of the method shows a downward trend with the increase of the number of experiments, and the reliability of the method is poor; Compared with reference [15], the average index



(a) Time-consuming of indexing of the method in reference [14]



(b) Time-consuming of indexing of the method in reference [15]



(c) Time-consuming of local feature indexing of multimedia video based on intelligent soft computing

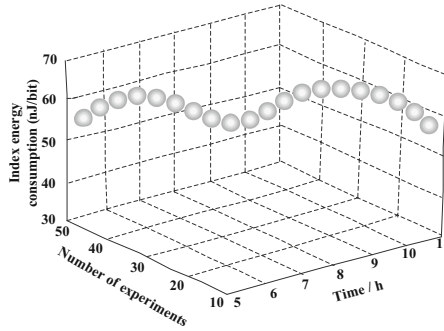
Fig. 4 Time-consuming comparison of different methods for indexing

precision of the method in reference [14] is about 80%. The index precision curve fluctuates greatly and has certain credibility; the index accuracy curve of local feature indexing method for multimedia video based on intelligent soft computing is the most stable. And the index accuracy is high, which can be maintained above 90%. The main reason for these differences is that the proposed method can segment multimedia video, improve the efficiency of video local feature index and improve the accuracy of feature index to a certain extent.

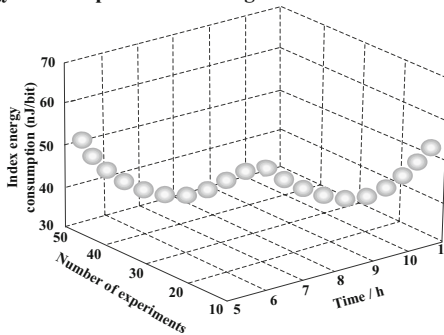
Recall rate is a reliable indexing to verify the comprehensiveness of retrieval. As can be seen from Fig. 3, the average retrieval recall rate of the methods in reference [15] is about 83%, the average retrieval recall rate of the methods in reference [14] is about 88%, and the average retrieval recall rate of the proposed methods is about 98%. And the recall rate of indexing by using the methods in reference [15] and reference [14] fluctuates little, and the local feature indexing method of multimedia video based on intelligent soft computing has higher overall

recall rate, which is more robust than the method in reference [15] and the method in reference [14]. The distance between feature descriptors of the proposed method can be obtained by online table lookup, which effectively improves the recall of video feature indexing.

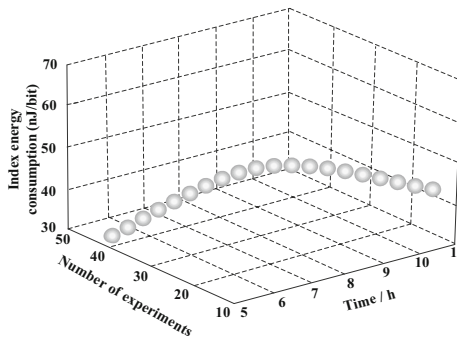
From Fig. 4, The time consumed by the method index in reference [15] is between 20 and 28 ugs, the time consumed by the method index in reference [14] is between 10 and 20 ugs, and the time consumed by the proposed method is between 1 and 6 ugs. It can be seen that the index time of the method in reference [15] and the method in reference [14] does not fluctuate greatly with the increase of the number of video images to be processed, but when the number



(a) Energy Consumption of Indexing of the Method in Reference [14]



(b) Energy Consumption of Indexing of the Method in Reference [15]



(c) Energy Consumption of Local Feature Indexing for Multimedia Video based on Intelligent Soft Computing

Fig. 5 Energy Consumption Comparison of Different Indexing Methods

of video images is 1400, the time-consuming curves of both methods show an upward trend, indicating that the practicability of the two methods is not very strong. When computing the cosine similarity between indexed video images and database video images, the local feature indexing method based on intelligent soft computing uses inverted indexing mode, which greatly improves the indexing efficiency of the proposed method and effectively reduces the indexing time.

Analysis of Fig. 5 shows that the energy consumption of the method index in reference [15] is between 43–62nJ/bit, and that of the method index in reference [14] is between 35–52nJ/bit, and that of the proposed method is between 30–42nJ/bit. The index energy consumption curve of the method in reference [15] is “M” type, and that of the method in reference [14] is “W” type. The energy consumption of the two indexing methods is not very stable under different running time and different experimental times. In order to reduce the energy consumption of video feature indexing, local feature indexing for multimedia video based on intelligent soft computing is used to reduce the dimension of feature descriptor, which reduces the energy consumption of indexing process and enhances the overall performance of the proposed method.

4 Discussion and exploration

In the discussion, whether the proposed method can guarantee the integrity of video information in the process of video image segmentation is analyzed. Discussions are made on the experimental platform. The reliability of the proposed segmentation method is verified by setting information feature points in the video image and the number of information feature points remaining in the image after segmentation by maximum entropy threshold method. The results of the discussion are as follows:

In Fig. 6, although the information integrity curve after video image segmentation is not very stable, it can be clearly seen from the image that the maximum entropy threshold method used to segment multimedia video images can ensure the information integrity of video images. This is mainly because in the process of video image segmentation, the definition of entropy in information theory is applied to image segmentation, and a suitable threshold is determined by using entropy. The continuous image is divided into two parts: target and background, which maximizes the amount of target and background information, thus ensuring the information integrity after video image segmentation.

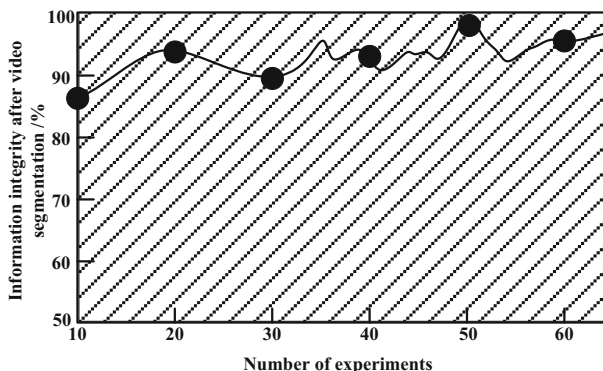


Fig. 6 Information Integrity after Video Image Segmentation

5 Conclusions

As the current research focus, video image retrieval is very important for the work and life of the public. Therefore, a local feature indexing method for multimedia video based on intelligent soft computing is proposed. Firstly, video image is segmented to facilitate video's local feature indexing, and then a series of intelligent soft computing is used to realize video's local feature indexing. Experiments and discussions prove that the proposed method is practical, and can provide reference for the development of this field. The following suggestions are put forward for the next research:

- (1) The most obvious feature of video is the huge amount of data. The next step is to introduce motion compensation or DCT transform into the indexing to further improve the real-time performance of the indexing.
- (2) Video is time-varying and dynamic. The next step is to integrate these two features into the indexing in order to adapt to the video's indexing transformation and improve the indexing performance.
- (3) Dynamic update indexing is a new technology based on the development of indexing technology. The next step is to introduce the current new concepts into video image indexing to meet the growing needs of users.

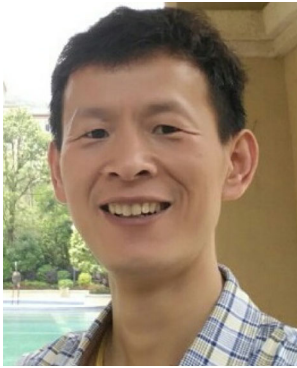
Acknowledgments This work was supported by National Key R&D Program of China (NO. 2017YFB0902100).

References

1. Gifford HC, Liang Z, Das M (2016) Visual-search observers for assessing tomographic x-ray image quality. *Med Phys* 43(3):1563–1575
2. Huang DM, Geng X, Wei LF et al (2016) A secure query scheme on encrypted remote sensing images based on Henon mapping. *Journal of Software* 27(5):1729–1740
3. Ke SW, Jin C, Zhu X (2017) Image retrieval based on two-dimensional shape features of salient region. *Journal of Guangxi University (Natural Science Edition)* 42(7):728–735
4. Li L, Feng L, Wu J et al (2016) Image retrieval based on semicircle local binary patterns structure correlation descriptor. *Journal of Dalian University of Technology* 56(1):532–538
5. Lu XY, Du LJ (2017) Fuzzy biological image feature extraction simulation research. *Computer Simulation* 34(2):397–400
6. Ming W, Ma J, Zhen Z et al (2016) Soft computing models and intelligent optimization system in electro-discharge machining of SiC/Al composites. *Int J Adv Manuf Technol* 87(4):1–17
7. Peng Y, Zhai X, Zhao Y et al (2016) Semi-supervised cross-media feature learning with unified patch graph regularization. *IEEE Trans Circuits Syst Video Technol* 26(3):583–596
8. Toliás G, Avrithis Y (2016) Image search with selective match kernels: aggregation across single and multiple images. *Int J Comput Vis* 116(3):247–261
9. Wang WH, Cheng B, Chen B (2017) The application research of histogram of SAR images processing. *Journal of China Academy of Electronics and Information Technology* 12(6):90–95
10. Xiang HY, Fu XW, Tian J et al (2016) Porosity evaluation for porous electrodes using image processing. *Chinese Journal of Power Sources* 40(8):572–574
11. Yan T (2016) Application analysis of graphic image processing in media communication. *Automation & Instrumentation* 75(4):212–213

12. Yan Y, Liu G, Wang S et al (2017) Graph-based clustering and ranking for diversified image search. *Multimedia Systems* 23(1):41–52
13. Yang X, Gao X, Song B et al (2018) ASI aurora search: an attempt of intelligent image processing for circular fisheye lens. *Opt Express* 26(7):7985–8000
14. Yu LX, Feng L, Zhang J et al (2016) An image feature extraction method based on adaptive fusion of object and background. *Journal of Computer-Aided Design & Computer Graphics* 28(6):1250–1259
15. Zhao X, Xu YY, Gong JY et al (2018) Image security retrieval method combining orthogonal decomposition and BoVW. *J Appl Sci* 36(2):299–308

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Zhongli Shen (1976-) received the PhD degrees in control science and engineering at North China Electric Power University, China, 2018. Now he is a assistant professor in Changsha University of Science and Technology. His scientific interests includes image processing, digital signal processing, intelligence algorithm design, and industrial process control.



Yuguang Niu (1964-) received his Ph.D degrees in Dynamic Engineering and Engineering Thermo-physicses from North China Electric Power University, Beijing, China, in 1997. He is currently a Professor with North China Electric Power University. He is also an Associate Director for the State Key Laboratory for Alternate Electric Power System With Renewable Energy Source. His research interests include machine vision inspection systems, image processing, stability analysis of combustion in power plants boiler by using flame image.



Yi Zuo (1982-) received the Ph.D degrees in control science and engineering at Hunan University, China, 2009. He was also a visiting student in University of Waterloo, Canada, in 2009. Now he is a assistant professor in Changsha University of Science and Technology. His scientific interests includes artificial intelligence, robust control, and nonlinear analysis.



Qiyue Xie (1980-) received his Ph.D degrees in Control Theory and Control Engineering from Shanghai Jiao Tong University 2010. He is currently an associate professor at Changsha University of Science & Technology. His research interests are focused on nonlinear system theory, intelligent algorithms.