Search by Image. New Search Engine Service Model

Conference Paper · October 2018

DOI: 10.1109/INFOCOMMST.2018.8632117

Search by Image. New Search Engine Service Model

Kirill Smelyakov
Kharkiv National University of Radio Electronics
Kharkiv, Ukraine, Nauka Ave, 14
kirillsmelyakov@gmail.com

Igor Ruban
Kharkiv National University of Radio Electronics
Kharkiv, Ukraine, Nauka Ave, 14
ihor.ruban@nure.ua

Denys Sandrkin
Kharkiv National University of Radio Electronics
Kharkiv, Ukraine, Nauka Ave, 14
denys.sandrkin@nure.ua

Martovytskyi Vitalii
Kharkiv National University of Radio Electronics
Kharkiv, Ukraine, Nauka Ave, 14
martovytskyi@gmail.com

Yury Romanenkov
National Aerospace University "Kharkiv Aviation Institute"
Kharkiv, Ukraine, Chkalova 17
KhAI.management@ukr.net

***Abstract* — Every year millions of photos and images appear in the Internet. Most of them are downloaded to personal cloud storage or become available to the public. With such a huge amount of information, the need for an effective search in the picture ripens. And if excellent tools have already been created for text search, image search remains an unresolved problem. The purpose of this publication is to develop a model for creating an effective image search service.**

Keywords—photo; Internet; search engine; search algorithm; artificial intelligence; neural network; image feature; image description; histogram; texture components; Haar features

I. Introduction

Nowadays, the total number of photos (images) uploaded daily to various online storages is estimated at millions; Only Instagram daily adds 52,000,000. In the presence of such a huge amount of information, the problem of searching for images not only by keywords, but also by the copy of the image (reverse image search) is very significant.

At the moment, the most known search engines (Search Engine), which provide the ability to reverse search the image, are Google, Bing and TinEye. In the series of experiments of image search from the test set, it was established that these services work with different efficiency.

In the experiment we used a test set of images found on the Internet. Images of famous places, people, animals, paintings were presented in the set. The search was done in the browser, with which images were found for reverse searching, and in a browser with a clean history and incognito mode turned on.

The "Search by image" service from Google well shows itself in recognizing the scene on the image, recognizing the place, or personality, but not always [1]. The downside is that the search for similar images depends on the result of recognizing the scene of the image and sometimes can produce completely incorrect results (Fig. 1). Also, the result is influenced by the presence of the image in the search history and the changes made to the image such as cropping, etc. When trying to recognize an image that already existed in the search history, there is only a general description of the picture, namely "Women, paintings " (Fig. 2), and when searching in another browser in the history of which this image is not present (Fig. 3), the search result is author's name.

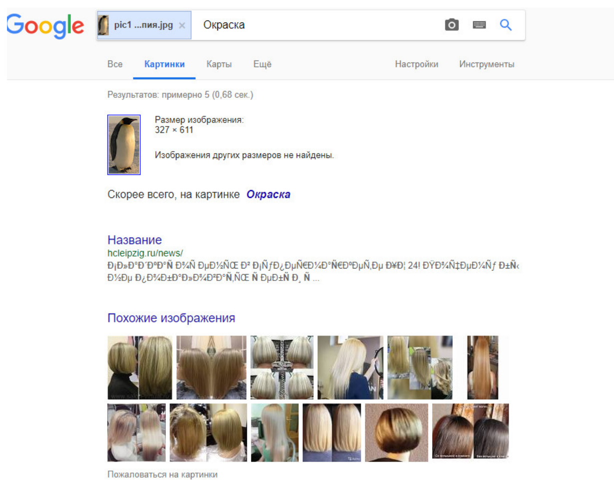


Fig. 1. Incorrect recognition of the object on the image (the penguin is recognized as a curl of hair)

978-1-5386-6611-1/18/$31.00 ©2018 IEEE

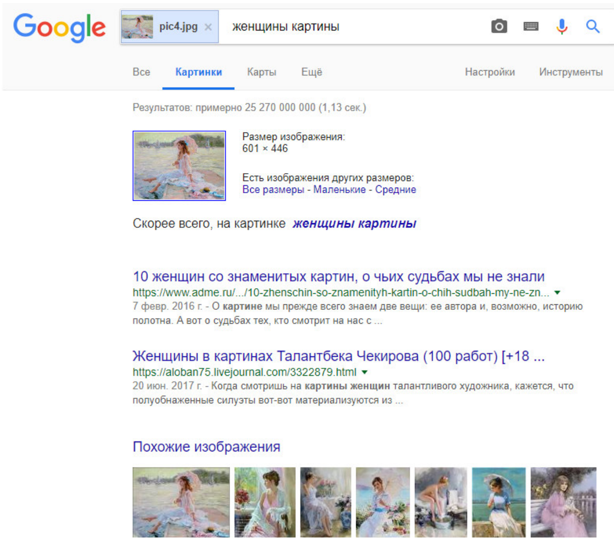


Fig. 2. Search result of an image that is in the search history.

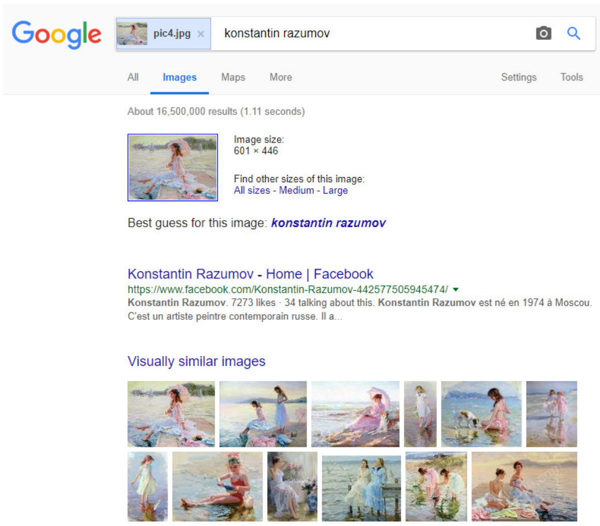


Fig. 3. Figure 3 - Image search result that is not in the search history

The "Bing Image Feed" service allows you to find sites on which the image we are looking for is present. With the task of finding sites it does a good job (Fig. 4), but when changes to the original image are done (cropping), the sites are not found. However, the functionality of searching for a similar image (Fig. 5) works.

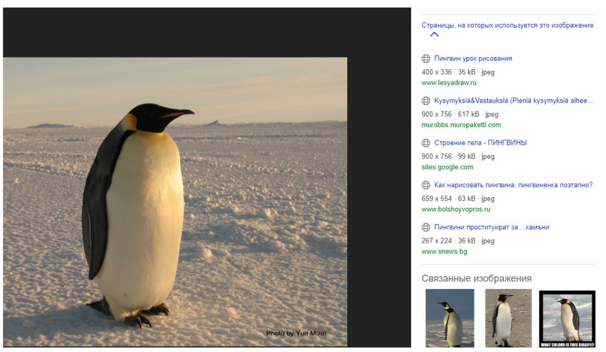


Fig. 4. Image search result in the "Bing Image Feed" service

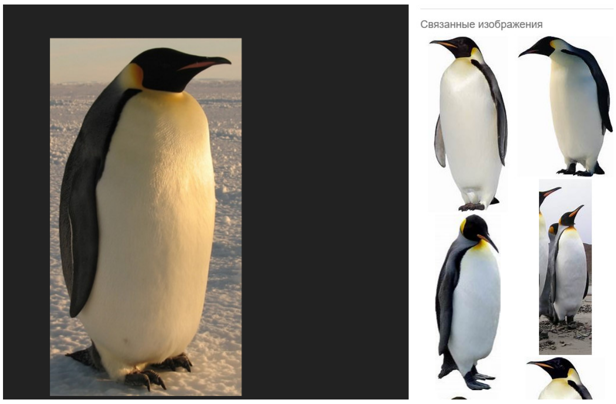
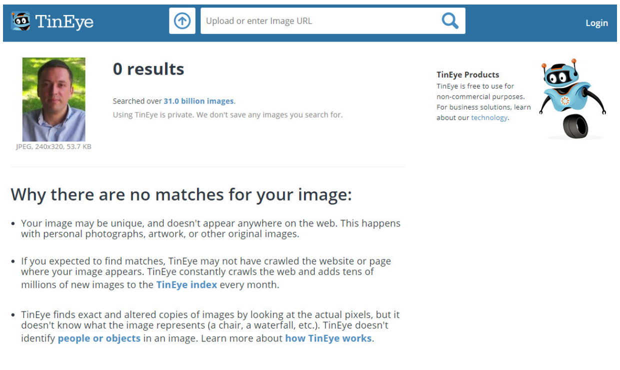
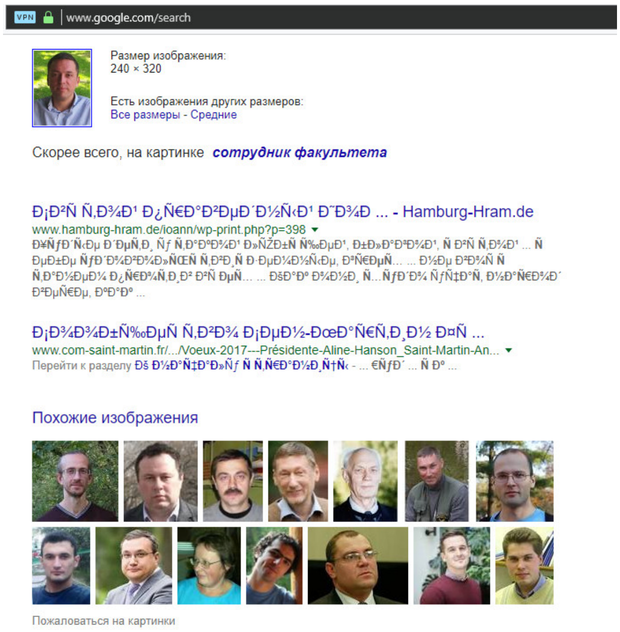


Fig. 5. The result of the search of the cropped image

TinEye does an excellent job of finding sites that contain the image you are looking for, even if you crop it, but its base is not as extensive as Google's, and in most cases the image that is not in TinEye can be found using Google Search by Image (Fig. 6); in this example, the image was found in the search history.



a)



b)

Fig. 6. Result of searching for the same image in TinEye (a) and Google Search by Image (b)

The carried-out experiment allows to draw a conclusion about the imperfection of modern systems of image search [2] The concept and the main ways of creating an effective service are offered further in the work.

II. THE PROPOSED APPROACH

An effective solution to the image search problem requires the development of a global distributed service "Search by Image", which uses a comprehensive approach and is based on the latest achievements of knowledge engineering in the areas of: computer vision, data mining, database organization, algorithms and data structures, expert systems and machine learning.

Search by image is based on the use of computer vision algorithms [3]. The success of such a search is based on the use of informative invariant rapidly computed features, as well as effective image classifiers.

Therefore, an effective search by image should be based on the data mining approach of lexical description of image context, which is called "Search by Keywords", image metadata, and calculated image features.

At this, keywords and metadata are mainly useful for operational narrowing of image search area. For this purpose it’s also useful to consider history and geolocation data.

Search effectiveness, in the first place, depends on the time complexity of computing the image features [4], and on the degree of image data structuredness [5]. Therefore, all search computations must be minimized. For this, when the image appears in the WEB for the first time, all required values of its features must be calculated and stored together with other elements of the image description. And to remain in a structured form, that maximizes the effectiveness of image search. To do this, it is required to develop a database model that is designed to store images links and contextual data about them.

The development of an images database is the first half of the matter. This is foundation. Based on this, data structures and image search algorithms are developed, taking into account the principles of Search Engine Optimization.

Due the constant changes In the WWW there is necessity to constantly adapt to the possible variations of user requests, search conditions and WEB modes. In this situation, the concept of using Artificial intelligence – is the continuous optimization of a search service by developing and applying the algorithms of: 1) expert evaluation and 2) machine learning.

Also the common methods and functions of description of image content should be used.

- Comparison of color content;
- Comparison of texture components;
- Image definition based on the geometric shape of objects;
- Use a SIFT-like signature;
- Use a perceptive hash;
- Use of Haar features;
- Use of artificial neural networks.

A. Comparison of the image using color content

Searching for images by comparing the color components is done by plotting the histogram (Fig. 7) of their distribution. At the moment, studies are provided to construct a description in which the image is divided into regions according to similar color characteristics, and their mutual arrangement is then taken into account. The description of the images with the colors that contain them is the most common, since it does not depend on the size or orientation of the image. The construction of histograms with their subsequent comparison is used often but is not the only way to describe the color characteristics.

The advantages are simplicity of implementation, speed. The main disadvantage of the method of color histograms is that it loses information about the spatial location of objects. Absolutely different pictures can have similar color histograms.

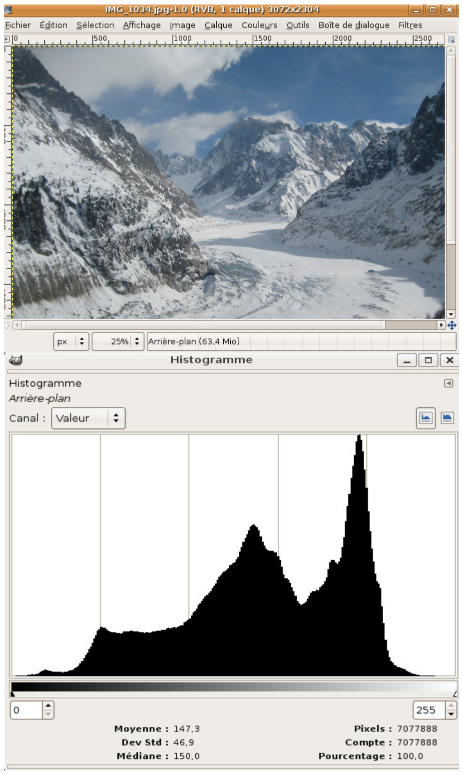


Fig. 7. Histogram

B. Comparison of texture components

The methods of such a description work by comparing the texture patterns present in the image and their relative positioning. To determine the texture, patterns that are grouped into sets are used. They contain not only information describing the texture, but also its location in the described image. Texture as an entity is difficult to describe formally, and it is usually represented as a two-dimensional array of brightness changes. Also, in the description sometimes include a measure of contrast, gradient orientation, regularity [6].

The main disadvantage is the lack of a clear criteria of the difference\proximity of two texture images.

C. *Image definition based on the geometric shape of objects.*

The description of the form implies the description of the geometric shape of the individual regions of the image [7]. To determine this, segmentation is first applied to the region (Fig. 8) or border selection (Fig. 9).

The main drawback is the need for frequent human intervention, since methods such as segmentation are difficult to fully automate for a wide variety of tasks.

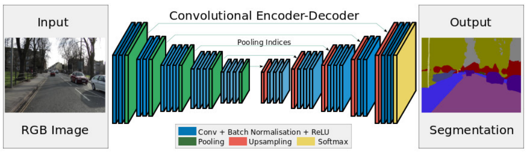


Fig. 8. Image segmentation process

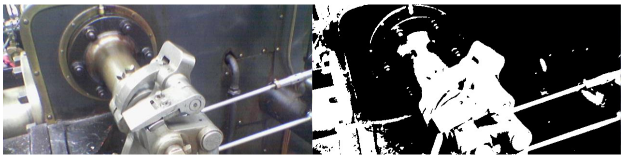


Fig. 9. Borders highlighting

D. *Use a SIFT-like signature*

The SIFT (scale-invariant feature transform) method determines a set of control points on the image and uses the relative position of the anchor points as a characteristic of the picture. This method can reliably identify objects even through noise or partially overlap, because the attribute descriptor is invariant to linear scaling as well as changes in spatial orientation or light, etc.

The advantage of this method is that it is well suited for determining whether one image is a deformed copy of another - however, to determine the similarity of two fundamentally different, although visually similar images, the method is not suitable (for example, two photographs of the same kitten in different poses will have little in common for such method). Therefore, in the search system, it can only play the role of determining whether there is a modification of the desired image in the database, but it cannot find similar ones in any other sense.

E. *The use of a perceptive hash.*

In this approach, using the discrete cosine transform, the so-called lower frequencies are left, in which more information on the shape of the image is concentrated than on its color characteristics. As a result, a large image turns into a 64-bit hash.

The algorithm of finding of hash:

- Decrease the size. In images, high frequencies provide detail, and low frequencies show the structure. A large, detailed photo contains a lot of high frequencies. In a very small picture there are no details, so it consists entirely of low frequencies. In the case shown in Fig. 10, the size is reduced to 8x8, so the total number of pixels is 64. You do not need to worry about the proportions and just crop it to a square eight by eight. Thus, the hash will match all the image options, regardless of the size and aspect ratio.

Fig. 10. Result of image size reduction

- Remove the color. A small image is converted to grayscale, so the hash is reduced in three times: from 64 pixels (64 values of red, 64 green and 64 blue) to just 64 color values.
- Find the mean. Calculate the average for all 64 colors.
- A chain of bits. For each color, we get 1 or 0, depending on whether it is more or less than average.
- Building a hash. Translate 64 separate bits into one 64-bit value. Order does not matter if it is kept constant.

Result: 8f373714acfcf4d0. The resulting hash will not change if the image is scaled, compressed or stretched [8]. Changing the brightness or contrast, or even manipulating colors, will not greatly affect the result. And most importantly: such an algorithm allows you to quickly get the result.

When comparing two pictures, you need to build a hash for each of them and count the number of different bits (Hamming distance). Zero distance means that it is most likely the same pictures (or variations of one image). Distance 5 means that the pictures are somewhat different, but on the whole, they are still pretty close to each other. If the distance is 10 or more, then this is probably a completely different image [9].

The advantages of this method include the speed of determination and the simplicity of the algorithm. The disadvantage is the reduction of the image to low frequencies and this gives a significant error in determining the result.

F. *Use of Haar features*

Characteristics of the digital image used in pattern recognition. Their name is due to an intuitive resemblance to the Haar wavelets. The Haar attributes were used in the first real-time face detector.

The Haar feature consists of adjacent rectangular areas. They are positioned on the image, then the intensities of the pixels in the areas are summed, after which the difference between the sums is calculated. This difference will be the value of a certain characteristic, a certain size, in a certain way positioned on the image (Fig. 11).

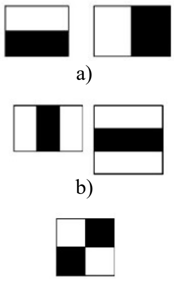


Fig. 11. Haar features: Edge (a), Line (b) and Four-rectangle (c)

For example, consider a database with human faces. Common to all images is that the area in the eye area is darker than the area around the cheeks. Therefore, the common hallmark of Haar for individuals is the two adjacent rectangular regions lying on the eyes and cheeks.

The simplest rectangular Haar feature can be defined as the difference of the sums of the pixels of two adjacent regions within a rectangle that can occupy different positions and scales in the image. This kind of feature is called 2-rectangular. Each feature can show the presence (or absence) of any particular characteristic of the image, such as borders or texture changes. For example, a 2-rectangular sign can show where the border between the dark and light regions lies (Fig. 12).

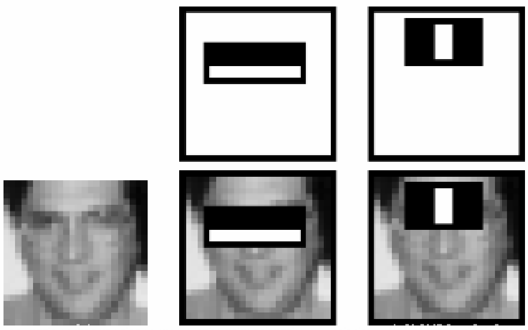


Fig. 12. Example of Haar features extraction

The key advantage of Haar's signs is the highest speed, in comparison with the remaining featuares. When using the integral image representation, the Haar attributes can be calculated with a constant time.

G. *Use of artificial neural networks*

At a qualitatively different level, the task of classifying images began to be solved in 2013. Then the ImageNet data set broke through the barrier in 15% of the classification errors of thousands of object types. Since then, for 5 years, a lot of different models of neural networks have been designed and trained, and a barrier in 5% of errors has been breached. The most successful of them are: VGG16, ResNet50, Inception, GoogLeNet and many others most of them are built on the basis of convolutional neural networks [10].

The schematic architecture of VGG16 is shown in Fig. 13.

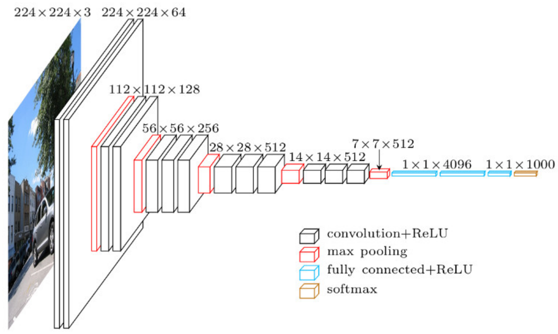


Fig. 13. Schematic architecture of VGG16

The layers of a neural network on the image consist of a set of different convolution-filters. [11] Each of the filters is responsible for finding a particular pattern, and when it finds a portion of the image that has this pattern, the filter sends a signal to the next layer. In turn, the signals of the previous layer form a new image for the next layer. Figure 11 of the VGG16 architecture shows that first there was a color RGB image of 224x224 pixels with 3 channels (red, green, blue). Then, after passing through the first layer of the bundle, we got an image of 224x224 pixels with 64 channels. These channels already represent not colors, but the results of each of the 64 filter-packages. And so on, up to an image of 7x7 pixels with 512 channels.

Building cascades of convolutional layers and teaching the model, we obtain layers containing the abstractions of images [12]. The first layers in themselves can contain small details: lines. Next come combinations of details - figures. The following layers can already contain forms, and in the end the whole objects (Fig. 14).

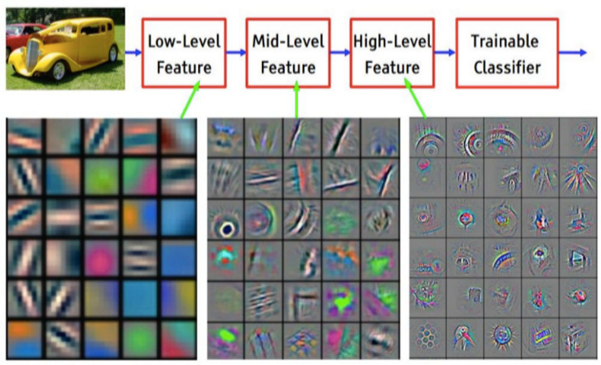


Fig. 14. Expanding the image into abstract layers

Another interesting feature of the convolutional layers in this model should be noted: each next layer is "thicker", since it has more filters, but "smaller", since the image is specially reduced by the MaxPooling operation [13]. Use this technique for the following reason: the fact of detecting a certain feature-object is more important than knowing the exact location of this object in the image [14]. That's why they take a maximum inside a small window, thereby creating a map of the location of the signs (Fig. 15).

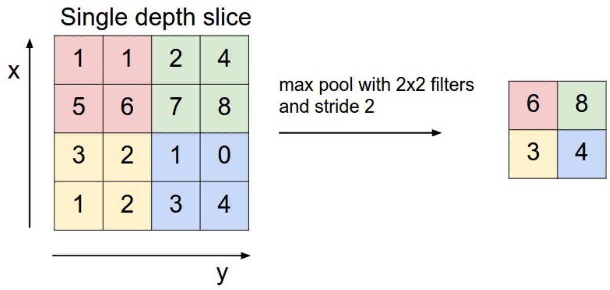


Fig. 15. The principle of creating location maps of features

Closer to the exit of the model, we have a small image - a map of signs of a size of 7x7 pixels with 512 filters. On this three-dimensional map it is still impossible to make predictions of the classes of objects in the image - a cat or a dog. In order to go already to the class predictions, this map is laid flat on the plane using the Flatten operation and connected to a fully connected hidden layer of 4096 neurons. And then the classical scheme: another hidden layer with 4096 neurons and an output layer with 1000 neurons, each of which gives the probability of belonging to one of the 1000 classes in the ImageNet problem.

The advantages are greater flexibility in customization and recognition accuracy. The disadvantage is the need for a large set of test samples and the probabilistic result of the comparison.

III. FUTURE RESEARCH

The next step of research is experimental approval of the proposed model efficiency and improvement of existing feature extraction algorithms. The final aim is creating of efficient search engine.

IV. CONCLUSION

The problem of searching by images is very urgent. An analysis of image estimation methods shows that the problem is complex, but solvable. At the moment there are several services that offer different solutions. Some of the services use user search history to find image, so the search result will be different for different users. Also, all services work badly with modified images. To find a way out of this situation, a model of service for searching by the image is suggested, which can allow to achieve a sufficient reliability of image search. The main components of this model:

- Computer vision. Search by image is based on the use of computer vision algorithms. The success of such a search is based on the use of informative invariant rapidly computed features, as well as effective image classifiers.
- Data mining. An effective search by image should be based on the data mining approach of lexical description of image context, which is called "Search by Keywords", image metadata and calculated image features.
- Database organization. Search effectiveness, in the first place, depends on the time complexity of computing the image features, and on the degree of image data structuredness. When the image appears in the WEB for the first time, all required values of its features must be calculated and stored together with other elements of the image description.
- Algorithms and data structures. Data structures and image search algorithms are developed, taking into account the principles of Search Engine Optimization. Speaking of data structures, it's mainly concerned with trees and hash tables for image search.
- Expert systems and machine learning. The concept of using Artificial intelligence – is the continuous optimization of a search service by developing and applying the algorithms of expert evaluation and machine learning.

REFERENCES

[1] L. Manso, C. Duque, P. Ribeiro, “Power Quality Waveform Recognition Using Google Image Search Engine (iPQ-Google),” International Conference on Harmonics and Quality of Power (ICHQP), IEEE, 2016, pp. 1010-1013.

[2] K. Sankar, G. N. K. Suresh Babu, “Implementation of Web Image Search Portal with automated concealed annotation,” Communication and Electronics Systems (ICCES), IEEE, 2016.

[3] S. Mashtalir, V. Mashtalir, “Sequential temporal video segmentation via spatial image partitions,” Data Stream Mining & Processing (DSMP), IEEE First International Conference, IEEE, 2016, pp. 239-242.

[4] D. Kinoshenko, S. Mashtalir, V. Shlyakhov, M. Stolbovyi, “Video Shots Retrieval with Use of Pivot Points,” International Conference on Theory and Applications of Fuzzy Systems and Soft Computing, Springer, Cham, 2018, pp.102-111.

[5] S. K. Pawaskar, S. B. Chaudhari, “Web image search engine using semantic of Images's meaning for achieving accuracy,” Automatic Control and Dynamic Optimization Techniques (ICACDOT), IEEE, 2017, pp. 99-103.

[6] S. Bogucharskiy, S. Mashtalir, "Image sequences texture analysis based on vector quantization," Radio Electronics, Computer Science, Control, no. 2, 2014, pp. 94-99.

[7] Zhengbing Hu, S. Mashtalir, O. Tyshchenko, M. Stolbovyi, “Video Shots’ Matching via Various Length of Multidimensional Time Sequences,” vol. 9, International Journal of Intelligent Systems and Applications, Modern Education and Computer Science Press, 2017, pp. 10.

[8] D. Tian, D. Tao, “Global hashing system for fast image search,” vol. 26, no. 1, IEEE Transactions on Image Processing , IEEE, 2016, pp. 79-89.

[9] Yuan Cao, Heng Qi, Jien Kato, Keqiu Li, “Hash Ranking With Weighted Asymmetric Distance for Image Search,” IEEE Transactions on Computational Imaging, IEEE, vol. 3, no. 4, Dec. 2017, pp. 1008-1019.

[10] S. Mashtalir, O. Mikhnova, “Detecting Significant Changes in Image Sequences,” Multimedia Forensics and Security, Springer, Cham, 2017, pp. 161-191.

[11] Dawei Liang, Ke Yan, Wei Zeng, Yaowei Wang, Qingsheng Yuan, Xiuguo Bao, “Deep hashing with mixed supervised losses for image search,” IEEE International Conference on Multimedia and Expo Workshops, IEEE, 2017, pp. 507-512.

[12] M. Mohmmadi Kashani, S. Hamid Amiri, “ Leveraging deep learning representation for search-based image annotation,” Artificial Intelligence and Sigral Processing Conference, IEEE, 2017, pp. 156-161.

[13] Jun-yi Li, Jian-hua Li, “Fast image search with deep convolutional neural networks and efficient hashing codes,” Fuzzy Systems and Knowledge Discovery (FSKD), IEEE, 2015, pp. 1285-1290.

[14] Weicheng Sun, Songhao Zhu, Yanyun Cheng, “Image search via semantic hashing learning,” Control And Decision Conference (CCDC), IEEE, 2017, pp. 1986-1990.