CONTENT-BASED IMAGE RETRIEVAL SYSTEM WITH RELEVANCE FEEDBACK

by Sidhant Bansal

Sreja Babu

CSS 584: Multimedia Database Systems

Min Chen

University of Washington Bothell

Bothell, WA

02/20/2022

Table of Contents

I. INTRODUCTION

The goal of this project is to create a system that retrieves images based on various properties of the images and user feedback. This application will allow the users to retrieve images similar to a query/selected image according to their color codes or their intensities or a combination of both, on top of that it will also allow users to provide feedback which will help in improving the retrieval accuracy.

II. SETUP INSTRUCTIONS

To run the program, please follow one of the three steps given below:

1. Extract the zip file which contains the report, source code folder named “src” and executable jar file named “CBIR” in a folder. Double-click on the “CBIR” jar file to execute the program.

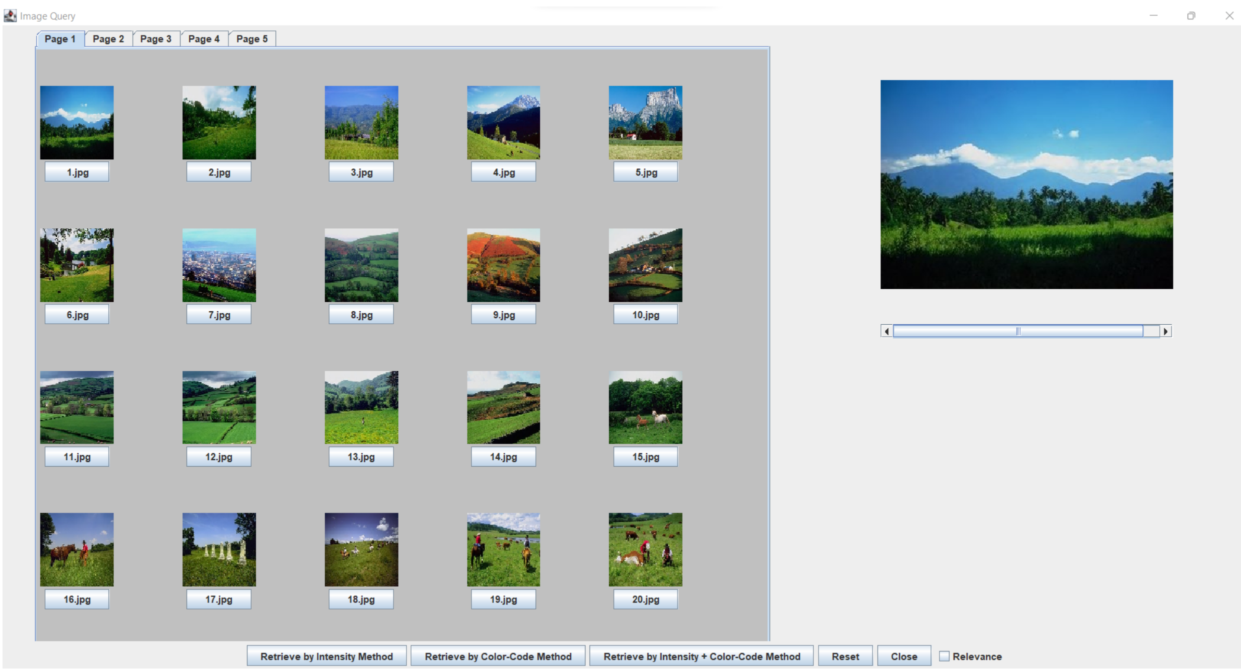
The “src” folder containing the source code should be present in the same directory for CBIR to execute. For a better experience please adjust the JFrame screen according to the machine’s screen size. The image titles/names and relevant checkboxes of each image are clickable, not the images themselves.

2. Execute it from the command prompt: Open it inside the current folder containing the src and jar file and then type “java -jar CBIR.jar”.
3. In Eclipse IDE, the project can be opened by clicking on “File”, then clicking on “Open Projects from File System”, then clicking on “Directory” and select the folder where the files are extracted. Once opened, the “Main” class can be run as a Java Application.

III. EXECUTION OUTPUTS

The whole user interface was developed from scratch in java with the help of java swings for the front end of the application, java for the backend logic.

Images were selected without any sorting/retrieval and with and without relevant checkboxes checked.



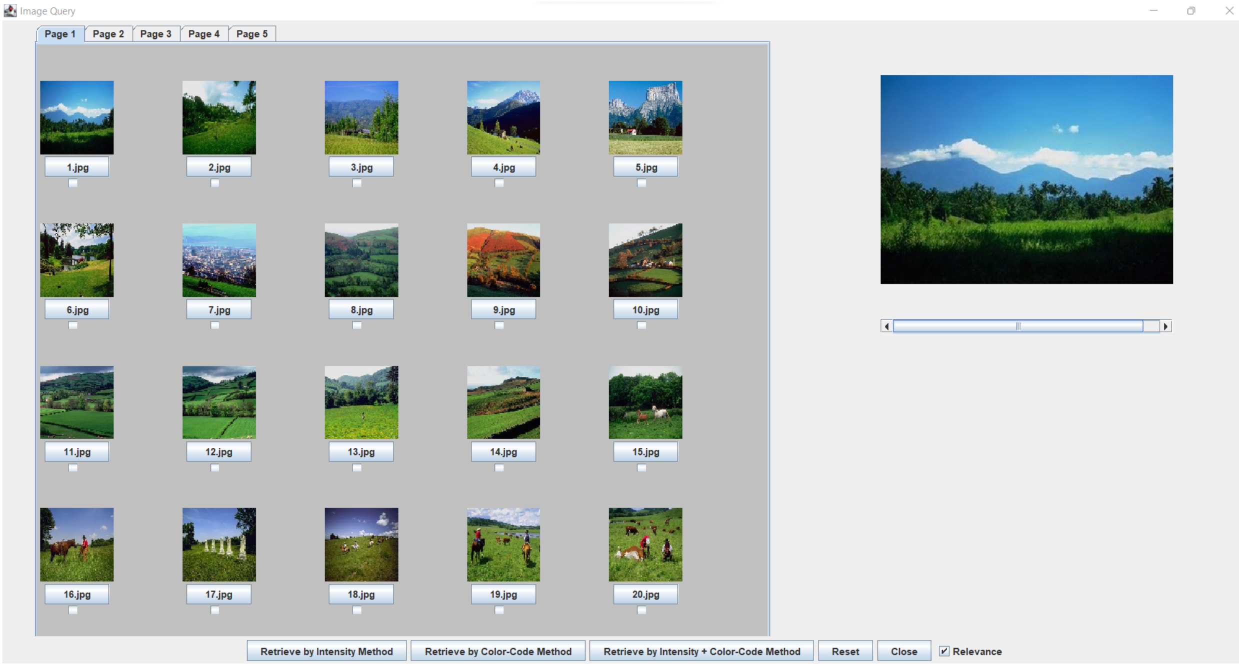
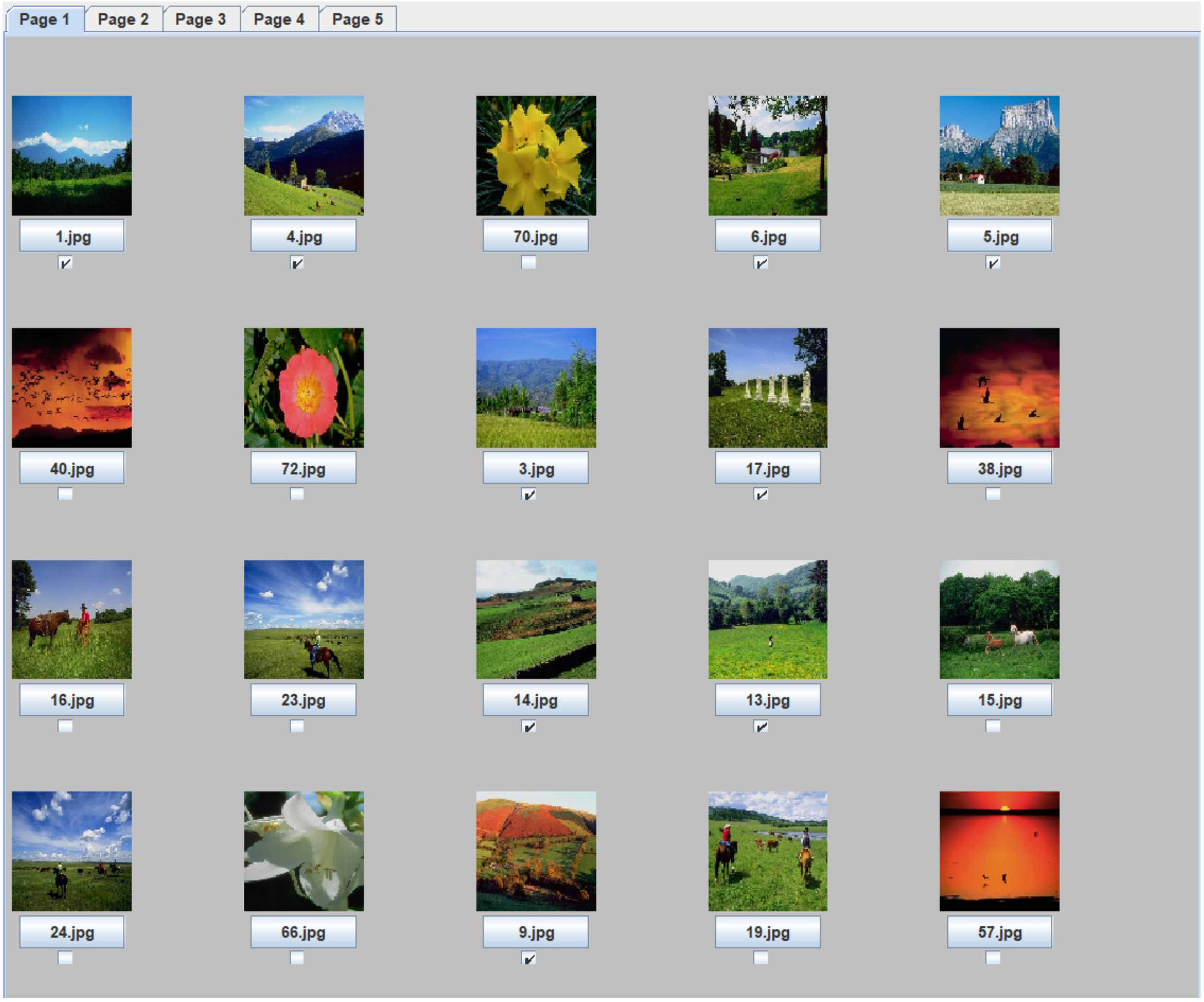


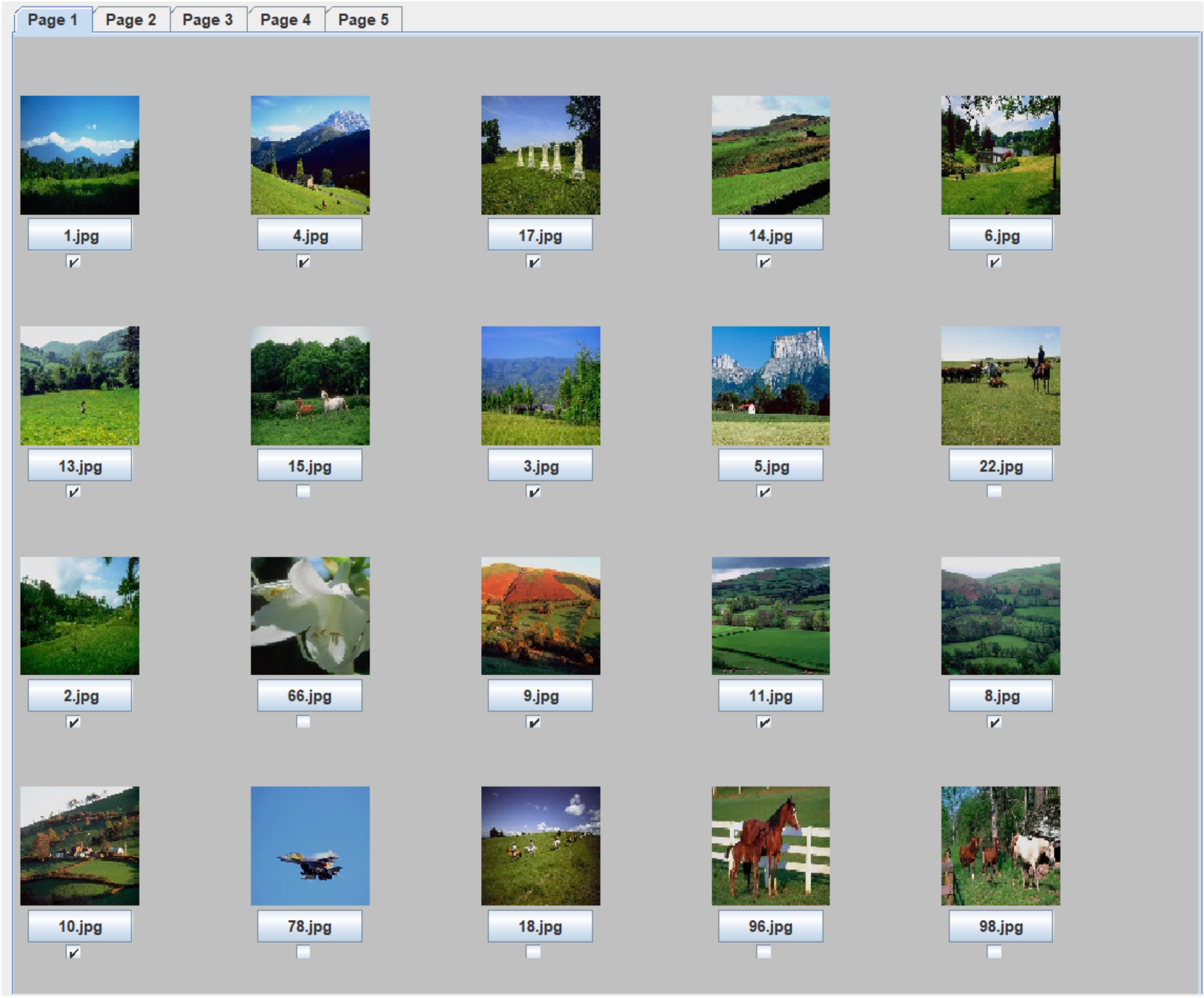
Image 1 query using intensity + color code + RF

Initial result (before doing RF, with the relevant image selected which will be used for RF analysis):



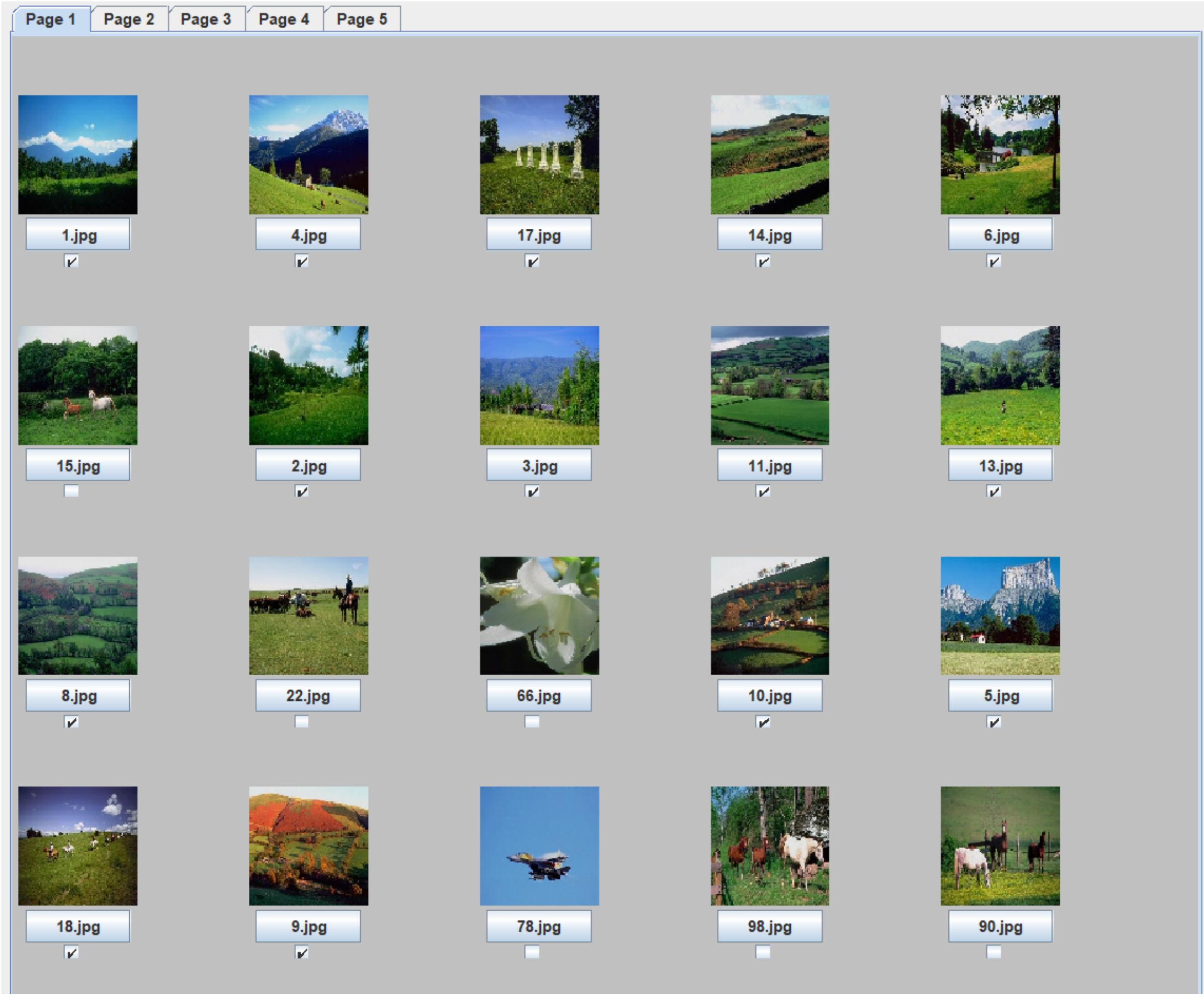
Precision: 9/20

1st iteration result (1st RF analysis results, with new relevant images, are selected for 2nd RF analysis):



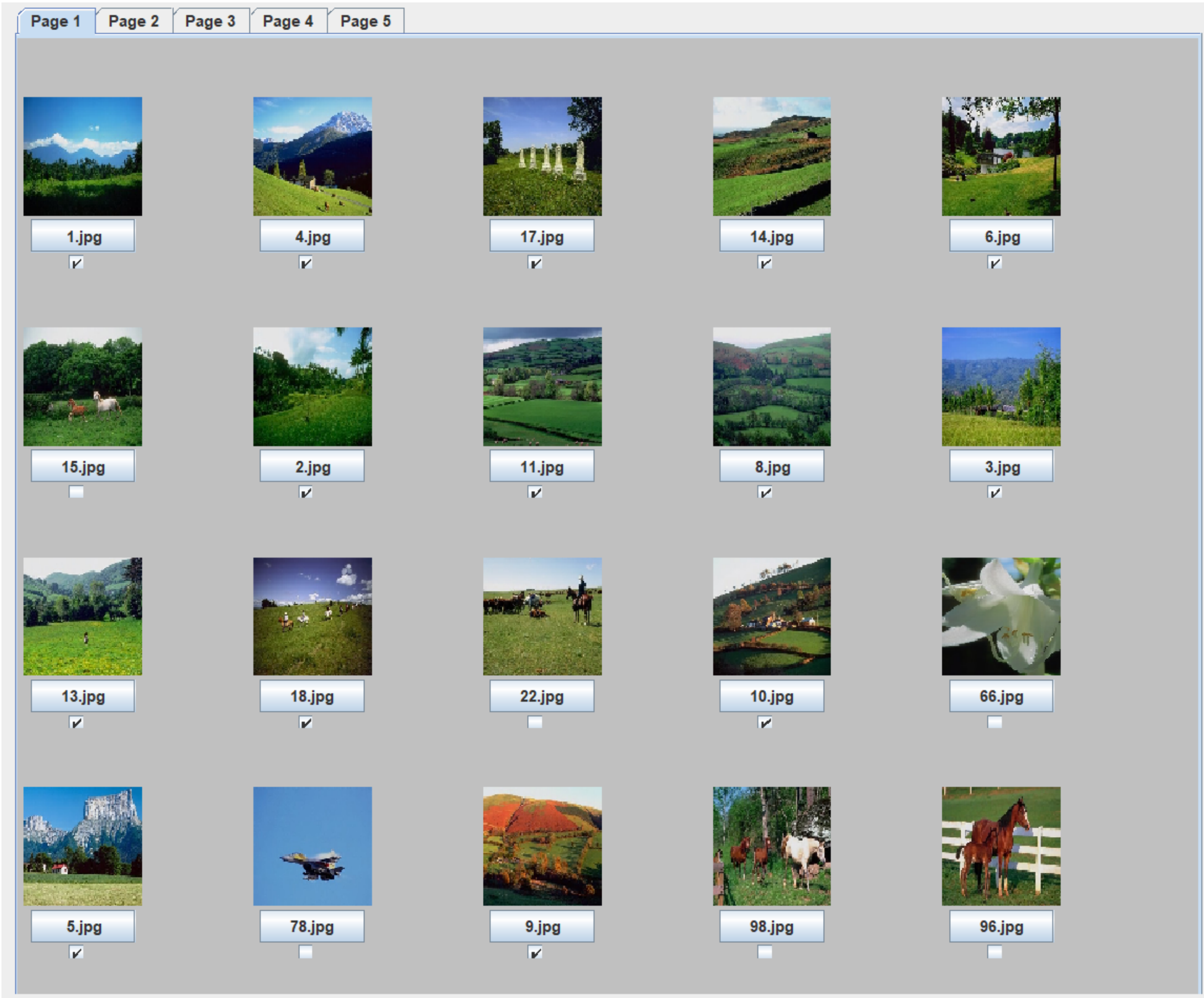
Precision: 13/20

2nd iteration result



Precision: 14/20

3rd iteration



Precision: 14/20

IV. ADVANTAGES

Developer

1. While developing this application, the design aspects of the program were convenient to visualize, which in turn helped in the smooth creation of the application. The complete application was developed from scratch.
2. A well-defined set of capabilities was defined ensuring that the purpose of the application was clear. After that, an architecture of the system was created with enough details for the

implementation of the system. Several services and components were also used in this, components included generating histograms(intensity or color-coded), and these components were wrapped in services.

3. These components were designed in such a way that they could be easily integrated into other applications similar to this one. The BufferedImage class made it easy to fetch the RGB values of a pixel of an image, which was useful because it made it easy to retrieve the correct and accurate features of a pixel of all the images in the dataset.
4. All the features of an image including intensity and color-coded histogram were stored in a custom array and were shared among the components that required this data to execute it. This ensured that the execution time of the program was less as different classes didn't have to separately calculate the features of an image again and again.

User

1. As a user, the end product seemed visually good.
2. Easy to navigate.
3. There was no ambiguity while using the application.

V. LIMITATIONS

Developer

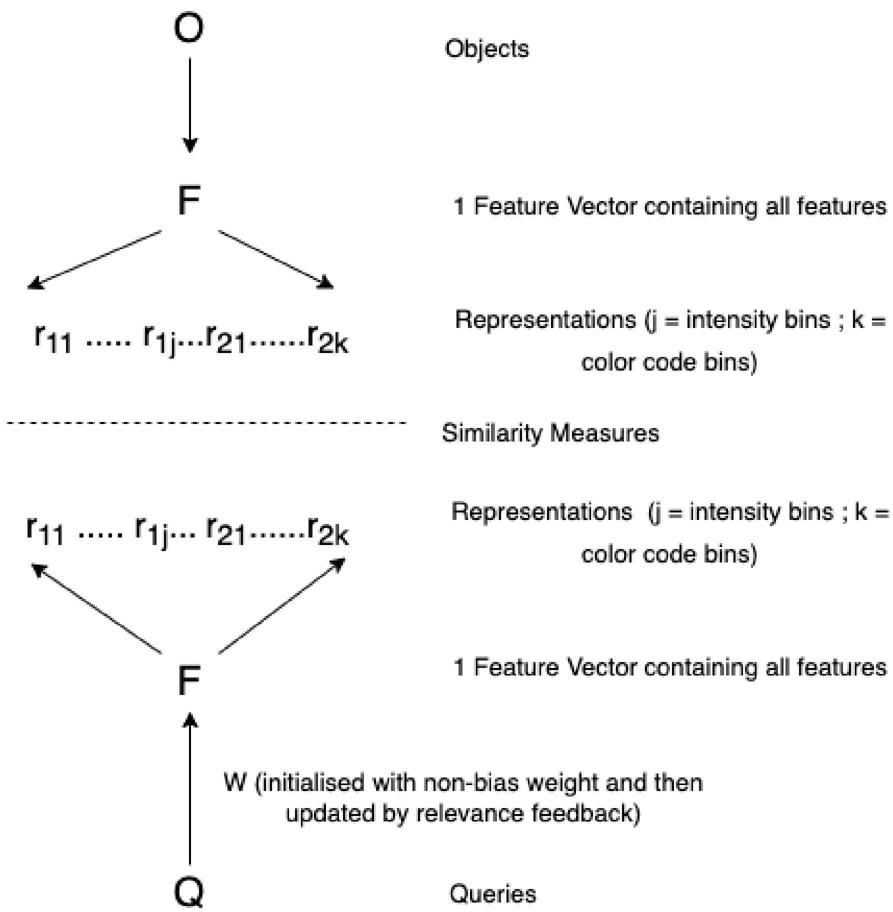
1. Faced difficulties in designing the user interface of the application.
2. The components and their placement didn't work as desired, this required intensive experimentation with different grid layouts before reaching the final design of the user interface.
3. Also, the dynamic nature of Java swing was difficult to implement which changed with each button click.

4. As all the images were displayed in a tabbed pane with the help of Arrays, the buttons didn't work normally in the initial stages of the design which was also overcome by experimentation.
5. The complete array of custom objects with image details was passed to both intensity and color-coded retrieval methods even though the intensity method didn't require the histogram for color-coded image retrieval and vice versa. This could have been avoided as we sent unnecessary information to both methods and increased the communication overhead.

User

1. The application view is perfect in full-screen mode on several screen resolutions, but the image view is not aligned with the window for some screen resolutions.
2. Some of the image retrievals didn't seem relevant to the naked eye after their respective retrieval methods.

VI. DISCUSSION



High-level Comparison	Simplified RF	RF in [Rui98]
Multi-modeling	2 features concatenated (1 Vector)	Multiple feature vectors
Interactive	✓	✓
Dynamic Feedback	✓	✓
Weight Updating	intra-weight	inter-weight and intra-weight
Normalisation	intra-normalisation	inter-normalisation and

		intra-normalisation
Removes burden from the user	✓	✓

VII. COMPARISON

The simplified RF and the RF framework proposed in [Rui98] took the two distinct characteristics of CBIR systems into consideration: 1) The gap between high-level concepts and low-level features; 2) the Subjectivity of human perception of visual content. These let the users interact with the system (interactive retrieval approach) and capture the user's high-level query and perception subjectivity by weights updated dynamically based on the user feedback. Computer-centric systems need the user to select the visual feature and specify weights for each feature. But, in simplified RF and the RF framework proposed in [Rui98], based on the user feedback (human-computer interaction), the query is adjusted for better performance.

The multimedia object model supports multiple features, representations, and weights. The RF framework proposed in [Rui98] has multiple representations such as color histogram and color moments representations for the color feature whereas the Simplified RF uses only the 2 features concatenated as 1 vector. The query has the same model as that of image objects since it is also an image object in nature. Different similarity measures are used such as Euclidean and histogram intersection for comparing vector-based representations and color histogram representations respectively.

In Simplified RF, each value in the histograms is divided by the size of the image to obtain the feature matrices of each histogram. Then, the two feature matrices are merged together to get the original merged feature matrix and the standard deviation is calculated. Finally, the Gaussian Normalisation formula is used to find the normalized feature matrix and the distance is calculated based on the dynamically calculated weights or the default no-bias weights (1/N). The RF framework proposed in [Rui98], also has a set of initial no-bias weights. The user's information need is distributed through the features and then through the representations, according to the weights. The object's similarity to the query is calculated based on the corresponding similarity measure and the weights. Each representation's similarity values are combined into the feature's similarity value and the overall similarity is obtained by combining individual similarity values. The objects in the database are ordered by their overall similarity and weight and the query is adjusted based on user feedback. The next iteration starts with the updated query.

A significant difference between the simplified RF and the RF framework proposed in [Rui98] is that in simplified RF, the user marks the images as relevant or non-relevant, but in [Rui98], the user marks most relevant, relevant, no opinion, non-relevant, or highly non-relevant for the images. Another difference is that the simplified RF uses Gaussian Normalisation (intra-normalization in [Rui98]) whereas the RF in [Rui98] uses two types of normalization: intra-normalization and inter-normalization. Intra-normalisation ensures equal emphasis of each component within a representation vector. Inter-normalisation ensures equal emphasis of each individual similarity value within the overall similarity value. In order to ensure no single similarity value overshadows others only because it has a larger magnitude, inter-normalization is applied. Queries in the RF algorithms are considered to be subjective because the weights can be dynamically updated and hence the burden of specifying weights is removed from the user.

In simplified RF, for updating weights, an intra-weight method is used where the inverse of standard calculation is used to calculate the updated weight, and the sum of updated weights is used to calculate the normalized weight by dividing the individual updated weights by the sum. In [Rui98], intra-weight and inter-weight methods are being used. In inter-weight, weight is first initialized to 0 and then the scores are added to calculate the weights (3, 1, 0, -1, -3). Similar to the intra-weight method, the calculated weights are added together, and then each weight is divided by the total weight. When there is more overlap of relevant objects, the weight seems to be larger.

VIII. LITERATURE REVIEW

Currently, most image retrieval methods rely on the comparison of metadata or textual tags associated with the images. Some of the activities related to video indexing for example tagging are mostly manual which has scalability issues for running this experiment on large-scale videos. This relies on human intervention to provide an interpretation of the image content so as to produce tags associated with the image. However, the use of Image data is growing tremendously in every field such as medicine, engineering design, fashion, interior design, education, etc. For this growing need for image data, we also need to have an effective tool for its retrieval. This led to the development of algorithms to augment and replace tag-based image retrieval with content-based image retrieval (CBIR). In CBIR, images are retrieved based on Low-level features (human vision-related), Middle-level features (object-related), and High-level features (semantic-related). This retrieval involves two phases, they are Feature extraction and Matching. There is another approach called Text-Based Image Retrieval, but CBIR is more widely used than TBIR because textual-based retrieval cannot append the perceptual significant visual features like color, shape, and texture.

These algorithms work by comparing the actual content of the images rather than text which has been manually annotated. There are a number of features that can be extracted from an image for comparisons based on their content such as video keyframes, audio sound, tagging, and auto-generated text. Once the specified feature has been extracted from the image, the images are compared [1]. Generally, the similarity between two images is based on computing Euclidean distance or histogram intersection between the respective extracted features of two images. Both these methods involve the mathematical definition of a distance between two objects [2]. The most common characteristics upon which images are compared in content-based image retrieval algorithms are color, shape, and texture.

There are several color descriptors used by the researchers for the image retrieval process such as color correlogram, color histogram, color moments, Dominant color descriptor, color coherence vectors, and color structure descriptor. The idea behind the paper by Devyani Soni; K.J. Mathai [3] is to present an effective method for image retrieval by combining color histogram and color correlogram methods. The advantage of the color correlogram is that it calculates spatial relationships which cannot be obtained by using the color histogram approach. This approach also uses Euclidean distance to compare query images with database images. According to this fusion approach, when the images are read from the database, their RGB values are extracted. A normalization histogram is created for each color channel, thus each image is having three histograms allied to it. Color Correlogram and Color Histogram are constructed based on which feature vector is created. Then, the RGB values are retrieved from the query image. For query image, normalization histogram, Color Correlogram, Color Histogram, and vector feature are constructed. Now, the feature vector of the query image is compared to the feature vector database and the images are sorted according to their distance.

Then the same steps are carried out for HSV color space. The proposed system gives the output image which is similar to the query input image. Retrieval Efficiency is obtained by calculating Recall and precision using 1000 natural color images.

The developed algorithm uses the same query image for retrieval in all four approaches and also compares the color spaces (RGB and HSV), to which color space gives better results as compared to others. The query image selected shows 17% accuracy and 100% recall values in the color histogram approach, while using the fusion of color histogram and color correlogram approach, the performance has been increased and the accuracy is 29% and Recall is 28%. It also shows the comparison of Precision and recalls against the number of retrieved images for Data sets in HSV Color Space. When only a single color histogram is used accuracy was 42% and the recall value was 88%. While using the fusion of two color descriptors, gives an accuracy of 100% and a recall value of 6%. It has been determined that the HSV color space gives better results as compared to the RGB color space. The RGB color space gives an accuracy of 28%, while the HSV color space gives 100%. During experimentation, both the HSV color model as well as RGB color model is used for the same process of retrieval. This paper [3] claims that the retrieval accuracy is close to 100% in the proposed algorithm. It has also been observed that HSV color space gives more accurate results when compared to RGB color space.

A paper by Mariofanna Milanova [4] uses Adaptive Inverse Pyramid Representation which claims that for a good visual information retrieval (VIR) system, apart from performing similarity matching on media objects represented by features vectors, user feedback is also necessary for the retrieval process. It was considered that relevant feedback would be a breakthrough in these regards but even the most sophisticated algorithms these days do not meet the need for similarity-based retrieval at a satisfactory level. The author presents a new approach for content-based image retrieval that is based on image representation with Adaptive Inverse Difference Pyramid (IDP) decomposition controlled by a neural network. It uses a multi-layer image retrieval approach. The use of the IDP decomposition permits the creation of efficient multi-view models. The important advantage of this approach is the multi-scale representation, based on the relations between transform coefficients in adjacent layers, which offer a significant reduction of the transform coefficients needed for the object model creation [5]. The introduction of flexible feedback in the process of object model creation and search makes this approach close to the human way of thinking. This method develops flexible models for some basic feature sets, such as texts, graphics, cartoon images, medical images, natural grayscale or color images, etc.,

Thus, it is determined that HSV colorspace is more robust towards external lighting changes because Hue values vary relatively lesser than RGB values. This implies that using HSV instead of RGB would give a more accurate result. [3]. A color histogram captures only the color distribution in an image and does not include any spatial correlation information. Thus, a new technique of combining Color histogram and Color Correlogram would provide an effective Content-Based Image Retrieval algorithm [3]. This would also improve precision and recall. The adaptive Inverse Pyramid Representation approach shows promising results with the introduction of flexible feedback in the process of object model creation and searches making this approach close to the human way of thinking [4].

Overcoming the Limitations

1. HSV colorspace is more robust towards external lighting changes. Hue values vary relatively lesser than RGB values. Thus, using HSV instead of RGB would give a more accurate result. [3]

2. A color histogram captures only the color distribution in an image and does not include any spatial correlation information. Thus, the paper [3] proposes a new technique of combining Color histogram and Color Correlogram to provide an effective Content-Based Image Retrieval algorithm. This would improve precision and recall.
3. The Adaptive Inverse Pyramid Representation approach shows promising results with the introduction of flexible feedback in the process of object model creation and searches making this approach close to the human way of thinking.

REFERENCES

[1] Sclaroff, S, A. Pentland, and R Picard. "Photobook: Tools for Content-Based Manipulation of Image Databases ." *SPIE Storage and Retrieval of Image & Video Databases II*. February 1994. http://vismod.media.mit.edu/vismod/demos/photobook/ (accessed November 5, 2008).

[2] Jeong, Sangoh. *Histogram-Based Color Image Retrieval*. Psych221/EE362 Project Report, Stanford University, 2001.

[3] Devyani Soni; K.J. Mathai, "An Efficient Content Based Image Retrieval System Based on Color Space Approach Using Color Histogram and Color Correlogram", 2015 Fifth International Conference on Communication Systems and Network Technologies.

[4] *Content Based Image Retrieval Using Adaptive Inverse Pyramid Representation*, Mariofanna Milanova, Roumen Kountchev, Stuart Rubin, Vladimir Todorov, and Roumiana Kountcheva, https://link.springer.com/content/pdf/10.1007%2F978-3-642-02559-4_34.pdf

[5] Kountchev, R., Kountcheva, R.: *Image Representation with Reduced Spectrum Pyramid*. In: Tsihrintzis, G., Virvou, M., Howlett, R., Jain, L. (eds.) New Directions in Intelligent Interactive Multimedia, pp. 275–284. Springer, Heidelberg (2008).