International Journal of Research in Engineering, Science and Management
Volume-3, Issue-6, June-2020
www.ijresm.com | ISSN (Online): 2581-5792


# Spark Based Real Time Imagepro

Shivprasad S. More[1], Prajakta Madhukar Varute[2*], Pooja Mahesh Shirguppe[3], Sankalp Sanjay Mali[4], Neha Pradip Kambale[5], Sneha Mukund Kokate[6]

[1]*Assistant Professor, Department of Computer Science and Engineering, Sanjay Ghodawat University, Kolhapur, India*

[2,3,4,5,6]*Student, Department of Computer Science and Engineering, Sanjay Ghodawat University, Kolhapur, India*

*Corresponding author: prajuuvarute@gmail.com

***Abstract*: The time, only when images had been used illegally and serious adverse consequences had appeared did. Therefore, preserving image protection and the rights of the image owner is of great importance of today's regular life. Most image security is actually passive. Most of the time, the image owner may only know the illegal use of their images. And there was no any facility to actually protect images. Spark-based real-time proactive image protection model is proposed to monitor the status of images under protection in real time.**

***Keywords*: Fingerprint, Image security, Spark, Database system, Crawling, OTP.**

## 1. Introduction

No one would like to use to his/her belongings unauthorized, especially their photos. For example, now days everyone shares the its recent Images on social media. In these circumstances, the rights of Image owner are hurt, and if they will determine the unauthorized use of their photos as early as possible, their losses are often minimized.

Nowadays, information spread sooner and widely, which makes the safety and privacy protection of data really important [1]. Once images are illegally utilized in inappropriate situations, the image owners may suffer from great trouble or loss [2]. Therefore, it's of great necessity to possess research on real-time proactive image protection to defend the rights of image owners. There are a huge number of images existing in websites, and lots of new images appear every day.

During this big data environment, traditional standalone image processing method can hardly guarantee the image safety in real time [3, 4]. There are lots of research leads to image processing, while to the simplest of our knowledge, there's no research on real time and image tracking protection model is proposed to seek out the illegal use of images and protect the images owner's legitimate rights.

Spark is an open-source distributed general purpose cluster computing framework. It utilizes in-memory catching and optimized query execution for fast queries against data of any size.

This model is deployed within the parallel computing frame spark to enhance the system's real-time performance.

## 2. Literature Survey

At present criteria, there is no proper existing system for real-time image tracking system, so we propose a new system to tracking the image is most important factor in now days so providing protection to Images is necessary. Protection of Images and their Image owner's is very important so we design ImagePro in real-time based on spark.

In this project implementation we can develop ImagePro in real-time based on spark. Which provide protection to Images and notify unauthorized use of images.

### *A. Existing System*

The most important topics related to our research are image security, authentication, and retrieval. Photo privacy is intended to provide high-security photos to preserve their confidentiality and integrity Watermarking and encryption are commonly used methods of image protection. Lots of research has been done in this area [5-8]. Hu et al. [9] Use impulsive neural network synchronization technique to protect intelligent image against illegal swiping and abuse.

Bhargava et.al [10] proposes a method to add invisible watermark to an image, with the user information hidden.

A similar project related as we propose is also available for commercial use.

- TinEye

TinEye is a search engine developed and offered by Idée, Inc. for reverse image. It is the first web-based image search engine to use image recognition rather than keywords, metadata or watermarks.

TinEye allows users to search for images rather than using keywords. TinEye generates a "one-of - a-kind and lightweight digital signature or fingerprint" of the image when it is uploaded and matches it with other indexed images.

*Disadvantages:*

1. This type of system requires highly configured machines to work, huge amount of database to be supported.
2. Also, as this is a commercial service, they charge a large amount of money to provide their services like $300 per month.
3. There is an upload limit size is 20 MB.

## 3. Proposed System

### A. Problem Definition

The main objective of this project is to notify image owner if any kind of misuse is detected. This model protects the ownership and privacy of image owners.

### B. Objectives

1. To provide image with high security to keep its confidentiality and integrity.
2. Spark based real-time ImagePro model (SRI) is used for real time image protection.
3. Spark based system greatly reduces the extraction time as compared to stand alone environment and thus increases efficiency.
4. To derive a cost effective system which be used by any normal person.

### C. System Architecture

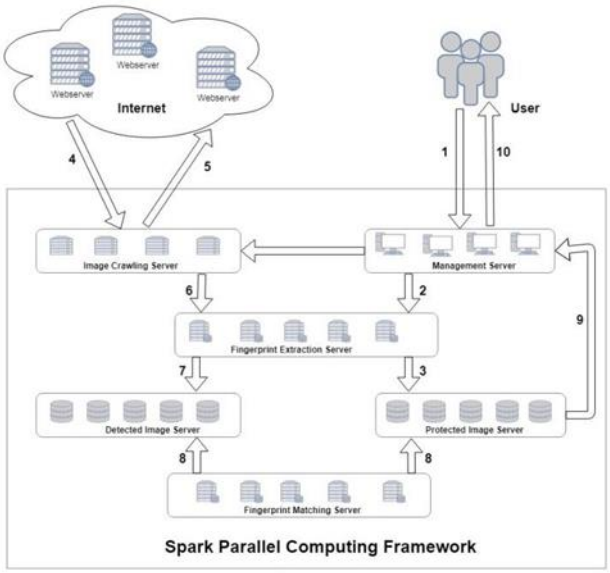



Fig. 1. System architecture of spark based real-time ImagePro

SRI's purpose is to protect image rights and privacy, and it will send warning to image owners as soon as their pictures are found to be used unlawfully. The detailed image protection process is as follows:

1) If a user feels it is necessary to protect his / her photos, he/she should apply the application to SRPITP for protection. The images submitted are checked to see if they were already in the database of the protected image. If not, the appeal will be accepted.
2) The photographs approved are fingerprinted and marked by the extraction servers for fingerprints.
3) In the secure image archive, the fingerprints of these photographs and related owner details are stored.
4) Photos are crawled in the pages that are being tracked. The administrator of the SRPITP system has the responsibility of deciding which websites should be tracked.
5) File crawling servers collect website files.
6) The fingerprint extraction servers must tag and fingerprint images collected from step (5).
7) In the detected image database, the fingerprints and related image information obtained from step (6) are saved. The related information contains up loader names and upload time.
8) The fingerprint matching servers use the image matching algorithm to decide if there is unauthorized use of the file
9) If inappropriate use of the picture is detected, a message will be sent to a maintenance system by the fingerprint matching application.
10) The management server immediately sends an alarm to the image owner, and detailed information about the illegal use is also sent to the owner. The picture owner can take appropriate steps to protect his/her interests.

Therefore, to maximize the performance of the device, Spark is added. One of the administration servers is the boss who controls the state of the entire system and is also responsible for the scheduling of jobs. All the other machines in our network function as worker nodes and operate in addition to maintaining ImagePro real-time capital.

Also the OTP system is added for the purpose to preserve owner rights.

### D. Modules

#### 1) Image Fingerprint extraction

One of the core algorithms in ImagePro is the fingerprint extraction algorithm. Compared with other image finger print extraction algorithms, the dHash algorithm has lower calculation complexity and higher accuracy, so it is adopted in our model for the reader's convenience.

#### 2) Crawler Image

Scrapy is selected in ImagePro to complete the collection of real time images from websites decided by the system administrator. Scrapy is a fast, high level screen capture and web crawler websites and extract structured data from pages.

Scrapy is deployed in the image crawling servers, and the servers execute the crawling at regular intervals [11]. When acquiring images from websites, the image crawling servers adopts the so called incremental crawling policy, i.e. The server's only crawl websites' new added images since last crawl.

#### 3) Database establishment and Tag classification

ImagePro has two types of database, i.e. the secure image database and the image database that has been identified. Protected image database contains image information submitted by the proprietors of the image. To increase the efficiency of image matching, this database has several data tables containing various types of images, such as photos of human and photographs of scenes or animals etc.

The structure of detected image database is quite similar to that of the protected image database, except that this database contains only one data table and it has a field to hold image tag values.

There is no need to Travers the entire protected image database during matching to get the result, and only the data table with same tag must be searched with the tag value of the detected image.

*4) Image matching*

SRI is based on efficient and accurate matching of the graphic. Because SRI has to manage large images in real time, Tag labeling approach is chosen to help ensure quicker image matching. During the image matching process, image files in the observed database are managed in order to maximize machine performance.

## Result and Analysis

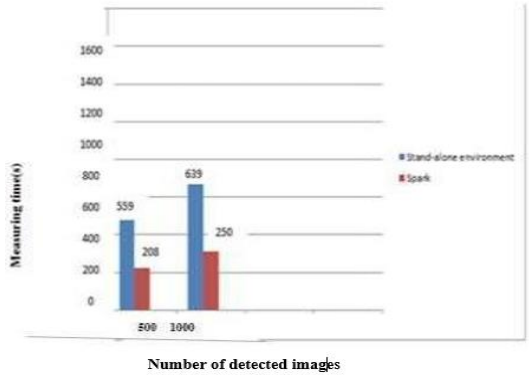
### *E. Matching accuracy*

Table 1
Image matching accuracy rate

| Number of protected Images | Number of detected images | Accuracy rate |
|---|---|---|
| 500 | 490 | 91.80% |
| 500 | 1000 | 94.22% |

In the protected image database, the total amount of data is 500, and the data volume in the detected image database is 500 and 1000, respectively. The results that match the image are compared with the results that are checked manually. Every experiment was performed five times, measuring the average precision rate, as shown in Table 1.

### *F. Comparison of Image matching time*




Since image matching efficiency is a key indicator of ImagePro, comparative experiments were performed with the application of Spark to show the system performance improvement. The total data volume of the protected image database was 500 in each experiment, and the maximum value was 1000 for the detected image database. Every experiment was performed five times, measuring the average precision rate, as shown in Table 1.

## 4. Future Scope

In future, we are planning to improve the recognition ability of fingerprint extraction algorithm to handle images with serious attack. Also, more efficient algorithms will be proposed and technologies will be adopted to further improve the performance of ImagePro.

## 5. Conclusion

In order to solve the problem of image privacy and security protection, a real-time proactive image tracking protection model ImagePro is proposed. Tag classification method and parallel computing framework Spark are adopted to enhance the efficiency of ImagePro.

## References

[1] D. G. Feng, Z. Min, L. Hao, Big data security and privacy protection. Chin. J. Comput. Phys. 37(1), 246–258 (2014).
[2] J. Liu, H. Wang, Image protection scheme based on privacy protection in content sharing environment. Comp. Appl. Software 32(7), 207–211 (2015).
[3] X. Lu, X. Han, Research on big data security and privacy protection technology architecture. Inform. Secur. Res. 2(3), 244–250 (2016).
[4] Y. Liu, T. Zhang, X. Jin, et al., Personal privacy protection in big data era. J. Comput. Res. Dev. 52(1), 229–247 (2015).
[5] I. J. Cox, J. Kilian, F.T. Leighton, et al., Secure spread spectrum watermarking for multimedia. IEEE Trans. Image Process. 6, 1673–1687 (1997).
[6] G. Xuan, J. Zhu, J. Chen, et al., Distortionless data hiding based on integer wavelet transform. IEEE Electron. Lett. 38, 1646–1684 (2002).
[7] H. T. Wu, J.L. Dugelay, Y.Q. Shi, Reversible image data hiding with contrast enhancement. IEEE Signal Process Lett. 22, 81–85 (2015).
[8] K. Kurihara, S. Imaizumi, S. Shiota, et al., An encryption-then compression system for lossless image compression standards. IEICE Trans. Inf. Syst. E100.D(1), 52–56 (2017).
[9] B. Hu, Z. Guan, N. Xiong, et al., Intelligent impulsive synchronization of nonlinear interconnected neural networks for image protection. IEEE Trans. Indus. Inform. 14(8), 3775–3787 (2018).
[10] N. Bhargava, M.M. Sharma, A. S. Garhwal, et al., in Proceedings IEEE International Conference on Radar, Communication and Computing. Digital image authentication system based on digital watermarking (2013), pp. 185–189.
[11] Scrapy 1.6 documentation. https://docs.scrapy.org/en/latest/. Accessed 04 Feb. 2019.