

# Criminal Data Retrieval System (CDRS) Through Image Processing

Pawar Ruchita<sup>1</sup> Mulla Tanveer<sup>2</sup> Sachdev Jayshri<sup>3</sup> Randive Amol<sup>4</sup>  
<sup>1,2,3,4</sup>Brahmdevdada Mane Institute of Technology, Belati, Solapur, India

**Abstract**— Content Based Image Retrieval (CBIR) is a significant and increasingly popular approach that helps in the retrieval of image data from a huge collection. Image representation based on certain features helps in retrieval process. Three important visual features of an image include Color, Texture and Shape. We propose an efficient image retrieval technique which uses dynamic dominant color, texture and shape features of an image. We will encrypt content data (Color, Texture and Shape) in the hash algorithm is secure enough for critical purposes whether it is MD5 or SHA-1. This hashing algorithm used for only encryption this content cannot be decrypted. That is the image content data cannot be leaked or hacked. In this project of ours we aim to create an application that is helpful the police department to identify criminals and store and retrieve their criminal record data conveniently. This is called Content Based Image Retrieval (CBIR). Image retrieval with multiple features also known as Query by Image Content (QBIC) and the visual information is the application of computer vision to the image retrieval problem, that is, the problem of searching for digital images in large databases. “Image retrieval” means that the search will analyze the actual contents of the image. The term ‘content’ in this context might refer colors, shapes, textures, or any other information that can be derived from the image itself. Without the ability to examine image content, searches must rely on metadata such as captions or keywords. Such metadata must be generated by a human and stored alongside each image in the database. Problems with traditional methods of image indexing have led to the rise of interest in techniques for retrieving images on the basis of automatically-derived features such as color, texture and shape – a technology now generally referred to as Content-Based Image Retrieval (CBIR). However, the technology still lacks maturity, and is not yet being used on a significant scale. In the absence of hard evidence on the effectiveness of CBIR techniques in practice, opinion is still sharply divided about their usefulness in handling real-life queries in large and diverse image collections.

**Keywords:** CBIR, CDRS, MD5 Algorithm

## I. INTRODUCTION

The logical image representation in image databases systems is based on different image data models. An image object is either an entire image or some other meaningful portion (consisting of a union of one or more disjoint regions) of an image. The logical image description includes: Meta, semantic, color, texture, shape, and spatial attributes.

Color attributes could be represented as a histogram of intensity of the pixel colors. A histogram refinement technique is also used by partitioning histogram bins based on the spatial coherence of pixels. Statistical methods are also proposed to index an image by color corelograms, which is actually a table containing color pairs, where the k-th entry for  $\langle i, j \rangle$  specifies the probability of locating a pixel of color j at a distance k from a pixel of color I in the image [6].

## II. SYSTEM ARCHITECTURE

The proposed system architecture of our project will be like given diagram.

The admin panel will send the request for criminal register record, after that cloud server will permit the admin for receiving registered criminal.

In the criminal data search panel i.e. User panel sends the request for criminal data to cloud server and cloud server will response with data related to search.

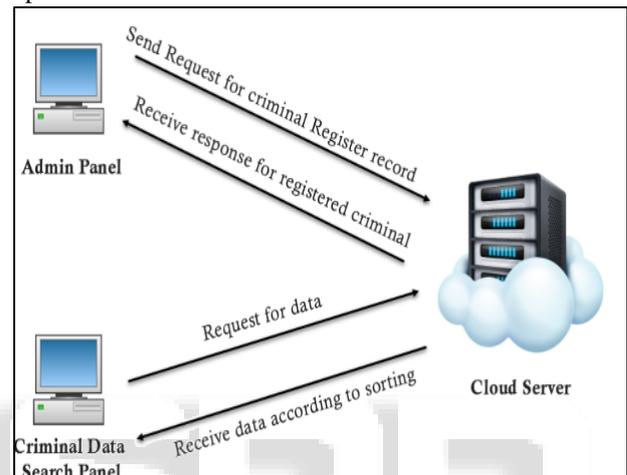


Fig. 1: System Architecture

Objective of this system is the extraction of text from any image and then displaying its related information on the mobile screen. Main goal of this system is that if a person doesn't have or know any specific thing then he/she could get its information with the help of this android application. Different modules used in this system are as follows:

### A. Text Extraction:

In text extraction feature text is being extracted from the natural scene or an image. Heretext extraction is done with the help of character description and stroke configuration [1]. Firstly the text will be detected, understood and then recognized.

### B. Searching:

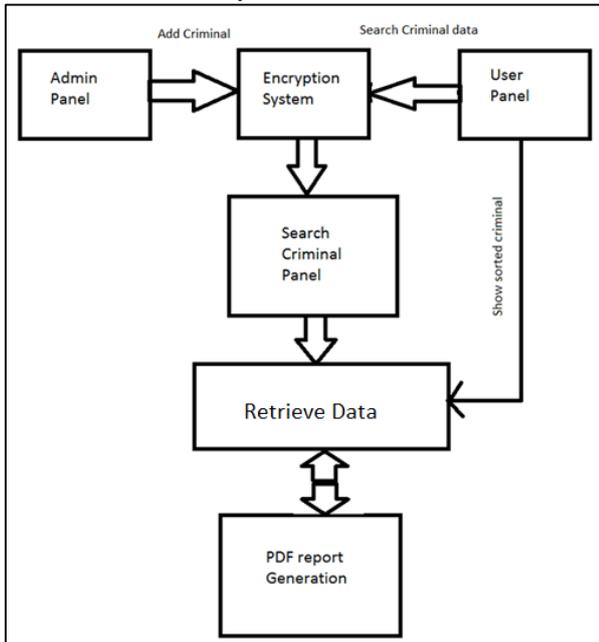
In searching process extracted text is being searched over net or in database. Here searching is done with the help of item ranking according to the item of interest. It basically derives meta data information about the item of interest by extending the user's given interest.

### C. Web Mining:

In this mining process required information is retrieved from the web or from database in an efficient manner. This is done with the help of Semantic and Synaptic web mining at low entropy [7]. After retrieving the information successfully it is displayed on the mobile screen.

### III. WORK CARRIED OUT

We describe the methods undertaken by us for our project Criminal data retrieval system.



From the above diagram we come to know that there are different panels like Admin Panel, User Panel, and Search Criminal Panel etc.

The Admin adds the Criminal data to the encryption system, then the data is sent to the Search Criminal Panel, after that the data is been retrieved and at the end it is checked by the PDF report generation

Or the other way can be the User Panel searches the criminal data in the encrypted system and it shows sorted criminal data then the data is been retrieved.

### IV. DECIDING THE ALGORITHM

In the past few years, there have been significant research advances in the analysis of hash functions and it was shown that none of the hash algorithm is secure enough for critical purposes whether it is MD5 or SHA-1. Nowadays scientists have found weaknesses in a number of hash functions, including MD5, SHA and RIPEMD so the purpose of this paper is combination of some function to reinforce these functions and also increasing hash code length up to 512 that makes stronger algorithm against collision attacks.

An Application of MD5 algorithm is implemented for the stream controlled transfer messages in the network. This would be a high security algorithm for data transfer in mobile networking with stream controlled logic. There may a vast number of applications for this algorithm in data transfer in various types of networks.

The MD5 message-digest algorithm is a widely used hash function producing a 128-bit hash value. Although MD5 was initially designed to be used as a cryptographic hash function, it has been found to suffer from extensive vulnerabilities.

#### A. Security:

The security of the MD5 hash function is severely compromised. A collision attack exists that can find collisions

within seconds on a computer with a 2.6 GHz Pentium 4 processor (complexity of 224.1). Further, there is also a chosen-prefix collision attack that can produce a collision for two inputs with specified prefixes within hours, using off-the-shelf computing hardware. The ability to find collisions has been greatly aided by the use of off-the-shelf GPUs. On an NVIDIA GeForce 8400GS graphics processor, 16–18 million hashes per second can be computed. An NVIDIA GeForce 8800 Ultra can calculate more than 200 million hashes per second.

These hash and collision attacks have been demonstrated in the public in various situations, including colliding document files and digital certificates. As of 2015, MD5 was demonstrated to be still quite widely used, most notably by security research and antivirus companies.

MD5 uses the Merkle–Damgård construction, so if two prefixes with the same hash can be constructed, a common suffix can be added to both to make the collision more likely to be accepted as valid data by the application using it. Furthermore, current collision-finding techniques allow to specify an arbitrary prefix: an attacker can create two colliding files that both begin with the same content. All the attacker needs to generate two colliding files is a template file with a 128-byte block of data, aligned on a 64-byte boundary that can be changed freely by the collision-finding algorithm. An example MD5 collision, with the two messages differing in 6 bits, is:

```

d131dd02c5e6ecc4 693d9a0698aff95c 2fcab58712467eab
4004583eb8fb7f89
55ad340609f4b302 83e488832571415a 085125e8f7cdc99f
d91dbdf280373c5b
d8823e3156348f5b ae6dacd436c919c6 dd53e2b487da03fd
02396306d248cda0
e99f33420f577ee8 ce54b67080a80d1e c69821bcb6a88393
96f9652b6ff72a70
d131dd02c5e6ecc4 693d9a0698aff95c 2fcab50712467eab
4004583eb8fb7f89
55ad340609f4b302 83e4888325f1415a 085125e8f7cdc99f
d91dbd7280373c5b
d8823e3156348f5b ae6dacd436c919c6 dd53e23487da03fd
02396306d248cda0
e99f33420f577ee8 ce54b67080280d1e c69821bcb6a88393
96f965ab6ff72a70
    
```

Both produce the MD5 hash 79054025255fb1a26e4bc422aef54eb4. The difference between the two samples is that the leading bit in each nibble has been flipped. For example, the 20th byte (offset 0x13) in the top sample, 0x87, is 10000111 in binary. The leading bit in the byte (also the leading bit in the first nibble) is flipped to make 00000111, which is 0x07, as shown in the lower sample.

Later it was also found to be possible to construct collisions between two files with separately chosen prefixes. This technique was used in the creation of the rogue CA certificate in 2008. A new variant of parallelized collision searching using MPI was proposed by Anton Kuznetsov in 2014, which allowed to find a collision in 11 hours on a computing cluster.

The main MD5 algorithm operates on a 128-bit state, divided into four 32-bit words, denoted A, B, C, and D. These are initialized to certain fixed constants. The main

algorithm then uses each 512-bit message block in turn to modify the state. The processing of a message block consists of four similar stages, termed rounds; each round is composed of 16 similar operations based on a non-linear function F, modular addition, and left rotation. Figure 1 illustrates one operation within a round. There are four possible functions; a different one is used in each round: crypt is the library function which is used to compute a password hash that can be used to store user account passwords while keeping them relatively secure (a passwd file). The output of the function is not simply the hash – it is a text string which also encodes the salt (usually the first two characters are the salt itself and the rest is the hashed result), and identifies the hash algorithm used (defaulting to the "traditional" one explained below). This output string is what is meant for putting in a password record which may be stored in a plain text file.

**B. Base 64:**

Base64 is a group of similar binary-to-text encoding schemes that represent binary data in an ASCII string format by translating it into a radix-64 representation. The term Base64 originates from a specific MIME content transfer encoding. Each Base64 digit represents exactly 6 bits of data. Three 8-bit bytes (i.e., a total of 24 bits) can therefore be represented by four 6-bit Base64 digits.

The particular set of 64 characters chosen to represent the 64 place-values for the base varies between implementations. The general strategy is to choose 64 characters that are common to most encodings and that are also printable. This combination leaves the data unlikely to be modified in transit through information systems, such as email, that were traditionally not 8-bit clean. For example, MIME's Base64 implementation uses A–Z, a–z, and 0–9 for the first 62 values. Other variations share this property but differ in the symbols chosen for the last two values; an example is UTF-7.

**Base64 table**

The Base64 index table:

Index	Char	Index	Char	Char
	Index	Char	Index	
0	A	16	Q	g
	48	w		
1	B	17	R	h
	49	x		
2	C	18	S	i
	50	y		
3	D	19	T	j
	51	z		
4	E	20	U	k
	52	0		
5	F	21	V	l
	53	1		
6	G	22	W	m
	54	2		
7	H	23	X	n

	55	3			
8	I	24	Y	40	o
	56	4			
9	J	25	Z	41	p
	57	5			
10	K	26	a	42	q
	58	6			
11	L	27	b	43	r
	59	7			
12	M	28	c	44	s
	60	8			
13	N	29	d	45	t
	61	9			
14	O	30	e	46	u
	62	+			
15	P	31	f	47	v
	63	/			

**1) Base64 can be used in a variety of contexts:**

Base64 can be used to transmit and store text that might otherwise cause delimiter collision

Spammers use Base64 to evade basic anti-spamming tools, which often do not decode Base64 and therefore cannot detect keywords in encoded messages.

Base64 is used to encode character strings in LDIF files  
Base64 is often used to embed binary data in an XML file, using a syntax similar to `<data encoding="base64">...</data>` e.g. favicons in Firefox's exported bookmarks.html.

Base64 is used to encode binary files such as images within scripts, to avoid depending on external files.

The data URI scheme can use Base64 to represent file contents. For instance, background images and fonts can be specified in a CSS stylesheet file as data: URIs, instead of being supplied in separate files.

The Free SWAN IPsec implementation precedes Base64 strings with 0s, so they can be distinguished from text or hexadecimal strings.

**2) Decoding Base64 with padding**

When decoding Base64 text, four characters are typically converted back to three bytes. The only exceptions are when padding characters exist. A single = indicates that the four characters will decode to only two bytes, while == indicates that the four characters will decode to only a single byte. For example:

Encoded	Padding	Length	Decoded
YW55IGNhcm5hbCBwbGVhcw==	1	any	carnal pleas
YW55IGNhcm5hbCBwbGVhc3U==	2	any	carnal pleasu
YW55IGNhcm5hbCBwbGVhc3VyNone	3	any	carnal pleasure

**3) Decoding Base64 without padding**

Without padding, after normal decoding of four characters to three bytes over and over again, fewer than four encoded characters may remain. In this situation only two or three

characters shall remain. A single remaining encoded character is not possible (because a single Base64 character only contains 6 bits, and 8 bits are required to create a byte, so a minimum of 2 Base64 characters are required: The first character contributes 6 bits, and the second character contributes its first 2 bits. For example:

Length	Encoded	Length	Decoded	
2	YW55IGNhcm5hbCBwbGVhcw	1	any	
carnal pleas				
3	YW55IGNhcm5hbCBwbGVhc3U	2	any	
carnal pleasu				
4	YW55IGNhcm5hbCBwbGVhc3Vy3	any		
carnal pleasur				

## V. SELECTION LANGUAGE

### A. Programming Language: PHP

PHP is a server-side scripting language designed for web development but also used as a general-purpose programming language. Originally created by Rasmus Lerdorf in 1994, the PHP reference implementation is now produced by The PHP Group. While PHP originally stood for Personal Home Page, it now stands for the recursive backronym PHP: Hypertext Preprocessor.

PHP code may be embedded into HTML code, or it can be used in combination with various Web template systems and web frameworks. PHP code is usually processed by a PHP interpreter implemented as a module in the web server or as a Common Gateway Interface (CGI) executable. The web server combines the results of the interpreted and executed PHP code, which may be any type of data, including images, with the generated web page. PHP code may also be executed with a command-line interface (CLI) and can be used to implement standalone graphical applications. 10 standard PHP interpreter, powered by the Zend Engine, is free software released under the PHP License. PHP has been widely ported and can be deployed on most web servers on almost every operating system and platform, free of charge. The PHP language evolved without a written formal specification or standard until 2014, leaving the canonical PHP interpreter as a de facto standard. Since 2014 work has been ongoing to create a formal PHP specification.

PHP (recursive acronym for PHP: Hypertext Pre-processor) is a widely-used open source general-purpose scripting language that is especially suited for web development and can be embedded into HTML.

Designing and Development are the steps that are important. PHP Programming the Languages mostly commonly used for Website and Web Application Development. PHP is a general purpose, server-side scripting language run a web server that's designed to make dynamic pages and applications.

When PHP receives the file, it reads through it and executes any PHP code it can find. After it is done with the file, the PHP interpreter gives the output of the code, if any, back to Apache. When Apache gets the output back from PHP, it sends that output back to a browser which renders it to the screen.

The designing and development are the two most promising steps which are crucial. ... It is given a thought to what has made PHP programming language as one of the

most commonly used languages and is also a good choice for web development. Using PHP as its language has many advantages like it supports Oracle, Sybase, etc There are a few of these to consider, including:

LAMP: Linux, Apache, MySQL and PHP.

WIMP: Windows, IIS, MySQL/MS SQL Server and PHP.

WAMP: Windows, Apache, MySQL/MS SQL Server and PHP.

LEMP: Linux, NGINX, MySQL and PHP.

Stands for "Windows, Apache, MySQL, and PHP." WAMP is a variation of LAMP for Windows systems and is often installed as a software bundle (Apache, MySQL, and PHP). It is often used for web development and internal testing, but may also be used to serve live websites.

Anything. PHP is mainly focused on server-side scripting, so you can do anything any other CGI program can do, such as collect form data, generate dynamic page content, or send and receive cookies. But PHP can do much more.

There are three main areas where PHP scripts are used.

Server-side scripting. This is the most traditional and main target field for PHP. You need three things to make this work: the PHP parser (CGI or server module), a web server and a web browser. You need to run the web server, with a connected PHP installation. You can access the PHP program output with a web browser, viewing the PHP page through the server. All these can run on your home machine if you are just experimenting with PHP programming. See the installation instructions section for more information. Command line scripting. You can make a PHP script to run it without any server or browser. You only need the PHP parser to use it this way. This type of usage is ideal for scripts regularly executed using cron (on \*nix or Linux) or Task Scheduler (on Windows). These scripts can also be used for simple text processing tasks. See the section about Command line usage of PHP for more information.

PHP can be used on all major operating systems, including Linux, many Unix variants (including HP-UX, Solaris and OpenBSD), Microsoft Windows, macOS, RISC OS, and probably others. PHP also has support for most of the web servers today. This includes Apache, IIS, and many others. And this includes any web server that can utilize the Fast CGI PHP binary, like lighted and nginx. PHP works as either a module, or as a CGI processor. So, with PHP, you have the freedom of choosing an operating system and a web server. Furthermore, you also have the choice of using procedural programming or object-oriented programming (OOP), or a mixture of them both.

With PHP you are not limited to output HTML. PHP's abilities include outputting images, PDF files and even Flash movies (using libswf and Ming) generated on the fly. You can also output easily any text, such as XHTML and any other XML file. PHP can autogenerate these files, and save them in the file system, instead of printing it out, forming a server-side cache for your dynamic content. One of the strongest and most significant features in PHP is its support for a wide range of databases. Writing a database-enabled web page is incredibly simple using one of the database specific extensions (e.g., for MySQL), or using an abstraction layer like PDO, or connect to any database supporting the Open Database Connection standard via the ODBC

extension. Other databases may utilize curl or sockets, like CouchDB.

PHP also has support for talking to other services using protocols such as LDAP, IMAP, SNMP, NNTP, POP3, HTTP, COM (on Windows) and countless others. You can also open raw network sockets and interact using any other protocol. PHP has support for the WDDX complex data exchange between virtually all Web programming languages. Talking about interconnection, PHP has support for instantiation of Java objects and using them transparently as PHP objects.

PHP has useful text processing features, which includes the Perl compatible regular expressions (PCRE), and many extensions and tools to parse and access XML documents. PHP standardizes all of the XML extensions on the solid base of libxml2, and extends the feature set adding SimpleXML, XMLReader and XMLWriter support.

### B. XAMPP

XAMP is a free and open source cross-platform web server solution stack package developed by Apache Friends, consisting mainly of the Apache HTTP Server, MariaDB database, and interpreters for scripts written in the PHP and Perl programming languages. XAMPP stands for Cross-Platform (X), Apache (A), MariaDB (M), PHP (P) and Perl (P). It is a simple, lightweight Apache distribution that makes it extremely easy for developers to create a local web server for testing purposes. Everything needed to set up a web server – server application (Apache), database (MariaDB), and scripting language (PHP) – is included in an extractable file. XAMPP is also cross-platform, which means it works equally well on Linux, Mac and Windows. Since most actual web server deployments use the same components as XAMPP, it makes transitioning from a local test server to a live server extremely easy as well.

XAMPP is regularly updated to incorporate the latest releases of Apache, MariaDB, PHP and Perl. It also comes with a number of other modules including OpenSSL, phpMyAdmin, MediaWiki, Joomla, WordPress and more. Self-contained, multiple instances of XAMPP can exist on a single computer, and any given instance can be copied from one computer to another. XAMPP is offered in both a full and a standard version (Smaller version).

Officially, XAMPP's designers intended it for use only as a development tool, to allow website designers and programmers to test their work on their own computers without any access to the Internet. To make this as easy as possible, many important security features are disabled by default. XAMPP has the ability to serve web pages on the World Wide Web. A special tool is provided to password-protect the most important parts of the package. XAMPP also provides support for creating and manipulating databases in MariaDB and SQLite among others. Once XAMPP is installed, it is possible to treat a localhost like a remote host by connecting using an FTP client. Using a program like FileZilla has many advantages when installing a content management system (CMS) like Joomla or WordPress. It is also possible to connect to localhost via FTP with an HTML editor.

### C. DATABASE: MySQL

MySQL is an open-source relational database management system (RDBMS), in July 2013, it was the world's second most widely used RDBMS, and the most widely used open-source client-server model RDBMS. It is named after co-founder Michael Widenius's daughter, My. The SQL acronym stands for Structured Query Language. The MySQL development project has made its source code available under the terms of the GNU General Public License, as well as under a variety of proprietary agreements. MySQL was owned and sponsored by a single for-profit firm, the Swedish company MySQL AB, now owned by Oracle Corporation. For proprietary use, several paid editions are available, and offer additional functionality.

MySQL is a popular choice of database for use in web applications, and is a central component of the widely used LAMP open source web application software stack (and other "AMP" stacks). LAMP is an acronym for "Linux, Apache, MySQL, Perl/PHP/Python." Free-software-open source projects that require a full-featured database management system often use MySQL. Applications that use the MySQL database include: TYPO3, MODx, Joomla, WordPress, phpBB, MyBB, Drupal and other software. MySQL is also used in many high-profile, large-scale websites, including Google (though not for searches), Facebook, Twitter, Flickr, and YouTube.

On all platforms except Windows, MySQL ships with no GUI tools to administer MySQL databases or manage data contained within the databases. Users may use the included command line tools, or install MySQL Workbench via a separate download. Many third party GUI tools are also available. MySQL is written in C and C++. Its SQL parser is written in yacc, but it uses a home-brewed lexical analyzer. MySQL works on many system platforms, including AIX, BSDi, FreeBSD, HP-UX, eComStation, i5/OS, IRIX, Linux, OS X, Microsoft Windows, NetBSD, Novell NetWare, OpenBSD, OpenSolaris, OS/2 Warp, QNX, Oracle Solaris, Symbian, SunOS, SCO OpenServer, SCO UnixWare, Sanos and Tru64. A port of MySQL to OpenVMS also exists.

### D. Web browser: Google Chrome

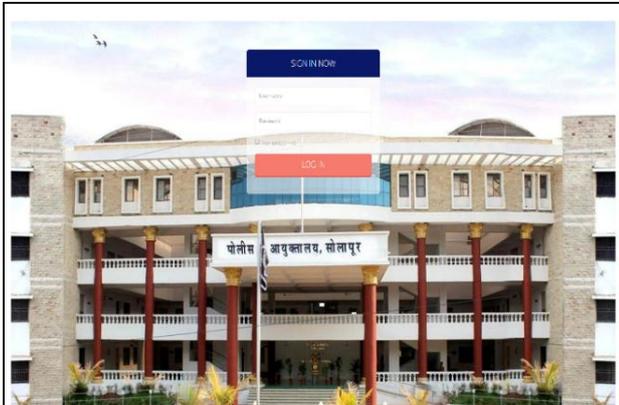
Google Chrome is a freeware web browser developed by Google. It used the WebKit layout engine until version 27 and with the exception of its iOS releases, from version 28 and beyond uses the WebKit fork Blink. It was first released as a beta version for Microsoft Windows on September 2, 2008, and as a stable public release on December 11, 2008. As of December 2015, Stat Counter estimates that Google Chrome has a 58% worldwide usage share of web browsers as a desktop browser. It is also the most popular browser for smartphones, and combined across all platforms at about 45%. Its success has led to Google expanding the 'Chrome' brand name on various other products such as the Chromecast. Google releases the majority of Chrome's source code as an open-source project Chromium.

## VI. TRIALS AND TESTING

Software testing is a process of executing a program or application with the intent of finding the software bugs. It can also be stated as the process of validating and verifying that

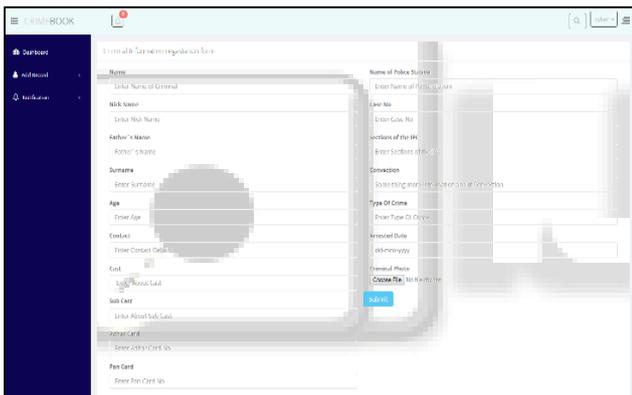
a software program or application or product: Meets the business and technical requirements that guided its design and development. In our software we providing validation for registration and Login forms because some like image or any photo.

### A. Snapshots of Project output



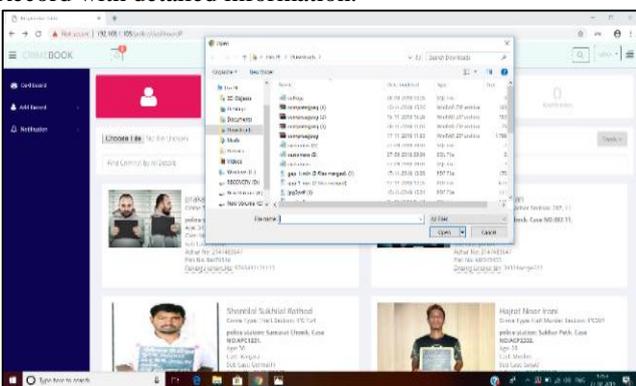
### B. Login Page

As we start our project, this is the login page. This authority is with the admin, and this can also be logged in through the user also.



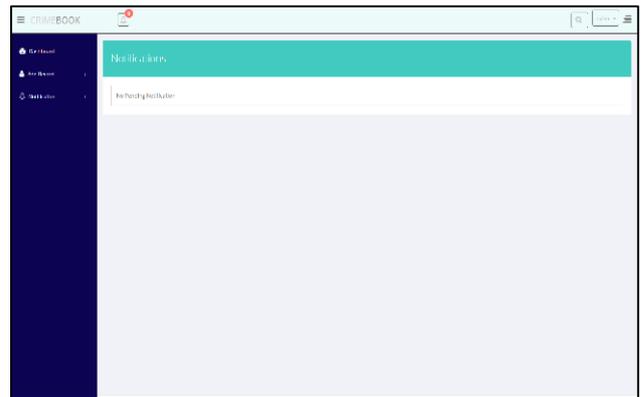
### C. Add Records Form

This is another form which tells us that we can add the criminals Record with detailed information.



### D. Searching form

This is the image searching process that is we choose a criminal Image from a folder and add it.



### E. Notification form

This is the notification form in this we can notify with the changes or Requirement of the criminal record.

## VII. RESULT AND DISCUSSION

Image retrieval with multiple features also known as Query by Image Content (QBIC) and the visual information is the application of computer vision to the image retrieval problem, that is, the problem of searching for digital images in large databases.

“Image retrieval” means that the search will analyze the actual contents of the image. The term 'content' in this context might refer colors, shapes, textures, or any other information that can be derived from the image itself. Without the ability to examine image content, searches must rely on metadata such as captions or keywords. Such metadata must be generated by a human and stored alongside each image in the database.

The logical image representation in image databases systems is based on different image data models. An image object is either an entire image or some other meaningful portion (consisting of a union of one or more disjoint regions) of an image. The logical image description includes: Meta, semantic, color, texture, shape, and spatial attributes.

The paper starts with discussing the fundamental aspects of CBIR. Another important issue in content-based image retrieval is effective indexing and fast searching of images based on visual features. Some fundamental techniques for content-based image retrieval, including visual content description, similarity/distance measures, indexing scheme, user interaction were introduced.

Therefore in criminal data retrieval system searching of image with its information is achieved.

## VIII. CONCLUSION/SUMMARY

The proposed project allows us to search criminal records and add new records of criminals. This will avoid wastage of papers and big files and mainly man work to find old criminal records .It will be Beneficial for users as it provide exact information regarding criminals by using image search.

## REFERENCES

[1] Eakins, John; Graham, Margaret. "Content-based Image Retrieval". University of North Umbria at Newcastle.

Archived from the original on 2012-02-05. Retrieved 2014-03-10.

- [2] Swapnalini Pattanaik<sup>1</sup>, Prof.D.G.Bhalke, Beginners to Content Based Image Retrieval, Department of Electronics and Telecommunication Tathawade, Pune, India
- [3] Ciampa, Mark Australia; United States: Course Technology/Cengage Learning. p.290.

