Privacy Issues in Content based Image Retrieval (CBIR) for Mobile Users, and a Proposed Solution

Kunal Agarwal[1]

[1]*Department of CSE, Guru Nanak Institutions, Hyderabad, India.*

---***---

Abstract - *Today, in the era of enhanced cameras and social media and cheap cloud storage, every smartphone user is bound to capture and store thousands of pictures and screenshots. However, when so much data gets collected, a new challenge comes up, i.e., the ability to efficiently surf through this data. It is turning into an annoying task to search through our mobile gallery and find a year-old picture owing to the number of photos we now have. CBIR is the application of computer vision technique to such image retrieval problem, that searches for digital images in large databases by analyzing the contents of the image rather than the metadata or descriptions associated with the image. But this convenience seems to be coming at the cost of User Privacy. For an artificially intelligent system to run a CBIR on a large number of photos, the data is uploaded on the clouds where the deep-learning neural network module processes the pictures on the server side, defiling the user's privacy. This is because the learned models in deep learning require significantly more memory, disk storage, and computational resources which today's mobile phones still lack. In this paper, we identify the challenges faced by the top companies providing photo storage applications in implementing their smart CBIR system, and learn about their legal crisis against the governments trying to safeguard citizens against unlawful collection and storing of user biometric information. Finally, we recommend a new approach to this problem by implementing 'Indistinguishability under Chosen-Plaintext Attack (IND-CPA)' secure CBIR framework and a pre-trained deep Convolutional Neural Network model, i.e., VGG-16.*

***Key Words*: metadata, cloud, neural network, retrieval**

1. INTRODUCTION

Owing to the rapid development of deep learning, a number of research areas have recently excelled, and along with the continuous improvement of convolution neural networks, computer vision has arrived at a new peak.[1] CNN also makes the application of computer vision greatly improve, such as face recognition, object detection, object tracking, semantic segmentation, and so on. Object detection is the technology related to image processing that deals with detecting instances of semantic objects of a certain class in digital images and videos, and has applications in many areas including image retrieval and video surveillance.

Image retrieval system is used for efficiently browsing, searching and retrieving images from a large database of digital images. Most conventional and common methods of image retrieval utilize some method of adding meta data such as keywords, captioning, title or descriptions to the images so that retrieval can be performed over the annotation words. However, manual image annotation is time-consuming, laborious and too expensive. To address this, there has been a large amount of research done on Content-based image retrieval (CBIR). Content-based signifies that the search analyzes the contents within the image instead of the metadata attached to it. The term "content" in this context might refer to colors, shapes, textures, or any other information that can be derived from the image itself.[4] CBIR is desirable because searches that rely completely on metadata are dependent on completeness and annotation quality.

Most of us love to take a million pictures and screenshots of every important (or not so important) part of our lives, but then scroll for ages through a bottomless self-generated photo feed and can never seem to find that one picture we really wanted.[5] Google has made it much easier to look for specific images with objects/texts through its Photos app. The 'Things' category scans photos for their subject matter: birthdays, buildings, cats, concerts, food, graduations, posters, screenshots, etc. In May 2017, Google announced that its free app Google Photos has over 500 million users, who upload over 1.2 billion photos every day.

There is a popular quote that says, "If You're Not Paying For It, You Become The Product". Notwithstanding the promise of rising AI technology to enhance many aspects of life, serious concerns about privacy, security, government monitoring, and the sale of people's personal, biological data have risen up. The Biometric Information Privacy Act (BIPA) was passed by the Illinois General Assembly on October 3, 2008, that safeguards citizens against the unlawful collection and storing of biometric information. The Act prescribes $1,000 per violation, and $5,000 per violation if the violation is intentional or reckless. Because of this damages provision, the BIPA has spawned many class action lawsuits involving Google's user privacy exploitation, Facebook's face-tagging feature, Snapchat Lens, Shutterfly, etc. Other nations too including England, China, Canada have since passed similar laws.[9]

We realized that the major tech companies' products implement Neural Networks system for developing and organizing a photo storage album, but parallelly exploit most of the users' privacy as they apply the deep-learning vision model using cloud-based API. We study and recommend a new approach to this problem proposed by a team of researchers from Communication University of China (CUC). Here, we implement an IND-CPA secure CBIR framework that performs image retrieval on the cloud without the user's constant interaction. A pre-trained deep Convolutional Neural Network model, i.e., VGG-16, is used to understand the deep features of an image. We apply our framework into three public image datasets. The experimental results show that our framework is efficient and accurate.

2. OBJECT DETECTION USING ARTIFICIAL INTELLIGENCE

Object detection as one of the important applications in the field of computer vision has been the focus of research, and convolution neural network has made great progress in object detection. Object detection is developing from the single object recognition to multi-object recognition. The meaning of the former is just to identify a single object from an image, it can be said that it is a problem of classification, and the meaning of the latter is not only to identify all the objects in an image, but also map the exact location of the objects. [1]

Getting to use modern object detection methods in applications and systems, as well as building new applications based on these methods is not a straightforward task. Previous implementations of object detection used classical algorithms, like the ones supported in OpenCV, the popular Computer Vision library. However, these algorithms did not achieve enough performance to work under distinct conditions.

The outrage and enormous adoption of deep learning in 2012 brought into existence modern and highly accurate object detection algorithms such as R-CNN, Fast-RCNN, Faster-RCNN, RetinaNet, yet highly accurate ones like SSD and YOLO. Using these methods and algorithms based on deep learning however requires a lot of mathematical and deep learning frameworks understanding.[2]

For a demonstration, a new training dataset is required, so we generated a random street view image dataset, which has multiple categories of objects like human, bus, truck, car, bicycle. Fig 1 is the image from which we will detect as many objects as possible, and Fig 2 is the resultant image with the detected objects marked using bounding boxes in distinct colors.

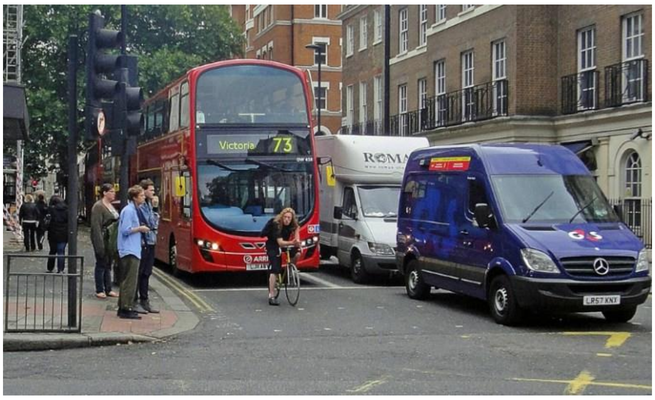

Fig 1: Before Detection [2]

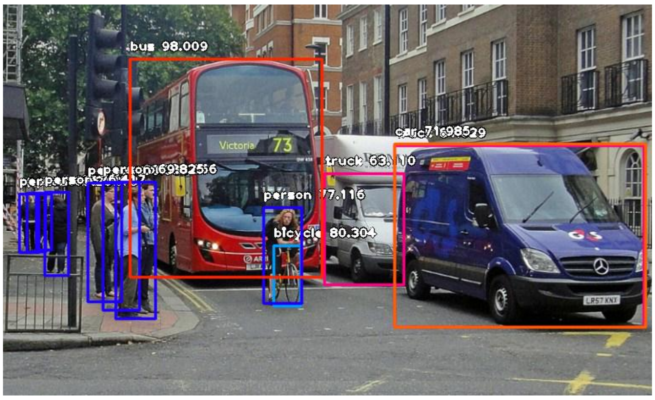

Fig 1: After Detection [2]

Red marked in the figure is the bus, blue marked are humans, pink marked is truck, cyan marked is bicycle, and orange marked is car. Many more features useful for customization and production capable deployments for object detection tasks have been developed over time.

- **Adjusting Minimum Probability**: In the above demonstration, objects detected with a probability percentage of less than 50 are not be shown or reported. We can increase this value for high certainty cases or reduce the value for cases where all possible objects are needed to be detected.

- **Custom Objects Detection**: Using a CustomObject class, you can tell the detection class to report detections on one or many numbers of unique objects.

- **Detection Speeds:** You can reduce the time taken to detect an object by setting the speed of detection speed to 'fast', 'faster' and 'fastest'. [2]

3. CONTENT-BASED IMAGE RETRIEVAL

Content-based image retrieval, also known as query by image content (QBIC) and content-based visual information retrieval (CBVIR), is the application of computer vision techniques to the image retrieval problem, that is searching for digital images in large databases. Content-based image retrieval is opposed to traditional concept-based approaches where the image annotations are searched for image retrieval. CBIR is desirable because searches that rely completely on metadata are dependent on completeness and annotation quality. Fig 3 shows a glimpse of well-trained CBIR system where an image query is described by using one or more example images, and low-level visual features (e.g., color, texture, shape, etc.) are automatically extracted to represent the images in the database.[3]

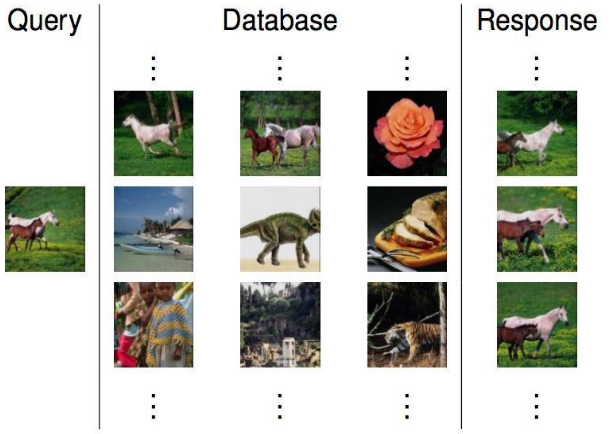

Fig 3: Content Based Image Retrieval

3.1 Google Photos

Every smartphone user today clicks hundreds of pictures of things and people around him, and dumps them into his phone memory. But when he later wants to roll back to a specific picture, he only ends up scrolling for ages through a bottomless self-generated photo feed and fails to find it. Google Photos has made it much easier to look for specific images with objects/texts. Over the years, the tech giant released various object recognition features powered by Google Lens for its backup application, so it doesn't end up as any photo dump where you can't find anything anymore.[6] The Photos service analyzes and organizes images into groups and can identify features such as beaches, parks, skylines, or "Tsunami in Mumbai". From this app's search window, users are shown recommended searches for groups of photos in 3 major categories: Places, People, and Things. It analyzes photos for similar faces and groups them together under the People category. Similarly, the Places category uses geotagging data and can also determine locations in older pictures by analyzing for major landmarks (e.g., photos containing the Taj Mahal). The Things category processes photos for their subject matter like birthdays, buildings, cats, concerts, food, graduations, posters, screenshots, etc. Users can manually remove categorization errors.

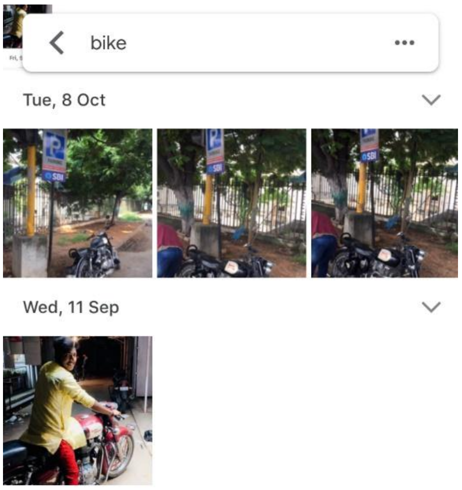

Fig 4: 'Bike' in Google Photos

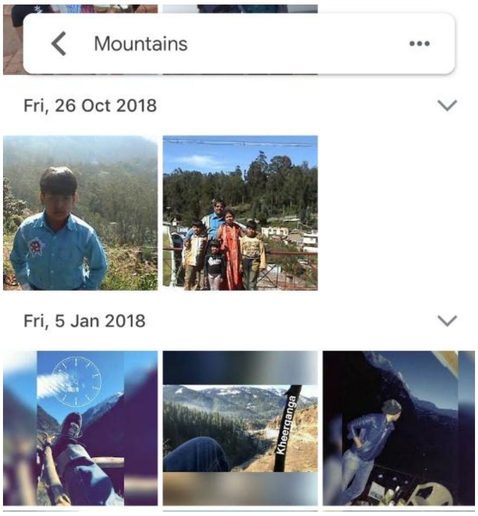

Fig 5: 'Mountains' in Google Photos

Google Photos recently rolled out Optical Character Recognition (OCR) on all photos within the Google Photos that now reads and searches texts in any image. Further, we can simply click the Lens button when we find the target image to be able to copy-paste text from it.

3.2 Apple Photos On-Device Face Recognition with Secured User Privacy

Apple's CEO Tim Cook, against the general cloud-based photo albums, noted that the customers should be in control of their own information. We might like these so-called free services, but we don't think if they are worth having our email, our search history and now even our family photos data mined and sold off for undisclosed advertising purpose. Someday, the customers will realize this but it would be too late then to do anything but regret.

Deep learning for face detection was introduced by Apple in iPhones in 2012 using the iOS 10. It even released the Vision framework, that could preserve user privacy and run efficiently on-device. Compared to traditional computer vision, the learned models in deep learning required more memory, disk storage, and computational resources. As advanced as today's mobile phones are, the typical high-end mobile phone was not an ideal platform for deep-learning vision models. Majority of the industry dealt with this problem by providing deep-learning solutions through a cloud-based API. In this solution, images are uploaded to a server for analysis using deep learning inference to detect similar faces. Cloud-based services typically use powerful desktop-class GPUs with large amounts of memory available. Very large network models, and potentially ensembles of large models, can run on the server side, allowing clients (mobile phones) to take advantage of large deep learning architectures that would be impractical to run locally.[8]

Apple iCloud Photo Album is a cloud-based solution for photo and video storage. However, due to the company's strong commitment to user privacy, they couldn't use iCloud servers for computer vision computations. All the data sent to iCloud Photo Library is encrypted on the device itself before it is uploaded, and can only be decrypted by devices that are registered with the iCloud account. Therefore, to bring deep learning-based computer vision solutions to their customers, they had to address directly the challenges of getting deep learning algorithms running on iPhone. [8]

They faced several challenges and came out with various upgrades. The deep-learning models were shipped as part of the operating system, taking up valuable NAND storage space. These models were also loaded into RAM, and required significant computational time on the GPU and/or CPU. Unlike cloud-based services, whose resources can be dedicated solely to a vision problem, on-device computation takes place while sharing these system resources with other running applications. Finally, the computation must be efficient enough to process a large Photos library in a reasonably short amount of time, but without significant power usage or thermal increase. [8]

4. BIOMETRIC PRIVACY INFORMATION ACT (BIPA)

To make the Photos app's sharing and tagging features work, Google had to analyze a photo subject's facial structure and create a unique "faceprint" for them. The company is currently fighting a lawsuit in Illinois being alleged that their facial-recognition technology violates a state law guarding citizens' biometric data, and the tech hasn't been rolled out in many parts of Europe believing that it might run afoul of their privacy laws. On 27 February 2017, a federal district court in Chicago rejected Google Inc.'s motion to dismiss a putative class-action lawsuit by two Illinois residents over Google's creation of face templates from photographs of the plaintiffs. The court held that information about people's physical traits derived from their photographs is covered under the BIPA, just as information derived from an in-person facial scan would be. This could make it likelier that more people will file BIPA lawsuits based on photograph-derived information, and possibly lead to the suppression of helpful technology-based advances and job creation.[9]

Over the last decade, biometric technology has increasingly found its applications in various fields. Facebook, Apple and Snapchat, for instance, use technology based on people's facial geometry and allow users to employ fingerprint technology to unlock phones. Some financial institutions even enable customers to conduct transactions online or at ATMs through such technology.

The Illinois General Assembly stated, although a person can change her Social Security number if it is stolen, its impossible to alter her distinct facial geometry once that identifier falls into the wrong hands. The BIPA includes "biometric information" as data derived from a "biometric identifier" such as "a retina or iris scan, fingerprint, voiceprint, or scan of hand or face geometry." Illinois' BIPA restricts any private entity to obtain a person's biometric identifier or information, unless the entity informs the person in writing and discloses the specific purpose and length of time for which the data are being collected, stored or used. The entity must even obtain written consent from the person to use or store his/her biometric information. The act further requires the party in possession of the data to protect the security of the information. [9]

The tech companies tried arguing that the face templates did not count as "biometric information" under the BIPA because they had created them from photographs, not in-person facial scans. The law-keepers said, "For every face template, the tech companies are creating a set of biology-based measurements that is used to identify a person, which is equivalent to biometric information." The court highlighted that the BIPA does not specify "how the biometric information can be obtained" and hence there is no statutory basis for excluding photograph-sourced measurements from the requirements of the BIPA. [9]

Several cases concerning the digital privacy of mobile users alerted Illinois lawmakers that the state's privacy law might be having a damaging effect on the advancing and helpful AI technology. On May 26, 2016, state Sen. Terry Link, D-Vernon Hills, author of the BIPA in 2008, proposed an amendment to specifically exempt physical and digital photographs and biometric information derived through them from the privacy protections of the act. Link's proposal could have affected the status of technology, but on May 31, 2016, after privacy advocates and the Illinois attorney general raised concerns about the proposed amendment, the measure was put on hold. Finally, the use of biometric information is an increasingly common facet of everyday life. Lawmakers, government bodies, businesses and consumers need to arrive at ways to protect this data.

5. SECURE CBIR SYSTEM

Secure content based image retrieval has attracted considerable interests recently due to users' security concerns. However, it still suffers from the challenges of relieving mobile devices of excessive computation burdens, such as data encryption, feature extraction, and image similarity scoring.[17]

In a paper by Fei Liu, Yong Wang, Fan-Chuan Wang, Yong-Zheng Zhang, and Jie Lin, they propose and implement an IND-CPA secure CBIR framework that performs image retrieval on the cloud without the user's constant interaction. A pre-trained deep CNN model, i.e., VGG-16, is used to extract the deep features of an image on device. The information about the neural network is strictly secured by using the lattice-based homomorphic scheme. They implement a real number computation mechanism and a divide-and-conquer CNN evaluation protocol to enable the framework to securely and efficiently evaluate the deep CNN with a large number of inputs. They further proposed a secure image similarity scoring protocol, which enables the cloud servers to compare two images without knowing any information about their deep features. [17]

Although users prefer systems to offer both functionalities, i.e. intelligent and secure image retrieval, the challenging task is to outsource the image retrieval onto a cloud without letting the cloud know anything about the image contents during the processing phase.[17] The detailed procedure to implement a secure content-based image retrieval system is summarized as follows:

- Implement a CBIR framework that shifts excessive computations onto the cloud servers, such as IND-CPA to deal with secure image re-encryption, deep feature extraction, and image similarity scoring. In this way, a mobile user only needs to encrypt his/her image with a lightweight encryption algorithm and upload the encryption onto the cloud. The latter performs Approximate Nearest Neighbor (ANN) image retrieval without further user's interaction.

- Apply a pre-trained deep CNN model, i.e., VGG-16, to extract the deep features of an image. The information about the neural network is strictly concealed by utilizing lattice-based homomorphic scheme. Implement a real number computation mechanism to achieve better accuracy without loss of its efficiency.

- A divide-and-conquer CNN evaluation protocol is used to deal with the problem of noise growth in the homomorphic scheme. This protocol makes it possible to homomorphically evaluate very deep CNN with a large number of inputs.

- Finally, follow a secure image similarity scoring protocol, which enables the cloud servers to compare two images without knowing any information about their deep features. [17]

We apply our framework into three public image datasets. The experimental results in Fig 6 show that the framework is efficient and accurate.

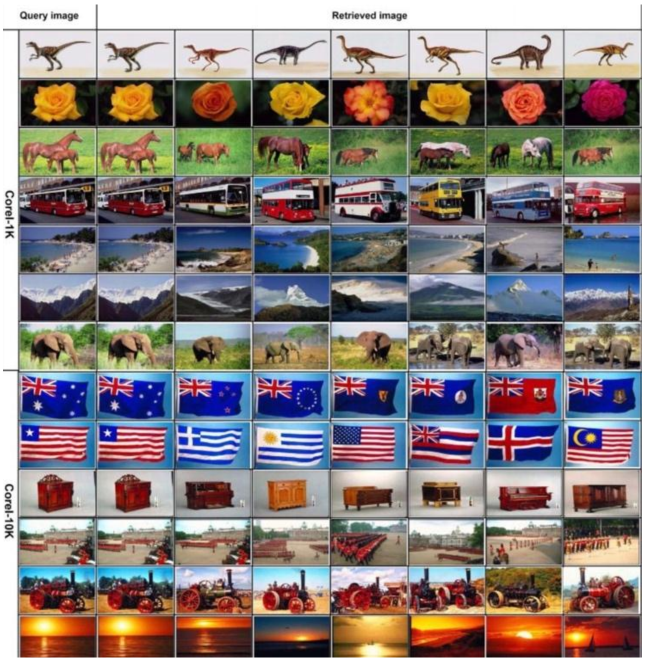

Fig 6: Test Results [17]

6. CONCLUSIONS

In the era of global digitalization and social media, every smartphone user is bound to capture and store thousands of pictures. And with the tremendous growth of smart mobile devices, the Content-Based Image Retrieval (CBIR) becomes more and more popular daily and has great market potentials.

In this paper, we highlight the major shortcoming of the state-of-the-art CBIR system run by Google and Apple, and also discuss the privacy laws impacting the growth of computer vision and deep learning networks. What we need today is a reliable photo storage which is efficiently organized and safe at the same time. The smart model understands our query and swiftly provides us with the exact images which we are looking for. Regarding the privacy issue, user data cannot be risked and needs to be safeguarded from organization with atrocious intentions.

Google aims to become the third half of our brain. But now think about it: Do we really want the third half of our brain to make a living by using our personal data to show us ads? I don't.

REFERENCES

[1] Xinyi Zhou, Wei Gong, WenLong Fu, Fengtong Du, "Application of Deep Learning in Object Detection", 2017 IEEE/ACIS 16th International Conference on Computer

and Information Science (ICIS), DOI: 10.1109/ICIS.2017.7960069.

[2] Moses Olafenwa, "Object Detection with 10 lines of code", https://towardsdatascience.com/object-detection-with-10-lines-of-code-d6cb4d86f606.

[3] Swati Killikatt, Vidya Kulkarni, Madhuri Bijjal, "Content Based Image Retrieval by Online and Offline" International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering Vol. 2, Issue 7, July 2013, ISSN (Online): 2278 – 8875.

[4] "Content-based Image Retrieval", www.wikipedia.org/wiki/Content-based_image_retrieval

[5] Dana Miller, "Google Photos Just Added a Feature That Can Search by Word Recognition", 24 Aug 2019, www.interestingengineering.com

[6] Mariella Moon, "Google Photos can now search for text in images", 08.23.19, www.engadget.com

[7] Paul Nieuwenhuysen, "Information Discovery and Images: A Case Study of Google Photos", 2018 5th International Symposium on Emerging Trends and Technologies in Libraries and Information Services (ETTLIS), 11 October 2018, DOI: 10.1109/ETTLIS.2018.8485238

[8] Computer Vision Machine Learning Team, "An On-device Deep Neural Network for Face Detection" November 2017, Apple Machine Learning Journal

[9] Illinois Policy, "Federal Court in Illinois rule Biometric privacy Lawsuit against Google can proceed", https://www.illinoispolicy.org/federal-court-in-illinois-rules-biometric-privacy-lawsuit-against-google-can-proceed

[10] Dylan Curran, "Here is all the data Facebook and Google have on you", 30 Mar 2018, www.theguardian.com

[11] Maria Tezlepi, "Deep convolutional learning for Content Based Image Retrieval", Neurocomputing, Volume 275, 31 January 2018, Pages 2467-2478

[12] Rehan Ashraf, Mudassar Ahmed, Sohail Jabbar, Shehzad Khalid, Awais Ahmad, Sadia Din,Gwangil Jeon, "Content Based Image Retrieval by Using Color Descriptor and Discrete Wavelet Transform", Image & Signal Processing, First Online: 25 January 2018

[13] Sachendra Singh Chauhan, Shalani Batra, "Efficient layer-wise feature incremental approach for content-based image retrieval system", J. of Electronic Imaging, 28(2), 023038 (2019). https://doi.org/10.1117/1.JEI.28.2.023038

[14] "An On-device Deep Neural Network for Face Detection", Vol. 1, Issue 7, November 2017, by Computer Vision Machine Learning Team

[15] Iftikhar Ahmad, Shafaq Abdullah, Serkan Kiranyaz, Moncef Gabbouj, "Content-based image retrieval on mobile devices", in Proceedings of SPIE - The International Society for Optical Engineering 5684, March 2005, DOI: 10.1117/12.596772

[16] Arif Rahman ; Edi Winarko ; Moh. Edi Wibowo, "Mobile content based image retrieval architectures", 2017 4th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI), DOI: 10.1109/EECSI.2017.8239111

[17] Fei Liu, Yong Wang, Fan-Chuan Wang, Yong-Zheng Zhang, Jie Lin, "Intelligent and Secure Content-Based Image Retrieval for Mobile Users", IEEE DOI: 10.1109/ACCESS.2019.2935