



# AN IMPLEMENTATION PAPER ON A FRAMEWORK AND TECHNIQUES FOR IMAGE-BASED SEARCH APPLICATION WITH AN E-COMMERCE DOMAIN

Yuvraj, Ishu Bansal, Sakshi Dhar, Gitanjali Nikam

**Abstract**-A large image database is growing rapidly as billions of images transferred every day at various chronicles like Instagram, snapchat, Facebook etc. and expected to continue in the future. With a particularly immense measure of data, the requirement for a successful pursuit in the image ages. Furthermore, if astounding apparatuses have effectively been made for text search, image search stays an uncertain issue.

Traditionally, most end-users use retrieval systems to write questions or queries and get text results. But end-users continuously expect search engines to be "intelligent" and they want a retrieval system to explore cyberspace using an image, as with the built-in camera on a mobile phone to send relevant resemblance, videos, and other formats. So, it becomes a major challenge to retrieve and processing images from this large database. Our paper introduced a novel methodology in content-based image retrieval (CBIR) by consolidating the low level component for example shape, texture and color features. CBIR helps in retrieving images from an huge database in an efficient manner it is fastest growing research area in this field. The study found that convolutional neural networks can use to break down divisions and retrieval issues. Image details are much bigger than text data, and we cannot identify ocular particular by old techniques designed to identify text details. Therefore, CBIR has acquired significant benefits from research society. In this paper we aim to frame the feature vectors of the entire image by consolidating the shape, texture and color features and to get a decent performance in terms of the precision and recall. The image retrieval system is used to find images as per user request from the database. In this paper we also explored the flutter framework with dart language. We aim to mitigate the fundamental issue that app developers have been facing for so long- maintaining multiple apps for multiple platforms.

**Keywords** – deep\_learning, C.N.N., C.B.I.R., Flutter, Image Processing, Feature extraction, HSV Color Histogram.

## I. INTRODUCTION

Currently, search engines play a vital part in every human life. According to google, google operates over 42,000 searches every second on average. As cyberspace contains an enormous amount of data along this data is increasing exponentially. As reported by the microfocus, information produced each day is about 2.5 X 10<sup>18</sup> bytes. Not just smartphones and laptops, but television, refrigerators, IoT devices, and anything linked to cyberspace are a cause of creating data.

Currently, the information itself does not restrict to the text form. Figures can exist in any form like text, tables, graphs, images, audios, and videos. These various forms and such massive data would lead to increase user expectations. But end- users continuously expect search engines to be "intelligent" and to be able to manage more than just written queries. An image is worth a ten hundred words and those thousand words now worth of thousand of searches for a resemblance and Humans eyes doesn't have an edge to capture such small details and classify and find thousands of images related to it. Portrayals have an irreplaceable role in all multimedia applications. Previously, the image search engine indicated text-based image retrieval. Unfortunately, this program has many drawbacks. The text description cannot capture visual content. Image illustration may be exhibit in heterogeneous ways are beyond the meaning of the words. To subdue the problems substitute search procedure developed for image mining named Content-Based Content.

Image searches became an essential activity of web consumers. According to Moz.com, 27% of web queries go through Google Images. It is difficult for consumers to find their desired product. Most of the content-based search engines only extract images by their color. These search engines also don't have an amount of data related to the needs of consumers. The existing online e-commerce platform has a limitation that a consumer cannot purchase products from a shop. They do not go to the shop to check the quality of the product before buying. Consumers also had to wait for the delivery of their ordered products.

Also, developing an e-commerce app is not an easy task. Developers are facing a fundamental issue for so long in maintaining multiple apps for multiple platforms. Later on, many cross-platform & hybrid app development solutions have surfaced. They compromise on performance and developer productivity.

This paper represents the importance of CBIR (Content-based image retrieval) in e-commerce. How it works, and the benefits of the flutter framework over other hybrid/native approaches.

### A. MOTIVATION

Nowadays, the entire variety of images transferred every day to varied on-line stockpiles is calculable at billions. The main motivation of this paper is within the sight of such an enormous quantity of knowledge, the issue of checking out pictures not solely by keywords, yet additionally by the duplicate of the picture is essential. As there were an old saying “Time is money”. So, nowadays users want more fast and accurate results. Traditionally, users use retrieval system to write question or queries to get results (i.e. Text-based retrieval system) which consumes much time to find desired result from the given output.

Most of the time user don't recall a word for an object, that's why we need search engine to be “smarter” which can handle more than just text queries and take visual parameters like images as input. Visual attributes can enhance the performance of the search engine by retrieving more accurate results. Hence, we need content-based retrieval system to fasten up the process of searching in terms of time and accuracy.

### B. STATISTICS OF SOCIAL MEDIA USAGE

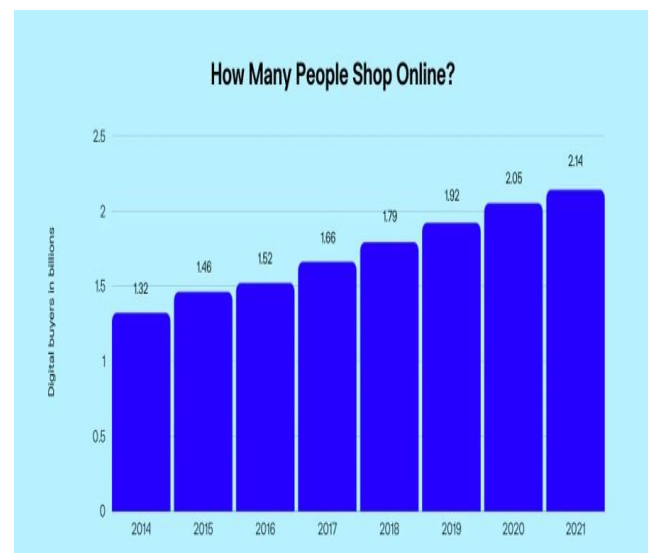
There were 2.4 billion web clients in 2014. The figure in 2016 was 3.4 billion and 300 million web clients were added till 2017. There are presently over 4.4 billion web clients as of June 2019. This is a 83% expansion in the quantity of individuals utilizing the web in only five years. There are 2.5 quintillion bytes of data made each day. On normal, Google currently measures in excess of 40,000 pursuits each second (i.e., 3.5 billion hunts each day).

These are a few numbers created all day long by various stages:

- 527,760 photographs are shared by Snapchat clients
- More than 120 LinkedIn accounts are created
- 4,146,600 YouTube recordings are watched
- 456,000 tweets are posted on Twitter

- 46,740 photographs are posted by Instagram clients .

With this expanding number of web clients it's assessed that there will be 2.14 billion worldwide advanced purchasers in 2021. This number has been rising quickly throughout the long term and is required to keep doing so. 81 percent of customers direct online exploration prior to making a buy online. One of the most obvious patterns in the realm of web based business is the phenomenal use of cell phones. In 2019, cell phones represented more than 60% of all retail site visits around the world. As the reception of cell phones is advancing at a fast speed, particularly in districts that need other computerized framework, portable mix will keep on forming the shopping experience of the future. With buyers progressively depending on internet shopping — it is assessed that 95% of buys will be made online by 2040. Online customers make buys through versatile more regularly than on PC. It is normal that versatile trade will take a piece of the overall industry of web based business of 72.9% by 2021.



*Fig. 1.1 stats of online buyer*

### C. AIM OF THE STUDY

The method of exploring and extracting an image from a database is known as image retrieval. The process consists of retrieving and processing the image. General image content containing both visual and embedded content. Visual content is more common and more specific to the background having color, texture, relationship. The aim of this paper includes:

- To explore various studies carried out in image-based retrieval.



- Explain the importance/ need of CBIR.
- Explain the working of the CBIR algorithm.

#### D. BENEFITS

*Scalability:* It can be a vital aspect in terms of ensuring lower computation cost in proposed work and it also ensure that the helpful information won't be bought from some place rather will be separated utilizing APIs from the various platforms.

*Real time analysis:* It leads to real-time retrieval of content that is to design a framework that can retrieve accurate results during a search. It is also a key factor in human perception.

## II. BACKGROUND

### A. TYPES OF IMAGE RETRIEVAL SYSTEM

A large amount of information on the web is stored in multimedia content, especially photographs. There is currently a focus on image search and there are three main ways to search for images.

- **Text Based Image Retrieval:** A text-based retrieval framework is one that utilizes the content identified with the image like the image record name and meta data acclimated portray the picture that is installed inside the markup language report. A client's inquiry is as a watchword or assortment of catchphrases. all through the recovery technique, the inquiry is contrasted and each text depiction and pictures whose text portrayal is most similar to the client's inquiry square measure recovered. Consequently, basically, the text based generally image retrieval framework utilizes common report recovery procedures. This can likewise be referred to as query by-text. Yahoo and Google square measure the preeminent in vogue and fruitful Image web indexes nowadays. each these web crawlers use text based generally image watching out.
- **Content Based Image Retrieval:** CBIR upholds picture substance or subordinate image choices. These alternatives exemplify shading, surface and structure contained in the picture. One altogether these choices or a blend of those choices is utilized to file pictures inside the picture information. Questions are communicated exploitation a model picture. This could be referred to as query-by-example.
- **Collection Based Image Retrieval:** Moreover, on top of 2 a third class exists that are known as Collection Based Image search engines. Collection-based engines attract with respect to a more modest pool of recorded pictures, once in a while from a data and for the most part filed by people. Huge advanced libraries and business

picture or clasps craftsmanship providers are shrewd examples of these collection-based search engines. Corbis is AN illustration of this class.

### B. DEEP LEARNING

Deep learning (deep organized learning, hierarchic learning or deep AI) is a part of AI upheld by a gathering of calculations that intend to display undeniable level deliberations in information by exploitation numerous interaction layers with confounded constructions, or in any case made out of different non-direct changes. Deep learning is a part of a more extensive group of AI methodologies upheld learning portrayals of information. Partner degree perception (e.g., a picture) will be depicted somehow or another like a vector of force esteems per picture component, or during an extra dynamic way as a gathering of edges, districts of explicit structure, and so forth A few portrayals assemble it simpler to learn undertakings (e.g., face acknowledgment or facial highlights acknowledgment ) from models. One of the possibilities of deep learning is recompense high quality choices with practical calculations for solo or semi-directed element learning and hierarchic component extraction .

Different deep learning models like deep neural organizations, convolutional deep neural organizations, deep conviction organizations and ceaseless neural organizations are applied to fields like PC vision, programmed discourse acknowledgment, etymological correspondence measure, sound acknowledgment and bioinformatics any place they need been appeared to give reformist outcomes on various assignments.

Deep learning can be described as a class of AI calculations that Use a course of the numerous layers of nonlinear cycle units for highlight extraction and change. each sequent layer utilizes the yield from the past layer as info. The calculations is additionally directed or unattended and applications embrace design investigation (unaided) and characterization (managed). More elevated level choices region unit got from lower level choices to make a hierarchic portrayal. Territory unit a piece of the more extensive AI field of learning portrayals of information. Get familiar with numerous degrees of portrayals that relate to entirely unexpected degrees of reflection; the levels kind a progressive system of thoughts.

These depictions share essentially:

- Several coating of the nonlinear interaction sections.
- The directed or unattended learning of feature depictions in each layer, with the layers outlining a hierarchy of leadership from low-level to verifiable level other options. The piece of a layer of nonlinear cycle units used in a significant learning rule depends upon the have an effect to be settled. Layers that are used in significant learning embrace concealed layers of an engineered neural organization and sets of complex propositional plans. They'll conjointly

acknowledge torpid variables facilitated layer-wise in significant generative models like the centers in Deep Belief Networks and Deep Ludwig Boltzmann Machines.

*Neural networks*

A neural network take its premise from the human sensory system where neurons go about as the principle data handling framework and the neurotransmitters between them are choose how and what measure of data will move between various neurons. This straightforward framework is equipped for tackling issues of any intricacy. The primary numerical model of a solitary neural network was created in the year 1943 by McCulloch and Pitts known as MCP neuron and every one of the neural network utilized today are altered adaptations of MCP.

Neural network is an network which contains a few hubs (neurons) present in each layer. Every hub of a layer is associated with each and every hub in the following layer and each edge interfacing them has a weight allocated to them. Actuation of every hub in the following layer relies upon the sigmoidal capacity which registers information in every hub in past layer by utilizing the loads of the edges interfacing them and some predisposition to give the enactment of the following hub. The hubs in layer one actuate the hubs in layer two and further the cycle proceeds until at long last a hub in the last layer is enacted, which is yield. A Neural Network contains these three layers:

- **Input Layer :** The input layer, for every perception gets the estimations of illustrative credits as the information esteem. By and large, estimation of informative ascribes are equivalent to the info hubs. The information isn't changed by the hubs of the info layer. The hubs get a solitary worth on their info and copy them to numerous yields and copies each worth and ships off completely covered up hubs. This layer simply serves the capacity of giving the data to covered up layers. No calculations are being performed here.
- **Hidden Layer:** The secret layer is liable for applying the changes to the info esteems. The name is covered up layer since the hubs in this layer are stowed away from the world. The neural organization abstracts them.
- They structure the computational unit of the neural network and henceforth measure the contribution to the yield.
- **Output Layer:** This is the substance of neural network for the external world. The data figured by the organization is introduced to the world through this layer. Secret layer is connected to the yield layer. Relating to the forecast of reaction variable it returns the estimation of a yield. In this layer, information is changed by the dynamic hubs who consolidates themselves to create the yield.

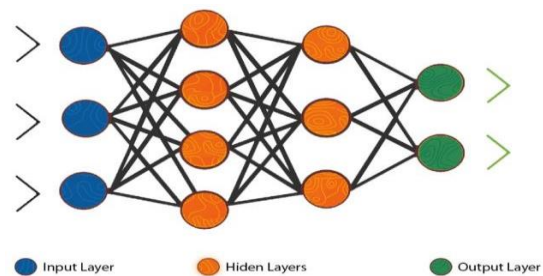
Three significant sorts of neural organizations that structure the reason for most pre- prepared models in profound learning:

- Artificial Neural Networks
- Convolution Neural Networks
- Recurrent Neural Networks

*Artificial Neural Networks (ANN)*

An anti-social neuron can be envisioned as a Logistic Regression. ANN is a accumulation of various perceptrons at each level. ANN is otherwise called a Feed- Forward Neural network since inputs are prepared distinctly the progressive ways:

- ANN comprises of 3 layers – Input, Hidden and Output. The information layer recognizes the information sources, the secret layer estimates the wellsprings of data, and the yield layer conveys the result. Fundamentally, each layer endeavors to get comfortable with explicit burdens.
- ANN can be utilized to take care of issues identified with:
  - Tabular information
  - Image information
  - Text information



*Fig 2.1. Layers of ANN*

*Advantages of Artificial Neural Network*

- Artificial Neural Network is fit for learning any nonlinear capacity. Consequently, these organizations are famously known as Universal Function Approximators. ANNs do have the ability to grasp loads that plot any contribution to the result.
- The principle purposes for all inclusive estimation is the activation function. Actuation capacities acquaint nonlinear properties with the organization. This assists the organization with learning complex connection among input and output.

*Challenges with Artificial Neural Network*

- Despite taking care of a picture grouping issue utilizing ANN, initial step is to change over a 2-d picture into a 1-d vector before preparing the model. This has two

downsides:

- The figure of teachable boundaries increments radically with an increment in the size of the picture
- ANN loses the contiguous spotlights of a image. Contiguous spotlights allude to the plan of the pixels in a picture.
- One ordinary issue in general these neural organization is the Vanishing and Exploding Gradient. This issue is connected with the backpropagation calculation. The heaps of a neural organization are revived through this backpropagation estimation by finding the points.
- ANN can't get progressive information in the data which is required for overseeing course of action data.

*Recurrent Neural Network (RNN)*

On the hidden state, RNN has a redundant relationship. This surrounding impediment ensures that progressive information is trapped in the information.

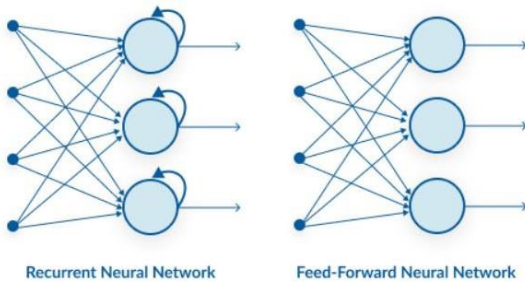


Fig. 2.2. RNN and Feed-Forward NN

We can utilize repetitive neural organizations to tackle the issues identified with:

- Time Series information
- Text information
- Audio information

*Merits of RNN*

- In the information, RNN catches the consecutive data for example reliance on word content in while performing expectations.
- At different time intervals RNNs share the boundaries. This results in lower limits to get ready and decreases the computational cost.

*Challenges with Recurrent Neural Networks*

- Deep RNNs moreover experience the evil impacts of the vanishing and exploding tendency which is a common problem that involves everything together the different kinds of N.N.

*Convolution Neural Network (CNN)*

CNN are for the most part the wrath in the profound learning local area this second. Various spaces and applications are utilized across these CNNs, but they are especially prevalent in video and image preparing projects.

Channels, also known as kernels, are the structure squares of CNNs. Using the convolution activity, kernels are used to extract the relevant features from the results.

Convolutional neural networks are well-versed in dealing with problems related to image data, and they also perform admirably on consecutive contributions.

*Advantages of Convolution Neural Network (CNN)*

- CNN picks up on the networks without explicitly mentioning them. These networks assist in removing the information's rights and relevant features.
- CNN identifies the spatial features in a photograph. The game plan of pixels and the relation between them in a picture are referred to by spatial features. They assist us in precisely identifying the object, the area of an article, and its relationship to other articles in an image.
- CNN also adheres to the principle of boundary sharing. To deliver an element map, a single channel is applied through different pieces of a contribution.

TABLE 2.1. Comparison between MLP, RNN and CNN

	MLP	RNN	CNN
Data	Tabular data	Sequence data(Time series, Text, Audio)	Image data
Recurrent connections	No	Yes	No
Parameter sharing	No	Yes	Yes
Spatial relationship	No	No	Yes
Vanishing and Exploding Gradient	Yes	Yes	Yes

**C. FLUTTER FRAMEWORK**

A method that's used for making mobile applications is by making them for a particular platform from the start . Another methodology is to write down a code base that may be compiled to many styles of platforms. it's {a



methodology| a way | a technique } known as cross-platform and may be a widespread method due to however versatile and quick a mobile application will be created.

Cross-platforming refers to a product or package that may be used on another platform than what it absolutely was developed for. In app development the principle of cross-platforming is to make and maintain only 1 code base, that is that the alluring part of victimization it since it saves time in development compared to native that's restricted to only 1 platform per code base. Samples of cross-platform frameworks/tools are Flutter, React Native and Ionic.

Flutter is an open-source framework for making apps that work best for both android and iOS. It uses one codebase for the applications for mobile, web, and desktop. Essential features of flutter contain AOT compilation and just-in-time compilation. The hot reload feature allows us to experiment, add features, build UI, fix bugs, and present them. It injects the uploaded ASCII text file into the DVM (Dart Virtual Machine). The framework redesigned the widget tree allowing it quickly to view results . It has an upscale number of widgets. There are two kinds of widgets countless widgets and state widgets in flutter .

Dart is used to writing all apps in flutter. Google created dart language to do programming & it is also conserved by google. Views in all-new frames with flutter are updated. This behaviour leads back to the fact that tons of objects which can hold less than one formed. Dart is designed to address this situation at the memory level with the aid of the "Generational Garbage Collection".

Flutter is a new tool and it continues to grow improbably quick. Conclusions are drawn that a Flutter application will contend with a native application once it involves central processor performance, however isn't as developed within the animation space. Flutter doesn't need advanced code for making an easy application and uses considerably less lines of code in development compared to native. the ultimate conclusion is that Flutter is best to use once building smaller to medium- sized applications, however encompasses a potential to grow to beat its current drawbacks within the animation department.

One of the papers (Heitkötter et al., 2012) discusses the variations between cross-platforms. The papers results showed that cross-platforms were higher for shorter time and budget. Native were most well-liked once interacting with the phone's integrated system.

A paper written by Amatya 2013, brings up the variations in cross-platforms and native by researching the platform, development setting and code base. Amatyas conclusion was that native may be a sensible acceptable heavier applications however isn't continuously the foremost appropriate alternative for all applications due to it being value serious and longer overwhelming than cross-compilers.

There is not rather more difficult code that's required to make a Flutter application compared to a native (both

operational systems), there are largely a distinction within the split and connections of the visual and therefore the code elements that is less complicated in Flutter for the foremost half.

*Comparison of React and Flutter: -*

- **Modularity**  
Both React Native and Flutter application regard the arranging of measured quality. That implies as opposed to composing applications page by page, engineers must be constrained to decouple pages into components thus collect them in sync with the needs. exploitation the format assessor, one will basically notice anyway a page is made by quantities of components.
- **Styling**  
React Native and Flutter holds totally various sentiments identifying with what styling essentially is and the manner in which it should be applied to sees. React, as a base structure of React Native, includes a powerful effect on anyway React Native applications are coordinated and created. Flutter, on the other hand, establishes and enforces explicitly the rule that gadget is the one and only ordered class for all, and that interpretations should be rendered and handled equally to look at components.
- **Scrolling**  
Both Flutter and React Native do radiant occupation identifying with to looking over. without a doubt circumstance, as over speed looking over, clear squares are delivered as placeholders and accordingly the Federal Protective Service in JavaScript string drops extensively. Flutter's Federal Protective Service, on the other hand, is somewhat secure.

### III. LITERATURE REVIEW

The major keywords we used in our research were deep\_learning, C.N.N., C.B.I.R., flutter, Image Processing, Feature extraction, HSV Color Histogram. To reduce the number of papers we did skimming of the papers distributed after 1985 and rejected those that did not match our problem statement.

We also scanned the papers by filtering its contents and accumulated 55 papers for in-depth reading. Out of the 55 papers we have chosen those 29 papers corresponding to our problem and strongly recommend a comprehensive review of the literature. We have cumulated the concise rundown of every one of those papers in section 3.1.

#### A. RELATED WORK

The present-day alternative of stochastic gradient descent algorithm used to form deep neural networks.

It involves arithmetic congestion. Two strategies of the simulation Model-Data, as well as Data uniformity, are applied for the parallel deep neural network framework. Data Parallelism partitioning data equally among several processing nodes (GPU) and Model-data Parallelism splits the model across multiples GPUs and executes each part on a different accelerator.[4]

Use of high-resolution images because it works more appropriately than single-scale training. In particular for determining small objects. On utilizing a high-resolution image of 1024x1024 pixels, lots of features can raise by every 512x512 crop. If ambient lighting switches, an entity will not switch its class. Specifically, this means that the deep learning model should not be sensitive to color, distribution, vignetting, and lens distortion [4].

The retrieval process consists of indexing the image present inside the database and then match the questioned image with the image in the database to extract the needed image. The search engine will take a feed-in the questioned image by searching the image database. Once selecting a resemblance, the application picks an active algorithm that accepts it. According to these definitions, the effectiveness of the system is assessed[5]. Users used Simple baseline methods.

So that user can state an object frame or numeric keys using a straight-line drawing method. If the user has data in fragments, then the split questioned photograph is utilized to obtain the same item. [6].

When an image retrieval system has created, The first step was to identify the database. The database indexing procedure includes measuring the database using an image dictionary for infusing attributes throughout the image.

The quantified image has governed by the image description. If the image inserts in the image descriptor, then attributes are obtained on the opposite side. Inside primary details, attribute simply merge numbers used to represent and measure image[1].



Figure 3.1. The pipeline of an image descriptor.

In this input image, an image identifier has used. The output can be a sequence of elements accustomed to measure the image[1].

Correlation of feature vectors can take place to similarities adopting distance metrics called matching function. The inputs given to Distance metrics and similarity functions are two feature vectors. They show the variety that means whence "similar" two attribute vectors are[1].

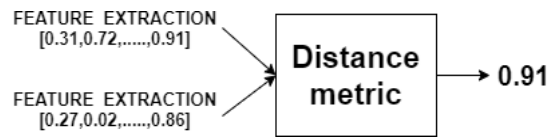


Figure 3.2. Correlation of vectors include feature vectors into the distance metric function.

The result of space function may be an individual floating-point amount represents the alikeness of these two images. The reverse image search method uses an image to find relevant images and details. When an image over cyberspace. Relevant information about the similar to the uploaded image in question from the image database. This online photo viewing tool also works with CBIR technology to find the same resemblance without any hindrances in the requested image [4].

Details found while performing reverse image searches include similar images, the same pictures with different sizes, and a detailed list containing that image.

Since cyberspace has so many images, it is difficult to find a specific picture, its source, and other relevant information from that huge number. So, at the same time, image search helps in finding the image from cyberspace. This tool returns images from six search engines, Google, Bing, Yandex, TinEye, Sogou, and Baidu, and gives all the matched resemblance in a few seconds[6].

There is no limit to having a specific device or smartphone to use this tool. It is a web-based tool that helps you search for an image on the device, any time.

Some classifiers used for image retrieval application:

### 3.1.1. Maximally Stable Extremal Regions

MSER uses very little memory and has a well-functioning housing that leads to faster use. The MSER algorithmic code is a regional detector first used in a wide-angle stereo that intervenes in the image of direct installation without slipping, which comes between the acquisition of each fine and bright frame. The algorithmic law keeps track of the pixel parts used by the union acquisition organization. The algorithmic rule that runs on individual pixel components results in less memory usage and quicker forcing more image elements across multiple grey-levels. Bulk can be aggregated and used as linked lists. The total value of the in-stack is that the digit of pixels. The grey levels are only one copy of the pixel that can found in the pixel boundary pixels. The pixel can only return to the pile as often as it has a neighbour. First, it takes log (m) to insert a pixel. The distress time to execute is bound[10]( Matas et al.)

$$O((n + e) \log(m)) \quad (1)$$

Where "m" stands for the digit of grey levels. "n" stands for the count of pixels, and e for edges in the image graph [10].



To understand MSER more clearly, create a vacant pixel lattice by eliminating all pixels for an image and then pixels are begin to install in their intensity order, for example in the beginning all the dark pixels are placed at their right areas until the total image is reestablished. Throughout this procedure more an more pixels are associated with other regions. A MSER is discovered when size of the dark areas is equivalent than in earlier intensity image.

MSER algorithm is basically split up into 4 main parts:

- **PREPROCESSING:** In this part for every intensity level, number of pixels are determined and are arranged in intensity order. A vector with pixel positions were needed as a base information which are arranged by intensity level. The sorting algorithm is used to sort intensity level is bin sort and radix sort.
- **Clustering:** In this part clusters at every intensity level are represented. Union- find algorithm is used to track the clusters of connected pixels. This algorithm utilizes memory, which has multiple positions as a query image, where each location resembles a pixel state in the same are in the earliest image.
- **MSER detection:** Size,  $|P|$ , of all clusters were followed and development rate,  $p$ , is assessed for local minimum which is given by size of region next intensity level minus, region size in previous intensity level divided by size of region in current intensity level. And if the result is in local minima then it means it is in MSER.
- **Display Result:** When a MSER is distinguished, all pixels that are essential for that district must be recognized and introduced as an output.

#### *Scale-Invariant Feature Transform*

SIFT may be a match of image scale and rotation want to observe and outline the native option in the digital images. It identifies bound points and provides details of measurements (called adjectives) will be for visual perception (D.Lowe, University of British Columbia, 2004).

#### *Speeded-Up Robust Features*

As a reputation suggestion, it is a speeded-up version of SIFT. The most interesting of the SURF approach lies in its quick computation of operators victimization box filters. Facultative period applications like tracking and visual perception (Bay, H., Tuytelaars, T., and Van Gool, 2006)[22].

#### *Principal component analysis*

PCA can also be way to deliver patterns during the database by converting variables. Its accustomed clean information sets to form it simple to explore and analyse. It also reduces the dimensionality of large data sets by transforming an outsized set of variables into a smaller one. Most of the knowledge within the large sets [25].

#### *Linear Discriminant Analysis*

LDA of a size-reduction process commonly used for surveillance problems. Used to model differences in groups like separating two or more classes. Used to print features in high-resolution spaces to low-level space.

#### *Flutter Framework*

Flutter is an open-source framework for making apps that work best for both android and iOS. It uses one codebase for the applications for mobile, web, and desktop. Essential features of flutter contain AOT compilation and just-in-time compilation. The hot reload feature allows us to experiment, add features, build UI, fix bugs, and present them. It injects the uploaded ASCII text file into the DVM (Dart Virtual Machine). The framework redesigned the widget tree allowing it quickly to view results [19]. It has an upscale number of widgets. There are two kinds of widgets countless widgets and state widgets are in flutter [20].

Dart is used to writing all apps in a flutter. Google created dart language to do programming & it is also conserved by google. Views in all-new frames with flutter are updated. This behaviour leads back to the fact that tons of objects which can hold less than one formed. Dart is designed to address this situation at the memory level with the aid of the "Generational Garbage Collection" [1].

## IV. PROBLEM FORMULATION

### A. GAPS IDENTIFIED

In the present study, we found that image-based retrieval mainly relies solely on the feature-releasing feature. Especially in a color that does not contribute to the positive results of your queries. We also found that in terms of efficiency, repeatability, and localization, MSER is the best algorithm as however, SIFT is best if robustness is the measurement criteria.

Table 4.1. Comparison of various classifiers used in image retrieval applications (rating out of 3 stars).





METHOD	SCALE	TIME	BLUR	AFFINE	ROTATION	ILLUMINATION
SIFT	***	*	*	**	***	*
MSER	***	***	*	***	**	**
PCA	**	**	***	***	**	**
LDA	*	**	**	***	*	***
SURF	*	***	**	**	*	***

Sr.no.	Parameter	React native	Flutter
1.	Released year	2015	2017
2.	Language	JavaScript	Dart
3.	Released by	Facebook	Google
4.	Learning curve	Easy for React and JavaScript developers	Need to read Dart first, so it can be a bit difficult at first
5.	Documentation and tools	Documentation is good but it can be challenging to start a project from scratch	Great documentation and starter toolkit
6.	Hot reloads	Yes	yes
7.	Libraries	Uses the native UI components of the OS	Rich library of widgets and plugins
8.	Architecture	Contingent on Flux and Redux, you need a bridge to connect to the native UI	Built-in BLoC, the UI is built with pre-Flutter features

*Table 4.2. Comparison between flutter and react-native*

A descriptive definition in mobile applications refers to applications designed to run on a specific platform or OS. Dart is a language that is comprehensive than any other programming language and was immediately used by developers. JavaScript bridges are used by React Native. React Native provides a way to connect with native modules. To communicate between two parties JSON messages need a smooth user interface. Whereas, Most of the necessities contained within Flutter. These rules eliminate the need for a bridge. Applications build in Flutter are stable. Better CLI support provides Flutter. Flutter also provides a convenient framework. Flutter integration time is short compared to React native[19].

**B. PROBLEM STATEMENT**

The aim of this work is to study basic framework and technique require towards the implementation of image

based search engine in e-commerce.

To Study the role of the content based image retrieval system that how we implement, what is the use of content-based image in real world, and what changes it can make in the world.

Develop a better e-commerce platform, where anyone become a seller without investing any money. We thought to provide a platform to the local vendors who is suppressed by the big online marketing sites like amazon, flipkart, and who is also get affected by demonetization and good and service tax. So, in our application anyone can become a seller and buyer can buy product directly from the shop. In this way seller doesn't need to invest any money to join online platform for selling.

**C. SEARCHING TECHNIQUES**

Image meta search is a type of retrieval system expert on searching animations, images, pictures etc. Like the text

search, image search is an information retrieval system designed to help to find information on the Internet and it allows the user to look for images etc. using keywords or search phrases and to receive a set of thumbnail images, sorted by relevancy.

There are three types of image search engines:

- Search by Meta-data
- Search by Example
- Hybrid Approach

*Search by Meta-data :*

Searching by meta-data is merely marginally totally different than your customary keyword-based search engines mentioned on top of. Search by meta- data systems seldom examine the contents of the image itself. Instead, they trust matter clues like (1) manual annotations and tagging performed by humans together with (2) machine-controlled discourse hints, like the text that seems close to the image on a webpage.

*Search by Example :*

Search by example systems, on the opposite hand, believe alone on the contents of the image — no keywords area unit assumed to be provided. The image is analyzed, quantified, and hold on in order that similar pictures area unit came back by the system throughout a look

*Hybrid Approach :*

A hybrid approach would be to correlate the options extracted from the image with the text. a picture program that might take each discourse hints along side a groundwork by Example strategy.

**D. SYSTEM OVERVIEW**

Content-based image retrieval is a mechanism that can resolve disparities in visual interpretation that can mislead the process of retrieval since it is focussed on the visual examination of content that is essential for the image of the inquiry and also involves four fundamental steps:

- Define Image description
- Identifies your dataset
- Defining your similarity metrics
- Searching

*Define Image Description*

In the first step, the aspect of the image to be defined is chosen.

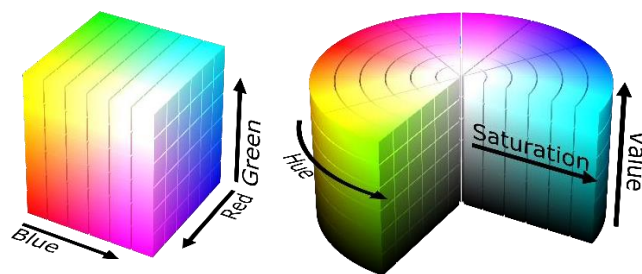
- The color

- The shape
- The texture of the image.

*Color Features*

RGB color system is most commonly used color system for image retrieval. Unfortunately, the RGB is not well suitable for describing colors in terms that are practical for human interpretation. Whereas, the HSV (hue, saturation, value) model is an ideal tool for developing image processing algorithms based on color descriptions that are natural and intuitive to humans [7]. For image retrieval, the image normally converts from RGB space to HSV color space.

*Fig. 4.1. (a) and (b) shows the RGB and HSV color space respectively.*



Harsh feeling of the density of pixel intensity are utilized by histograms.

Basically, histogram will appraise the likelihood destiny.

By experiment we came to know that if a histogram has less bins then it will have smaller features and can distinguish between images with a very unique color distribution. On the other if a histogram has more bins then it will lose the capacity to “generalize” images with the same corresponding coverage because one of the tops & ridges of the histogram should be identical so that the two images are considered "identical".

So, by experimental approach we came to know that number of bins for histogram descriptor depends upon The size of dataset Similarity of the color distributions in the dataset.

Now, Rather than inserting a 3D HSV histogram for the whole image, calculating the 3D HSV color histogram for different image regions would be a better option.

Utilizing regions-based histograms in compare to global-histograms permits us to mimic location in a color distribution below two images shows an entire

image and image after region distribution.

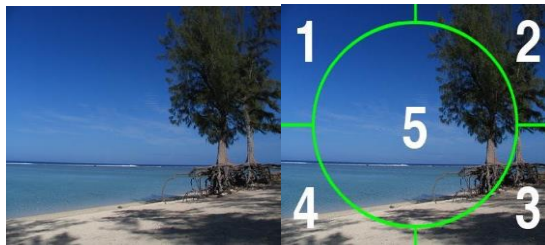


Fig. 4.2. (a) is an example of global image and (b) is an example of region based image.

By utilizing a global histogram in fig. 4.2.(a) we would simply realize that some levels of brown and blue are exists there. So, better approach is to use these regions as shown In Fig. 4.2.(b), we'll have the option to mirror a rough type of

localization, having the option to address above sea shore picture as in the upper right and upper left corners, there are some shades of sky and earthy colored land in the base right and base left corners, and afterward a blend of sky and earthy colored land in the middle area.

After considering these factors and perform feature It is vital to standardize shading histograms. Every histogram is addressed by a % calculation for particular bin & not a numerical count for every bin.

#### Texture Feature

##### Block Difference of Inverse Probabilities

Local intensity minima territories are addressed by valleys and edges are address the regions which include unexpected difference in intensity. In human vision these are the vital highlights and, particularly, 355 valleys are essential thing in the shape of an object in terms of visual impression [27, 28]. [26] is a surface highlights that successfully extricates valley and edges. Block difference of opposite probabilities, which is one of the proposed surface highlights, is characterized as the contrast b/w the quantities of pixels in a block [25].

$$BDIP = M^2 - \frac{\sum_{(i,j) \in B} I(i,j)}{MAX_{(i,j) \in B} I(i,j)} \quad (2)$$

When B defines a block size M x M. There is an increase in intensities when in a block, the BDIP value increases. Where I(i, j) signifies the worth at a pixel (i, j) in the image I [25].

##### Block-Based Local Correlation

BVLC [27] is used to measure texture smoothness well and it represents the variation of block-based local correlation coefficients according to four orientations. Each local correlation coefficient is defined as local covariance normalized by local variance.

$$\frac{1}{M^2} \sum_{(k,l) \in 04} I(i,j)I(i+k,j+l) - \mu_{0,0} \mu_{k,l} \quad (3)$$

$$\sigma_{0,0} \sigma_{k,l}$$

Where  $\mu_{0,0}$  and  $\sigma_{0,0}$  represent the local mean value and standard deviation of the block with size MxM. The (k,l) term denotes four orientations (90°, 0°, 45°, -45°). As a result,  $\mu_{k,l}$  and  $\sigma_{k,l}$  represent the mean value and standard deviation of the shifted block, respectively. The larger BVLC value indicates that the ingredients in the block are rough [27].

##### Shape Feature

Shape is an impressive element utilized for picture grouping, ordering and recoveries. Shape highlight will be addressed by region or edge based. The direction, position and size of an objects doesn't effect the numerical details of the image. Through the canny edge operator [27], an edge histogram is produced, to provide it as a shape feature for the continuation of the search engine [25]. The canny edge detection algorithm works in 5 different steps.

- Smoothing. Obscuring to eliminate commotion.
- Find gradients. Edges ought to be checked where the gradient of the pictures are huge in size.
- Non-maximum suppression. Just edges and neighborhood maxima ought to be set apart.
- Double threshold. Thresholds control the potential; edges.
- Edge track by hysteresis. Last edges are determined by stifling all edges which are not associated with a specific edge.[25]

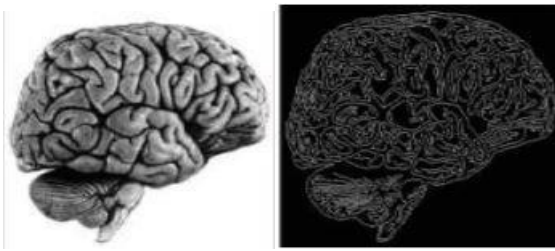


Fig. 4.3. (a) Normal Image and (b) Edge Detected Image

- i. Identifies your Dataset

In the previous step, the descriptor of the image is defined, and now it is applied to each image present in the database, extract elements, and write and save those features for later comparison.

#### *Instantiate Your Descriptor*

Since it specifies that the exact same descriptor with the exact same parameters is applied to each and every image in my dataset, I used a class instead of a function. This is particularly useful if I need to write my descriptor to disc and then load it back up later, such as when a user performs a new operation. To compare two images, you must use your image descriptor to represent them in the same way. If you want to compare two images for similarity, it's not a good idea to extract a histogram with 32 bins from one and a histogram with 128 bins from the other.

#### *Feature Extraction*

Feature extraction may be part of a size reduction process. The starting set is split and bring them into manageable groups. Feature detection helps to appeal to the simplest feature in those big data sets by selecting and mixing variables in features [1].

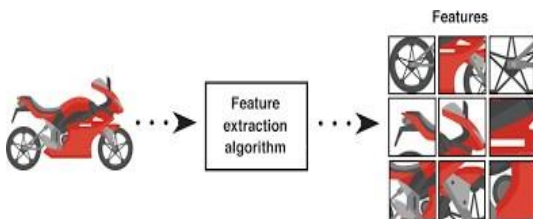


Fig. 4.4. Example of Feature Extraction

#### Defining your Similarity Metric

In this step, you would like the way to match those extracted features equally. Popular options include Euclidean distance, the sum of squared absolute difference, and Canberra distance. but the alternative depends largely on the database and

features extracted[1].

#### *Sum of squared absolute Difference (SSAD)*

The sum of square deviation of the total vector estimation of the 2 elements is calculated. It can be derived (Selvarajah and Kodi) Squaring, however, still returns a positive worth, yet a significant gap is highlighted. The SSAD distance metric can be utilized for the pixels and changed areas, yet the measured value in the transformed domain relies on the compression quality. (yuwakku, 2011) as:

$$\Delta d = \sum_{i=1}^n (|Q_i| - |D_i|)^2 \quad (4)$$

In contrast to SAD, it has some computational complexity due to a square of differences. Squaring, though, still gives a positive value, yet a significant gap is highlighted. The SSAD distance metric can be utilised for pixels as well as transformed domains, yet the measured value in the transformed domain relies on the compression quality [25].

#### *Euclidean distance*

It is often used to measure similarity in image extraction due to its salient features. Measure the space between two image vectors by deriving the total root of the total square difference and it are often derived (Szabolcs, 2008) as:

$$\Delta d = \sqrt{\sum_{i=1}^n (|Q_i| - |D_i|)^2} \quad (5)$$

#### *Canberra distance*

Canberra distance is the distance between vectors of database feature and query. It also does not give a large number of identical images. The Canberra distance can be derived (Szabolcs, 2008) as:

$$\Delta d = \sum_{i=1}^n \frac{|Q_i - D_i|}{|Q_i| + |D_i|} \quad (6)$$

#### Searching

The end-user will send an image to our search engine and our work is to extract the features and create an index and use your image index to match the index of the query image. then, it directly returns to the most suitable results in terms of our matching work(Ruchi Jayaswal, 2018)[22].

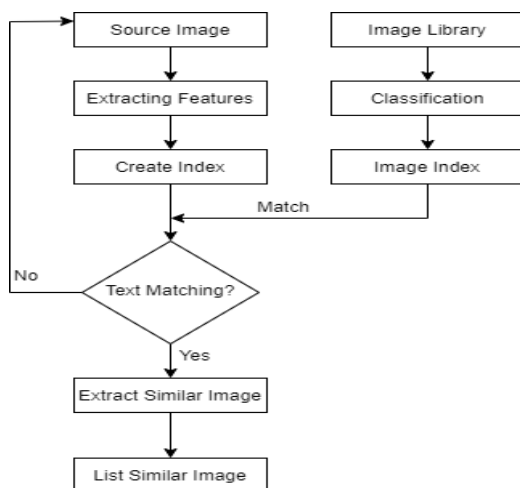


Fig. 4.5. Flow chart of Searching

## V. MATERIAL AND METHODS

### A. EVALUATION CRITERIA

To test the performance of our proposed model we calculate the realities correct, false, inaccurate, precisely generated by a system or algorithm then use this with multiple scales to evaluate and compare performance of various programs.

The transactional volume of fraudulent transactions identified as systematic fraud is known as True Positive. False Positive transactions are ones that were legal but were mistakenly classified as fraud. The various metrics for evaluation are:

Accuracy is part of a well-defined task. It is one of the most powerful and widely used test metrics. It basically tells us how many predictions were correct in all input transaction.

$$\text{Accuracy (ACC)/Detection rate} = (TN + TP) / (TP + FP + FN + TN) \quad (7)$$

Precision, usually as the identification scale, refers to a number of activities that are clearly identified as true or false. Indicates how many key results were returned and how many missing results were returned.

$$\text{Detection rate/Hit rate} = TP / (TP + FP) \quad (8)$$

Sensitivity, refers to the percentage of irregular records (those with the greatest risk of being fraudulent) that the method correctly classifies. It essentially tests how sensitive our system is by counting how many fraudulent transactions were correctly defined as such.

$$\text{True positive rate/Sensitivity} = TP / (TP + FN) \quad (9)$$

Specificity Specification measures a fraction of the standard records (records with the least chance of fraud) that are well categorized by system. It measures how many legal actions are classified as actual.

$$\text{True negative rate /Specificity} = TN / (TN+FP)$$

False Alarm rate False Alarm rating comes from complete categories that are categorized as fraud how many are incorrectly classified. It should be noted that the false alarm rate should be minimal otherwise it damages the spirit of the actual credit card users.

$$\text{False Alarm Rate} = FP / (FP+TN) \quad (11)$$

### B. TOOLS AND TECHNIQUES

The following are the methodologies, APIs, and libraries used in the implementation of the proposed system:

Techniques: The most common search terms in this area are search by meta data, for example. Among them the hybrid method does the best. As a consequence, the techniques used in the proposed hybrid model will be defined.

Languages and Libraries: Python, Java, Kotlin, dart or R language are some of the options that could be used. We modified our system using Python Language Recovery Image based on content and dart of the e-commerce platform.

Python: It is an open source translation, a high-quality, common sense language developed by Guido van Rossum in. It is widely used in machine learning because of its fast and compact code and supports many reading libraries. We can use a variety of built-in tasks to do almost any difficult task in just a few steps.

Numpy library: stands for 'python of numbers'. Recommends a python with rich data resources using multidimensional and matrix layouts. This module provides a large library of high-level mathematical operations to work on these matrices and arrays.

CV2 library: The OpenCV Python library is a collection of Python bindings for solving computer vision problems. Numpy is used by OpenCV-Python., which is a highly optimized library for numerical operations with a MATLAB-style syntax. All the OpenCV array structures are converted to and from Numpy arrays.

Imutils: Imutils are a series of simple tasks for basic image processing tasks such as translating, rotating, enlarging bone size, and easily displaying Matplotlib images with OpenCV

Argparse: The argparse module renders creating a simple command line interface an ease. The programme determines the arguments it needs, and argparse deduces how to extract them from sys.argv. When users send the software invalid arguments, the argparse module automatically produces support and usage messages and issues errors.

Glob: The glob module in Python is used to find files/pathnames that fit a template. The glob pattern rules are the same as the Unix route expansion rules. It is also expected that, based on benchmarks, it can fit pathnames in directories faster than other approaches.

CSV: Classes for reading and writing CSV tabular data are provided by the csv module. It enables programmers

to say things like "write this data in the format preferred by Excel" or "read data from this file provided by Excel" without having to know the specifics of Excel's CSV format. Programmers may also identify their own specialpurpose CSV formats that are understood by other applications. Programmers may also define their own specialpurpose CSV formats or identify the CSV formats that other applications understand.

## VI. EXPERIMENTAL RESULTS

This section deals with performance assessment results, including image specific and test measures, in order to determine the efficacy of the redundancy programme.

### A. EVALUATION MATRICES

For retrieval efficiency we have consider two parameters namely recall and precision. Precision can be defined by the ratio of number of relevant images retrieved to total number of images retrieved. Recall can be defined by number of relevant image retrieved to total number images in the database. We calculated recall and precision value in both case output after applying the HSV color Histogram for color feature extraction. For the similarity measurement we have used the chi squared distance Metrics. In our experiment, the precision and recall are calculated as[30]:

$$\text{Precision} = \frac{\text{Number of relevant images retrieved}}{\text{Total number of images retrieved}} \quad (12)$$

$$\text{Recall} = \frac{\text{number of relevant images retrieved}}{\text{Total number of relevant images in the database}} \quad (13)$$

### B. IMAGE DATABASE

The image dataset is downloaded from the <https://www.kaggle.com/mbkinaci/fruit-images-for-object-detection>. This image database consists of 240 images with 4 object categories in which 3 object categories contain 76 images and the remaining one contain 20 images. The cluster names of these images are : Apple, Banana, Orange and Mixed. Fig shows some of the sampled images from the dataset[30].



*Fig. 6.1. Some Sampled images from data set*

### C. RETRIEVAL RESULTS

On the picture website mentioned above, we used the proposed process.[30] Table 6.1 shows the results of the tests based on median accuracy and memory using the proposed form.

TABLE 6.1. Average Precision and Recall

	APPLE	ORANGE	BANANA	MIXED
PRECISION	72.727	89.830	79.661	88.888
RECALL	52.631	69.736	61.842	80.000

## VII. CONCLUSION AND FUTURE SCOPE

This chapter summarises the research project's major accomplishments. It addresses potential research that can be undertaken in order to reach the ultimate objective of developing a powerful and cost-effective software based on Content-based Image retrieval system.

### A. OVERALL CONCLUSION

The purpose of this paper is to study the basic framework and technique requirements for the implementation of an image-based search engine. The paper also identified gaps and existing studies. We reviewed many algorithms and ways to implement content-based image retrieval. Some use the MSER algorithm and some a few scale-invariant feature transform to obtain features of the image. These various



algorithm has some innumerable merits such as MSER performs better in terms of scaling and time complexity whereas PCA performs better when the image is blur or in preserving line and parallelism in image i.e. affine and SURF performs better in terms of time complexity and illumination and LDA performs better in terms of affine and illumination. As everything has its price. So, these algorithms had its own limitations too. As MSER does not give appropriate results when the image is blur and PCA gives some bad result in terms of time complexity, scale and illumination whereas LDA gives undesirable output when image is rotated. So, considering there merits and demerits we come to the conclusion that most of the algorithm detects images by the colors only which could not give the best results as different objects in different images may have same colors histogram. So, we figure it out if we manage to obtain the color, shape, and pattern of an image. We could get the best results because no different object in different images will match all these features. We also study a framework like React and flutter. To build a hybrid application but they compromise on performance and developer productivity. These hybrid approach will remove the burden of learning different languages for different platforms like objective C and Swift for IOS; Java and Kotlin for android by using these framework developers only required to learn single language for multiple platforms. These frameworks done well in terms of their use and its very hard to select one from them. There is only one point which makes flutter better than React that it is easy to build from scratch. We also identified loopholes in existing studies and proposed a framework that fills them in while also providing an accurate CBIR system with a mobile application.

## B. FUTURE SCOPE

For Future Work we intended to presented a novel approach in content-based image retrieval (CBIR) by combining the low level feature i.e. color, texture and shape features. Firstly, RGB color model is transformed to HSV color model and then color histogram is extracted to form color feature vector. Next, for extracting the texture feature, Block Difference of Inverse Probabilities (BDIP) and Block-Based Local Correlation (BVLC) moment are used. At last, we applied Canny edge detection to extract the shape features. Finally, all three features are combined to form the feature vectors of the entire image. Instead of just using Colour feature extraction.

## VI. REFERENCES

- [1] Latif, A. Rasheed, U. Sajid, J. Ahmed, N. Ali, N. I. Ratyal, B. Zafar, S. H. Dar, M. Sajid and
- [2] T. Khali, "Content-Based Image Retrieval and Feature Extraction: A Comprehensive Review,"
- [3] Hindawi, *Mathematical Problems in Engineering*, p. 21, 2019.
- [4] M. M, A. Kumar and P. M, "Cross-platform development using flutter," *International Journal of Engineering Science and Computing*, vol. 9, no. 4, pp. 21497-21500, 2019.
- [5] R. Wu, S. Yan, Y. Shan, Q. Dang and G. Sun, "Deep Image: Scaling up Image".
- [6] P. M. Chutel and A. Sakhare, "Evaluation of compact composite descriptor based reverse image search," *International Conference on Communication and Signal Processing, India*, pp. 1430- 1434, 2014.
- [7] D. Nist'er and H. Stew'enius, "Linear Time Maximally Stable Extremal Regions," *European Conference on Computer Vision*, pp. 183-196, 2008.
- [8] W. Wu, "React Native vs Flutter, cross-platform mobile application frameworks," *Metropolia University of Applied Sciences Thesis*, 2008.
- [9] G. A. Papakostas, D. E. Koulouriotis and V. D. Tourassis, "Feature extraction based on wavelet moments and moment invariants in machine vision systems," *intechopen*, 2012.
- [10] G.-H. Liu, Z.-Y. Li, L. Zhang and Y. Xu, "Image retrieval based on micro-structure descriptor," *Elsevier Ltd*, p. 2123–2133, 2011.
- [11] X. Wang, Z. Chen and j. Yun, "An effective method for color image retrieval based on texture," *Computer Standards & Interfaces* , vol. 34, no. 1, pp. 31-35, 2012.
- [12] R. Ashraf, K. Bashir, A. Irtaza and M. T. Mahmood, "Content-based image retrieval using embedded neural networks with bandletized regions," *entropy journal publication* , vol. 7, no. 6, pp. 3552-3580, 2015.
- [13] A. Irtaza and . M. A. Jaffar, "Categorical image retrieval through genetically optimized support vector machines (GOSVM) and hybrid texture features," *Springer-Verlag London*, 2014.
- [14] S. Fadaei, R. Amirfatahi and . M. R. Ahmadzadeh, "Local derivative radial patterns: a new texture descriptor for content-based image retrieval," *Signal Processing*, 2017.
- [15] W. Xingyuan and W. Zongyu, "A novel method for image retrieval based on structure elements' descriptor," *Journal of Visual Communication and Image Representation*, vol. 24, no. 1, pp. 63- 74, 2013.
- [16] D. Zhang and G. Lu, "Review of shape representation and description techniques," *Pattern Recognition*, vol. 37, no. 1, pp. 1-19, 2004.
- [17] F. Malik and B. Baharudin, "Analysis of distance metrics in content-based image retrieval using statistical quantized histogram texture features in the DCT domain," *Journal of King Saud University Computer and Information Sciences*, vol. 25, no. 2, pp. 207-218, 2013.
- [18] R. Jayaswal and J. Jha, "A Hybrid Approach for Image Retrieval Using Visual Descriptors," *International Conference on Computing, Communication and Automation*, pp. 1125-1130, 2017.
- [19] J. Cruz-Mota, I. Bogdanova, B. Paquier, M. Bierlaire and J.-P. Thiran, "Scale Invariant Feature Transform on the Sphere: Theory and Applications,"



International Journal of Computer Vision, vol. 98, pp. 217-241, 2012.

[20] F. Kristensen and W. J. MacLean, "Real-Time Extraction of Maximally Stable Extremal Regions on an FPGA," IEEE International, pp. 165-168, 2007.

[21] M. Olsson, "A Comparison of Performance and Looks Between Flutter and Native Applications," Department of Software Engineering at Blekinge Institute of Technology, 2020.

[22] T. Yeh, K. Grauman, K. Tollmar and T. Darrell, "A Image is Worth a Thousand Keywords,"

[23] Extended Abstracts Proceedings of the 2005 Conference on Human Factors in Computing Systems, CHI 2005, Portland, Oregon, USA, 2005.