ISSN (ONLINE):2456-5717
International Journal of Advanced Research in Basic Engineering Sciences and Technology (IJARBEST)
Vol.4, Issue.4, April 2018

Improvising the Efficiency of Image Search Based on Descriptors in Big Data Analytics

[1] Sasikala C ,[2] Priyanka S , [2] Rohini M , [2] Sathyapriya E, [2] Sumithra D.
[1]Assistant Professor, Department of Computer Science and Engineering,
[2] UG Scholars, Department of Computer Science and Engineering,
VSB College of Engineering Technical Campus, Coimbatore, Tamil Nadu – 642 109.

Abstract - Digital images have many applications in different fields like medical imaging and diagnostics, weather forecasting, space research, military etc. The number of images available and their wide variety increases with the ease of acquiring, storing and sharing digital images due to the advances in technology. As a result, the significance of image retrieval algorithms and systems has been considerably increased. Many researches on content-based image retrieval (CBIR) are being carried out. In this paper, a fast image retrieval algorithm called feature levels is proposed. Feature levels algorithm works with the classification of image features to different categories or levels, feature extraction in terms of levels and feature similarity comparison of the query image with database images. The system retrieves images from the associated database. The database is re-written after each level according to Database Revision (DR) algorithm.

Index Terms: **Local and Global features, Database Revision Algorithms, Distance Matching Algorithm, Indexing, CBIR.**

I. INTRODUCTION

A large number of images are available on the internet. Efficient and effective retrieval system is needed to retrieve these images by the contents of the images like color, texture and shape. This system is called content based image retrieval (CBIR). CBIR is an intensive and difficult area of research.

CBIR is performed in two steps: indexing and searching. In indexing step contents (features) of the image are extracted and are stored in the form of a feature vector in the feature database. In the searching step, user query image feature vector is constructed and compared with all feature vectors in the database for similarity to retrieve the most similar images to the query image from the database.

Availability of the huge number of images due to the rapid development and improvement of the internet, image capture devices and computer hardware cause the problem of storage and manipulation of images. To overcome the problem of space and manipulation time, at present almost all images are represented in compressed formats like JPEG and MPEG.

The features of image can be extracted directly from the compressed domain. To extract the low level features from the compressed images, first the images are decoded from the compressed domain to the pixel domain. After that, image processing and analysis methods are applied to images in the pixel domain. This process is inefficient because it involves more computations and increases the processing time. The low level texture features are extracted from 8×8 DCT transformed blocks using DC and AC coefficients in nine different directions which represent nine feature vectors and grayscale level distribution in the image.

II. LITERATURE SURVEY

A. A Fast Image Retrieval Method Designed for Network Big data

In 2017 Bin Jiang and Baihua Li et.al proposed the feature extraction algorithm. In the field of big data applications, image information is widely used. The value density of information utilization in big data is very low, and how to extract useful information quickly is very important. So should transform the unstructured image data source into a form that can be analyzed. In this paper, proposed a fast image retrieval method which designed for big data. First of all, the feature extraction method is necessary and the feature vectors can be obtained for every image. Then, it is the most important step to encode the image feature vectors and make them into database, which can optimize the feature structure. Finally, the corresponding similarity matching is used to determine the retrieval results. There are three main contributions for image retrieval in this paper. New feature extraction method, reasonable elements ranking, and appropriate distance metric can improve the algorithm performance. Experiments show that the method has a great improvement in the effective performance of feature extraction and can also get better search matching results.

B. An Efficient content based Image retrieval using EI classification and color Features

In 2014 M. Yasmin and M. Sharif et.al proposed the corner matching technique. An efficient method for image search and retrieval has been proposed in this study. For this purpose images are decomposed in equal squares of minimum 24x16 size and then edge detection is applied to those decomposed parts. Pixels classification is done on the basis of edge pixels and inner pixels. Features are selected from edge pixels for populating the database. Moreover, color differences are used to cluster same color retrieved results. Precision and recall rates have been used as quantification measures. It can be seen from the results that proposed method shows a very good balance of precision and recall in minimum retrieval time, achieved results are comprised of 66%-100% rate for precision and 68%-80% for recall.

C. An Improved Hash based Radio Frequency Identification (RFID) Two way Security authentication protocol and Application in Remote Education

In 2016 Lvqing Yang and Qingqiang Wu et.al proposed the Burrows -Abadi – Needham (BAN) logic. The increasingly expanding application of RFID technology, the security and privacy issues of RFID system has also been a widespread concern. The existing RFID security protocol can't simultaneously meet the requirements of low-cost, low computing, high efficiency and high security. In this paper, through the analysis of common principles and shortcomings of RFID security protocols based on Hash Function, improve the security protocols, so as to achieve the objectives of two-way authentication. This paper theoretically proves the protocol security through BAN logic. And the new protocols can effectively solve the security and privacy problems such as replay attack, fake attack, location privacy, anonymous tags and so on. The paper proposes a kind of authentication and communication security mechanism that makes full use of the functions of the conditional Access Module (CAM) which exist in the original remote education system and when authenticating the nodes of the system, use the proposed security scheme. Analysis showed that the mechanism is safe, reliable, strong compatibility, economic and applicable.

D. Beginners to content based image retrieval

This paper gives an overview idea of retrieving images from a large database. CBIR is used for automatic indexing and retrieval of images depending upon contents of images known as features. The features may be low level or High level. The low level features include color, texture and shape. The high level feature describes the concept of human brain. The difference between low level features extracted from images and the high level information need of the user known as semantic gap. A Single feature can represent only part of the image property. So multiple features are used to enhance the image retrieval process. This paper has used color histogram, color mean, color structure descriptor and texture for feature extraction. The feature matching procedure is based on their Euclidean distance.

E. Color Image coding and indexing using BTC

In 2003 Guoping Qiu et.al proposed the Block Truncation coding (BTC) Technique. Image indexing and retrieval is a very important research topic which has attracted great interests in recent years. Image coding is a successful field which has been studied for more than three decades. All image coding techniques attempt to extract visually most important features and represent them as compactly as possible. Content-based image indexing involves the extraction of compact and discriminative features and uses them as indexing keys. To believe the two fields, image coding and image indexing, are closely related, especially, image indexing can exploit fruitfully the results and experiences of over 30 years research in the field of image coding. This paper presents a new application of a well-studied image coding technique, namely BTC. It is shown that BTC can not only be used for compressing color images, it can also be conveniently used for content-based image retrieval from image databases. Study a number of single bitmap BTC techniques for color image coding in different color spaces. From the BTC

compressed stream (without performing decoding), derive two image content description features, one termed the Block Color Co-occurrence Matrix (BCCM) and the other Block Pattern Histogram (BPH). Use BCCM and BPH to compute the similarity measures of images in the retrieval of images in database applications. Experimental results demonstrate that BCCM and BPH are comparable to similar state of the art techniques in image retrieval.

F. Content Based Image Retrieval: Survey and comparison of CBIR System based on combined features

In 2015 Savita Gandhani and Nandini Singhal et.al proposed the K-means clustering.. In image processing, computer vision and pattern recognition, the Image retrieval is a most popular research area. In this paper, performance of various CBIR systems, based on combined feature i.e., color texture and shape, are compared.

G. Discriminant Eigen features for Image Retrieval

In 1996 Daniel L. Swets and John (Juyang) Weng et.al proposed the Discriminant analysis algorithm. This paper describes the automatic selection of features from an image training set using the theories of multidimensional discriminant analysis and the associated optimal linear projection. To demonstrate the effectiveness of these Most Discriminating Features for view-based class retrieval from a large database of widely varying real-world objects presented as "well-framed" views, and compare it with that of the principal component analysis

H. Image retrieval with interactive query description and database revision

This paper has a further exploration and study of visual feature extraction. According to the HSV (Hue, Saturation, Value) color space, the work of color feature extraction is finished, the process is as follows: quantifying the color space in non-equal intervals, constructing one dimension feature vector and representing the color feature by cumulative histogram. Similarly, the work of texture feature extraction is obtained by using gray-level cooccurrence matrix (GLCM) or color co-occurrence matrix (CCM). Through the quantification of HSV color space, we combine color features and GLCM as well as CCM separately. Depending on the former, image retrieval based on multi-feature fusion is achieved by using normalized Euclidean distance classifier. Through the image retrieval experiment, indicate that the use of color features and texture based on CCM has obvious advantage.

I. Image Retrieval by Examples

In 2000 Roberto Brunelli and Ornella Mich et.al proposed the Multi Dimensional Scaling (MDS)TechniqueOrdered Dither Block Truncation Coding (ODBTC). A currently relevant research field in information sciences is the management of nontraditional distributed multimedia databases. Two related key issues are to achieve an efficient con-tent-based query by

example retrieval and a fast response time. This paper presents the architecture of a distributed image retrieval system which provides novel solutions to these key issues. In particular, a way to quantify the effectiveness of low level visual descriptors in database query tasks is presented. The results are also used to improve the system response time, which is an important issue when querying very large databases. A new mechanism to adapt system query strategies to user behavior is also introduced in order to improve the effectiveness of relevance feedback and overall system response time. Finally, the issue of browsing multiple distributed databases is considered and a solution is proposed using multidimensional scaling techniques.

J. Sketch based and content based Image retrieval

In 2016 Ankita Gaikwad and Dipali Gite et.al proposed the Object Boundary Selection (OBS) algorithm. Sketch based and content based Image retrieval. In Sketch-Based Image Retrieval the appearance gap between sketches and photo-realistic images is a fundamental challenge in sketch based image retrieval (SBIR) systems. The existence of noisy edges on photo-realistic images is a key factor in the enlargement of the appearance gap and significantly degrades retrieval performance. To bridge the gap, propose a framework consisting of a new line segment-based descriptor named histogram of line relationship (HLR) and a new noise impact reduction algorithm known as object boundary selection. HLR treats sketches and extracted edges of photo-realistic images as a series of piece-wise line segments and captures the relationship between them. Based on the HLR, the object boundary selection algorithm aims to reduce the impact of noisy edges by selecting the shaping edges that best correspond to the object boundaries. Multiple hypotheses are generated for descriptors by hypothetical edge selection. The selection algorithm is formulated to find the best combination of hypotheses to maximize the retrieval score; a fast method is also proposed. To reduce the distraction of false matches in the scoring process, two constraints on spatial and coherent aspects are introduced. Tested the HLR descriptor and the proposed framework on public datasets and a new image dataset of three million images, which we recently collected for SBIR evaluation purposes. Compared the proposed HLR with state-of-the-art descriptors (SHoG, GF-HOG). The experimental results show that HLR descriptor outperforms them. Combined with the object boundary selection algorithm, our framework significantly improves SBIR performance.

In Content-Based Image Retrieval (CBIR) due to increase in large image database, the storage of such data is expensive, so that the image compression techniques come into picture. Content-based image indexing and retrieval has been an important research area, in which indexing and retrieval is performed on the basis of the contents of the images. The contents are like color, shape or texture of that image. Trying to retrieve similar images from the compressed image database is tedious job. So introduce a technique which index and retrieve the images from such database. This paper implement a halftone based Ordered-Dither Block truncation Coding (ODBTC) technique to compress an image. The benefit of low complexity, ODBTC generate an image content descriptor for content based image retrieval (CBIR). In the encoding

step, compress an image block into corresponding quantizers and bitmap image. Two image features namely color co-occurrence feature (CCF) and bit pattern features (BPF) are used to index an encoded image by involving the visual codebook, and this features are generated directly from the encoded data streams without performing the decoding. An efficient approach to retrieve similar images from compressed database using hierarchical clustering algorithm is proposed. Hierarchical clustering algorithm is bottom-up approach to compute similar images with improved efficiency. So this scheme is not only provide image compression, because of its simplicity, but also simple and effective descriptor to index images in CBIR system.

III. PROPOSED APPROACH

In this paper, in order to have a fast image retrieval, feature extraction is classified into groups. Three levels of feature extraction are employed here for the three categories of features and database of any size is reduced to 50 images with high feature similarity in all the three levels of search. In the proposed system, the feature extraction is categorized into different levels. After each level of feature extraction and similarity matching, the database is modified with an algorithm called Database Revision (DR). The existing database is replaced after each feature search on the basis of the previous feature search results. The following topics explain the proposed system. The System also presents the different levels of feature extraction and the DR method.

A. Querying

The user provides a sample image as the query for the system.

B. Similarity computation

The system computes the similarity between the query image and all the database images according to the aforementioned visual features.

C. Database rewriting

The images in the database are rearranged in ascending of their similarity to the query and only the top n-images among this list are selected to rewrite the database. The database size is now reduced to n-images. 30

D. Retrieval

The top k-images from the rewritten database are displayed to the user. Here, 20 images are presented.

E. Relevance Feedback

The user can interact with the system and add his/her satisfaction for each of these displayed images by giving a user satisfactory value between 1.0 and 0.0 for each of these images.

F. Database Revision

The images in the database are re-ranked based on the user-satisfaction value entered for each image. Relevant images are those images which are given a satisfactory value 1.0 by

the user while images with zero user-given value are irrelevant images. Relevant images will remain in the database and irrelevant images will be discarded from the database. Images with satisfactory values between 1.0 and 0.0 will be placed in the database depending on the similarity value and user satisfactory value.

G. Retrieval

Top 20 images of the revised database are then displayed.

H. If the user is not satisfied with the result, the process continues. Then the user has to again enter satisfactory level for the displayed images and DR occurs accordingly. This continuous until the satisfied with the displayed images or the search exceeds a predefined limit.

IV. Distance Matching Algorithm

The effective content-based image retrieval (CBIR) needs efficient extraction of low level features like color, texture and shapes for indexing and fast query image matching with indexed images for the retrieval of similar images. Features are extracted from images in pixel and compressed domains. However, now most of the existing images are in compressed formats like JPEG using DCT (discrete cosine transformation). In this paper we study the issues of efficient extraction of features and the effective matching of images in the compressed domain. In our method the quantized histogram statistical texture features are extracted from the DCT blocks of the image using the significant energy of the DC and the first three AC coefficients of the blocks. For the effective matching of the image with images, various distance metrics are used to measure similarities using texture features. The analysis of the effective CBIR is performed on the basis of various distance metrics in different number of quantization bins. The proposed method is tested by using Corel image database and the experimental results show that our method has robust image retrieval for various distance metrics with different histogram quantization in a compressed domain.

V. Implementation

This system consists of three main modules.

A. Color Search
B. Texture Search
C. Shape Search

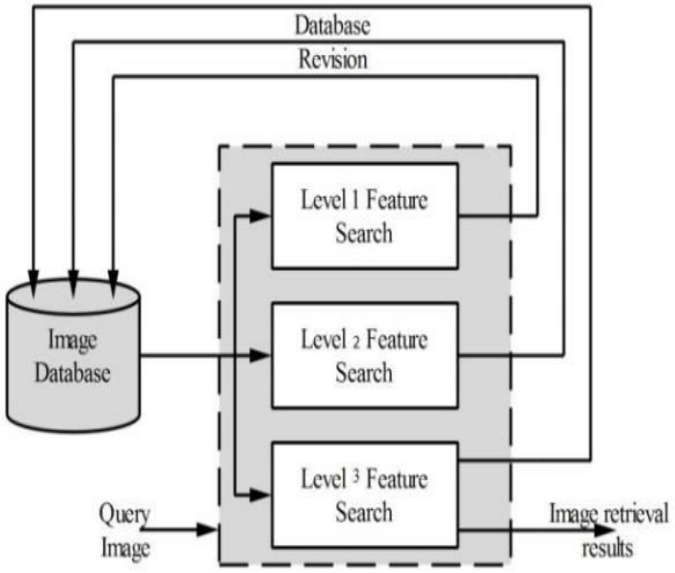


Fig 1. Feature Extraction at Different Levels

A. Color Search

Color is a very important aspect of any image. It is the most easily noticeable characteristic in images. This is because color feature possess higher human attention than any other image feature. Color feature is commonly used in all the CBIR systems.

Color features are generally divided into two groups:

1. Global color descriptors.
2. Local color descriptors.

The local color descriptors are advantageous compared to the global color descriptors as the pixel-level color information is not represented by them. The local color descriptors represent the color of the image with respect to its spatial location. In this paper, the local color features as well as some of the global color descriptors are utilized. Mean and standard deviation are the global color features. Binary bitmap using block truncation coding and color histogram are the local color descriptors. The HSV color space is used for the binary bitmap and global color descriptors and RGB color space for the color histogram. The color histogram feature search works RGB images as follows. First, the image is cropped to find the histogram of only the central portion of the image to concentrate on the localized color feature of the object depicted in the image, discarding the surroundings. Then the color histogram of the cropped portion is extracted. Color histogram represents the distribution of intensity of the color in the image.

B. Texture Search

Texture feature of an image is derived from a combination of pixels that reoccur several times in the image. The significance of extracting the texture is that it differentiates between objects with same backgrounds. Gray level co-occurrence matrix (GLCM) is used in this system to represent the texture feature. GLCM is a matrix. Each value of GLCM shows the number of reoccurrences of two pixels and separated by a distance and at an angle in the image.

C. Shape Search

Another distinguishing feature of images is their shapes. Shape is an important descriptor. An important shape feature is edge histogram descriptor (EHD). It represents the relative frequency of occurrence of the four types of edges, vertical horizontal, diagonal and anti-diagonal, in the corresponding 4x4 sub-image blocks of the image. The normalized representation of edges produces EHD. The shape search is carried out only for the 100 images that qualified the level 1 and level 2 searches. This considerably increased the speed of retrieval without any compromise on the results. The database is again revised to 50 top images based on similarity qualifying the shape search. Theses 50 images form the final database after the feature extraction.

VI. SYSTEM ARCHITECTURE

In this architecture, the user gives a query as an image and searching the relevant image using three levels like shape, color and texture. By using these three levels we can retrieve the most relevant images from the database.

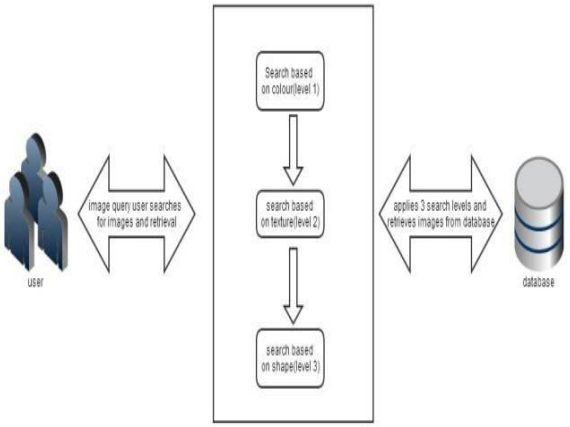


Fig 2. System Architecture

VII. RESULTS AND DISCUSSIONS

In this paper, experimental data set contains 1000 images from the database of images, divided into 10 categories, each category has 100 images. Experimental images including landscapes, animals, plants, fruits, transport (cars, planes) and so on. Selection of each type in the 80 images as training samples, 20 samples for testing. Feature extraction techniques based on the color, shape and texture feature extraction technology. Texture feature extraction techniques including two different extraction methods such as Gray Level co-occurrence Matrix(GLCM) and Color Co-occurrence Matrix(CCM). Color and texture and shape are describing the characteristics of images. Image database varies some images dramatic ups and downs in gray-level, showing a very strong texture characteristic, and some images from a number of smooth but the colors are different regional composition.

We consider the color and texture features combining not only be able to express more image information, but also to describe image from the different aspects for more detailed information in order to obtain better search results. The similarity measure from three types of characteristic features, including color features, texture features and shape feature. Three types of characteristics of images represent different aspects of property. In three feature extraction mode, color, shape and texture has effective performance.

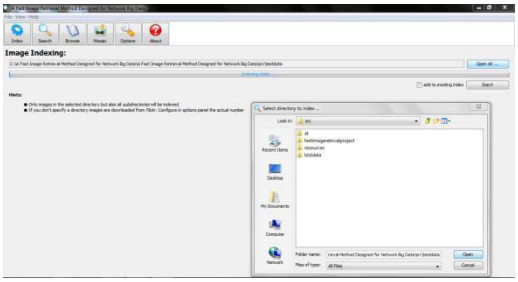


Fig 3. Screenshot for setting path

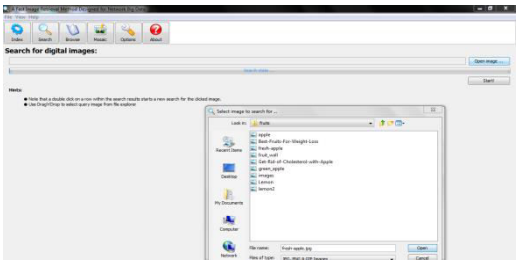


Fig 4. Screenshot for Selecting the Input

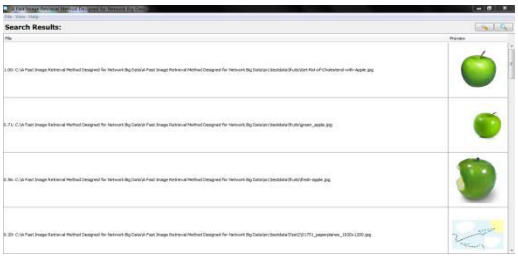
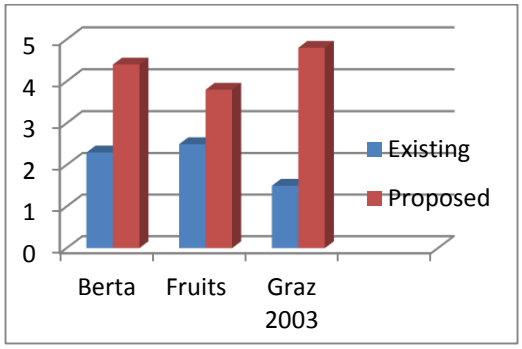


Fig 5. Screenshot for Output



Graph 1. Accuracy comparison for Existing and proposed Systems

VIII. Conclusion

The project presents an improved system by introducing a new algorithm based on feature categorizing into levels. Here, each image from all the image classes is compared. Both conventional and proposed methods are executed for retrieval. The significance of feature levels based search is verified. It is much faster than conventional method and as precise as the existing methods. The system is efficient and user-friendly because it gives good results without time wastage.

IX. Future Enhancements

In future, additional descriptors like blobs and patch will also be added which are very useful to extract the most relevant images based on the similarity measures. This will also reduces the size of the database as well as time to retrieve images. Future scope also includes implementing the CBIR system considering more low-level image descriptors and highly efficient deep learning neural network, which might prove to be very fast and précised one

References

[1] A Fast Image Retrieval Method Designed for Network Big Data Jiachen Yang, Member, IEEE, Bin Jiang, Baihua Li, Kun Tian, and Zhihan Lv, Member, IEEE

[2] A. Hoyos-Idrobo, P. Weiss, P. Massire, A. Amadon, and N. Boulant, "On variant strategies to solve the magnitude least squares optimization problem in parallel transmission pulse design and under strict sar and power constraints," IEEE Trans. Medical Imaging, vol. 33, no. 3, pp. 739– 748, 2014.

[3] A. Mehmood and H. Song, "Smart energy efficient hierarchical data gathering protocols for wireless sensor networks," SmartCR, vol. 5, pp. 498–509, 2015.

[4] A. Sarmah, R. Mandal, and M. Bhattacharya, "Content based image retrieval using color, shape and texture," Int. J. Adv. Res. Comput. Sci., vol. 3, no. 10, pp. 1–7, Oct. 2014.

[5] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, no. 12, pp. 1349–1380, Dec. 2000.

[6] B. Wu, Z. G. Xiong, Y. Z. Chen, and Y. D. Zhao, "Classification of quickbird image with maximal mutual information feature selection and support vector machine," Procedia Earth Planetary Sci., vol. 1, no. 1, pp. 1165–1172, 2009.

[7] B. Zhang, Y. Gao, S. Zhao, and J. Liu, "Local derivative pattern versus local binary pattern: Face recognition with high-order local pattern descriptor," IEEE Trans. Image Process. Publication IEEE Signal Process. Soc., vol. 19, no. 2, pp. 533–544, Feb. 2010.

[8] C. Lin, Z. Song, H. Song, Y. Zhou, Y. Wang, and G. Wu, "Differential privacy preserving in big data analytics for connected health," J. Med. Syst., vol. 40, no. 4, pp. 1–9, 2016.

[9] C. H. Lin, D. C. Huang, and Y. K. Chan, "Fast color-spatial feature based image retrieval methods," Expert Syst. Appl., vol. 38, no. 9, pp. 11412–11420, 2011.

[10] C. H. Lin, R. T. Chen, and Y. K. Chan, "A smart content-based image retrieval system based on color and texture feature," Image Vis. Comput., vol. 27, no. 11, pp. 658–665, 2009.

[11] C. Zhang, J. Y. Chai, and R. Jin, "User term feedback in interactive textbased image retrieval," in Proc. Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, Aug. 2005, pp. 51–58.

[12] D. L. Swets and J. Weng, "Using discriminant eigenfeatures for image retrieval," IEEE Trans. Pattern Anal. Mach. Intell., vol. 18, no. 8, pp. 831–836, Aug. 1996.

[13] E. Baccarelli, N. Cordeschi, M. Panella, S. Mohammad, and J. Stefa, "Energy-efficient dynamic traffic offloading and reconfiguration of networked data centers for big data stream mobile computing: Review, challenges, and a case study," IEEE Netw., vol. 30, no. 2, pp. 54–61, Mar./Apr. 2016.

[14] F. Li, Q. Dai, and W. Xu, "Improved similarity-based online feature selection in region-based image retrieval," in Proc. IEEE Int. Conf. Multimedia Expo, 2006, pp. 349–352.

[15] G. Cena, I. C. Bertolotti, A. Valenzano, and C. Zunino, "Evaluation of response times in industrial wlans," IEEE Trans. Ind. Inf., vol. 3, no. 3, pp. 191–201, Aug. 2007.