Pixort: A Novel Approach for Effective Photo Album Clustering

Cheryl L Mathias¹, Crystal F D’Souza¹, Job Alexander¹, Mariah S Hudson¹ and Renuka Tantry¹

¹Department of Computer Science and Engineering, St Joseph Engineering College, Mangaluru, Karnataka, India

E-mail: crystallaydsouza@gmail.com

Abstract. Retrieving digital records on the internet and capturing, uploading images on social media or other platforms has become such a notable part of one’s daily existence. Digital content and videos constitute majority of this ever-growing deluge of information. Major problems are faced by people on a daily basis when they have to manually scan large databases to search for particular images thereby exhausting valuable time and resources. To overcome this, Content Based Image Retrieval in addition to Emotion Detection must be performed to efficiently store and retrieve very specific images. An effective image clustering application should be capable of accurately extracting images based on the given query image. It should also provide functionalities like Facial Detection and Recognition for customized usage and retrieval. In this study, texture-based feature extraction followed by the usage of a feed forward backpropagation neural network has been proposed for CBIR. For emotion detection, correlation has been proposed in addition to the Viola Jones algorithm which has been used for facial detection.

1. Introduction
Digitization of the photography industry has led to an incredible ease in the acquisition of a camera thereby leading to the generation of a huge collection of images and data stored in a digital format [1]. Due to the massive growth in the amount of visual information available, information systems have to be upgraded to be able to manage this data. Most of the search engines on the Internet retrieve the images based on text that rely on headings as input [1]. Manually labelling the images in existing image archives containing millions of images is nearly impossible. Content Based Image Retrieval (CBIR) can thus be an effective solution to the above-mentioned problems [2], however, it is only used to retrieve inanimate objects and scenarios due to its inability to identify faces and detect emotions.

Human faces primarily convey information about the identity. Facial expressions are the most natural and immediate means for communication of emotions. At the moment, there are numerous commercial systems for facial detection and recognition. While these systems have displayed decent and acceptable level of performance in simple and controlled environments, their performance has been found to be much worse when deployed in real-time environments due to variation in factors such as pose, illuminative conditions, facial angles and viewpoint [3]. Thus, the goal of this research is to minimize the influence of these unpredictable factors.

CBIR can be carried out using machine learning and deep learning. In comparison to deep learning techniques which require high computational power, memory and specialized GPU usage, machine learning outperforms deep learning in terms of retrieval performance. Machine learning techniques that
can be used to implement CBIR are shown in figure 1. A trained network can solve highly complex problems with remarkable accuracy and performance [4][5].

![Figure 1. Machine Learning Techniques to Implement CBIR [1]](image)

The significant operations in a facial detection, recognition and emotion detection system are, image pre-processing, face detection, feature extraction, matching and classification.

A significant facial detector is the Viola-Jones detector. It performs well in real time due to three reasons. First, it uses the integral image to compute Haar-like features rapidly. Second, it is able to identify and select the best features by using classifier learning with AdaBoost. Third, it employs the attentional cascade structure which is instrumental in rejecting and discarding a major portion of the sub-windows in the early layers of the detector, thereby increasing the efficiency of the detection process [6]. This makes the Viola-Jones detector robust and fast with an accuracy of 0.74 [7] and has been implemented in various applications.

Some of the popular techniques for feature extraction include Gabor Wavelets, Speeded Up Robust Features (SURF) and many others. Due to the simplicity and speed of Local Binary Patterns (LBP), it is prominently used in facial detection and recognition systems.

Various techniques have been proposed for facial recognition and classification based on emotions such as Support Vector Machines (SVM), Random Forests and Artificial Neural Network. Overfitting, proper kernel function are issues with SVM. Random forests are difficult to visualize. ANN requires a large amount of training data [8]. To overcome these issues, digital image correlation can be used. Correlation techniques are easy, result in good computational efficiency and are implemented in multiple commercial and open source libraries like MATLAB, OpenCV and many more [9].

In this study, for CBIR, texture-based feature extraction followed by classification using a feed forward backpropagation multi-layer neural network has been proposed. For facial detection, the Viola-Jones detector and for facial recognition and categorization of images based on emotions, correlation has been proposed.

The paper is organized as follows: Section 1 introduces CBIR and emotion detection and provides justification for the different techniques used. Section 2 discusses the methodology used to achieve the desired result, including the different algorithms involved. Section 3 describes the proposed system to carry out CBIR and emotion detection. Section 4 discusses the results and its accuracy followed by conclusion and future work, explained in section 5.
2. Methodology

2.1. Feature Extraction using Local Binary Patterns (LBP)

The Local Binary Patterns (LBP) method was proposed by Ojala et al. [10] in 1996. It is an efficient image operator which is used to transform an image into an array or image consisting of integer labels describing textures and pixel intensity of the image. Local binary pattern operator is generally applied on a 3×3 pixel,

\[
\text{LBP}(x, y) = \sum_{i=0}^{7} s(g_i - g_c)2^i
\]

where, \( s(x) = \begin{cases} 
1, & x \geq 0 \\
0, & x < 0 
\end{cases} \)

This operator uses the value of the centre pixel as threshold and works around the eight neighbouring pixels. 1 is assigned to a neighbouring pixel if its grey value is greater than or equal to the centre pixel and 0 is assigned otherwise. The eight ones or zeros are then concatenated, resulting in a binary code, which is the LBP code for the centre pixel as shown in figure 2. [3]

![Figure 2. Computation of LBP](image)

2.2. Feed Forward Back Propagation Neural Network (FFBPNN)

After feature extraction, FFBPNN is employed for pattern recognition and computing the parallelism between the query image given as input and the images in the database [11]. FFBPNN is a robust and efficient approach for classification. The structure of a FFBPNN is shown in figure 3.

Choosing an appropriate FFBPNN predictor involves selection of various design parameters including usage of optimal: i) number of hidden layers ii) number of neurons in the input and hidden layers iii) activation function for each layer iv) learning rate and v) number of epochs. The network functions in both directions, forward and backward. Input data is processed in the forward direction and the output of neurons from one layer is linked to the corresponding input of neurons of the next layer.
In backpropagation, the errors from the neurons of the output layer are sequentially backpropagated via the hidden layers and then to the input layer, that is, from right to left [11].

![Diagram of FFBPNN](image)

**Figure 3.** Structure of a FFBPNN

During training, random values are used to initialize the weights and, in every iteration, the weight is adjusted by minimizing the error using backpropagation. As a result, the weights will gradually converge to the optimal value after a variable set of epochs. A well-generalized model requires a large set of training data but is always preferred. Poor generalization leads to overfitting which can be caused due to excessive training thereby leading to poor performance on the testing data [12].

2.3. Facial Detection using Viola-Jones Algorithm

The Viola-Jones Algorithm is used for real-time detection of faces in an image. For easier detection and management, it requires full view frontal upright faces. The algorithm is robust [7][13], providing a high true-positive rate and a low false-positive rate. The algorithm has four stages:

2.3.1. Haar Feature Selection. Haar-like features are digital image features that are primarily used for object detection and recognition in images as shown in figure 4. The value of the rectangular features can be calculated as: [6]

$$\sum_{i \leq N \leq M} \sum_{j \leq N \leq M} I(i,j)_{white} - I(i,j)_{black}$$

(2)

![Haar Feature Similar to Bridge of the Nose](image)

**Figure 4.** Haar Feature Similar to Bridge of the Nose
2.3.2. Creating an Integral Image. An Integral image simplifies the process of calculation of the Haar rectangle feature. For every location (i, j), the integral image computes the summation of the pixels above and to the left of (i, j), inclusive as shown in figure 5 [6].

![Integral Image](image)

**Figure 5.** (a) Original Image (b) Integral Image

2.3.3. AdaBoost Training. AdaBoost Training is employed to eliminate the redundant features and choose a smaller number of likely features. A weighted linear combination consisting of weak classifiers is used to build a complex classifier where every feature is regarded as a weak classifier.

2.3.4. Cascading Classifiers. An attentional cascade as shown in figure 6 is used to correctly classify the misclassified images from the previous stages, thus achieving a higher detection rate.

![Attentional Cascade Classifier](image)

**Figure 6.** Attentional Cascade Classifier

2.4. Facial Recognition and Classification into emotions using Pearson Product-Moment Correlation Coefficient (PPMCC)

For every pixel between two consecutive images, the difference between the value of intensity at that pixel and the average intensity of the whole image is computed. The correlation coefficient is returned in terms of values that lie in between -1 and 1, where 1 suggests a strong positive relationship, -1 suggests a strong negative relationship and 0 indicates no relationship. This value can then be used for recognition and can aid in the classification of the images based on emotion.

There are various formulas for the correlation coefficient calculation. In MATLAB, corr2 function computes the correlation coefficient using,
\[ r = \frac{\sum_m \sum_n (A_{mn} - \bar{A})(B_{mn} - \bar{B})}{\sqrt{\left( \sum_m \sum_n (A_{mn} - \bar{A})^2 \right) \left( \sum_m \sum_n (B_{mn} - \bar{B})^2 \right)}} \]  

where \( \bar{A} = \text{mean2}(A) \), and \( \bar{B} = \text{mean2}(B) \), mean2 is used to compute the mean of the elements of a matrix.

3. Proposed System

3.1. CBIR

The system proposed for CBIR is shown in figure 7. The GUI for communication with the retrieval engine is designed using MATLAB GUIDE and the development is carried out using MATLAB IDE.

![Proposed System for CBIR](image)

Figure 7. Proposed System for CBIR

Before feature extraction, the images used for training have to be pre-processed by resizing them into 256x256 size followed by conversion from sRGB to L*a*b* space to retain the lightness parameter and
leave out the color. These images are then subjected to LBP feature extraction algorithm for extraction of texture related features. The mean LBP value for all images in the database is computed and stored in an array which is then given as input to a FFBPNN in addition to the target array consisting of the labelled target values.

The design of the FFBPNN plays an integral role in the efficiency of training and image retrieval. Three hidden layers are used with 140, 150 and 1 node each and the logsig activation function is used for the first two layers and purelin for the third. The process of training is carried out using trainrp constant used in MATLAB for resilient backpropagation. It is slower than its contemporaries but results in improved memory efficiency. A learning rate of 0.03 is maintained at 20000 epochs. The FFBPNN is simulated using the Neural Network Toolbox in MATLAB and the trained network is stored for further classification.

When a query image is given as input, the same procedure of pre-processing followed by feature extraction is carried out. The pretrained FFBPNN is then invoked for classification. Thus, the network identifies the object and fetches relative images from the database.

3.2. Facial Detection, Facial Recognition and Emotion Detection

3.2.1. Creation of Templates. Figure 8 describes the steps involved in the creation of templates. The first step is the creation of two templates using the images from the database. All the images are subjected to the system object detector created by MATLAB’s vision.CascadeObjectDetector to yield all the faces in the images. The faces are then organized into two folders: i) according to person and ii) according to emotion. The faces in both folders are then resized and subjected to the LBP algorithm for feature extraction. The resulting values are then converted to its unsigned int8 equivalent to reduce it to a value between 0 and 255, creating two templates that will assist in the facial recognition and in the emotion segregation process.

![Figure 8. Creation of Templates](image)

3.2.2. Facial Recognition and Segregation based on emotion using PPMCC. Figure 9 outlines the sequence of steps performed to accomplish facial recognition and segregation based on emotion. When
the query image is provided to the application, the image is resized and subjected to the system object detector created by MATLAB’s *vision.CascadeObjectDetector*. This detects the faces in the query image. The face is then cropped from the image, resized and converted from sRGB to grayscale. The conversion is essential to eliminate the saturation and hue components of the image leaving behind only the luminance component. The image is further converted to L*a*b* colour space. The next step is the feature extraction involving the calculation of the LBP value for the face. The value is then converted to its unsigned int8 equivalent.

![Proposed System for Facial Detection, Facial Recognition and Emotion Detection](image)

**Figure 9.** Proposed System for Facial Detection, Facial Recognition and Emotion Detection

The person in the query image can be identified by using the correlation coefficient that measures the strength of the relationship between the LBP value of the query image given as input and the LBP values in the template organized by person. This step is performed using the corr2() function in MATLAB. The index of the maximum value of the coefficient will identify the person in the query
image. The same process is replicated on all the images in the database to retrieve images of the person in the query image.

After the images have been retrieved, they are classified into images depicting happiness and images depicting sadness. The correlation function corr2() is then used to evaluate the strength of the relationship between the LBP values of the retrieved images and the LBP values in the template organized by emotion. If the index of the maximum value of the coefficient crosses a threshold, the image is classified as happy. If not, the image is classified as sad. The threshold is the boundary index of the LBP values of the images in the template organized by emotion.

4. Results and Discussion
4.1. CBIR
The proposed system is trained using 400 images divided into 4 categories with 100 images of each category. A sample of the images is shown in figure 10.

![Sample of Training Images](image)

**Figure 10.** Sample of Training Images

The images have been divided into two sets for training and testing. The performance of CBIR can be measured using regression analysis.

![Regression Analysis](image)

(a) Overall Regression (b) Confusion Chart

**Figure 11.** (a) Overall Regression (b) Confusion Chart
The graph in figure 11 (a) shows the relationship between the predicted output and the target data in terms of the best fit. Hence, it is concluded that for the overall system, the output is classified correctly in comparison to the target data with a regression rate of 0.81597. The confusion chart is also plotted as shown in figure 11 (b).

A good average accuracy of 81% is achieved using the proposed CBIR system. A sample of the result is shown in figure 12.

![Figure 12. (a) Query Image (b) Relative Images Fetched](image)

### 4.2. Facial Detection, Facial Recognition and Emotion Detection

For the proposed system, 248 images are used as training data from which 305 relevant faces are extracted to identify 9 different faces and stored in the first template and these 305 images are further split based on 2 emotions - happy and sad and stored in the second template.

The performance of this system can be evaluated using precision, recall and accuracy.

\[
\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \\
\text{precision} = \frac{TP}{TP + FP} \\
\text{recall} = \frac{TP}{TP + FN}
\]

| Face | TP | TN | FP | FN | Precision | Recall | Accuracy |
|------|----|----|----|----|-----------|--------|----------|
| 1    | 31 | 3  | 1  | 0  | 0.96      | 1.0    | 0.97     |
| 2    | 56 | 5  | 7  | 0  | 0.88      | 1.0    | 0.89     |
| 3    | 39 | 7  | 5  | 5  | 0.88      | 0.88   | 0.82     |
| 4    | 16 | 8  | 1  | 2  | 0.94      | 0.88   | 0.88     |
| 5    | 11 | 2  | 3  | 0  | 0.78      | 1.0    | 0.81     |
| 6    | 34 | 5  | 5  | 0  | 0.87      | 1.0    | 0.88     |
| 7    | 10 | 3  | 2  | 0  | 0.83      | 1.0    | 0.86     |
| 8    | 8  | 4  | 2  | 0  | 0.80      | 1.0    | 0.85     |
| 9    | 14 | 5  | 3  | 2  | 0.82      | 0.87   | 0.79     |

TP is the total number of happy images classified correctly, TN is the total number of sad images classified correctly, FN is the number of happy images that were classified as sad, FP is the number of
sad images classified as happy. Thus, an overall precision of 0.86, recall of 0.96 and accuracy of 0.86 is achieved. A sample of facial detection, selection and segregation of images based on emotion is shown in figure 13.

![Figure 13](image)

**Figure 13.** (a) Facial Detection (b) Selecting a Face (c) Sample of Happy and Sad Images

5. **Conclusion and Future Work**

The main focus of this research was to demonstrate the significance of content based retrieval in image retrieval systems. In contrast to conventional approaches that are based on text, identifying and fetching images based on its content is an accurate and resourceful method for image retrieval. The proposed system was able to identify and retrieve images similar to the query image with an overall accuracy of 0.81. The application was also able to recognize faces and categorize the images based on two emotions using correlation with a precision of 0.86, recall of 0.96 and accuracy of 0.86.

CBIR is presently a dynamic research area in the deep learning and computer vision community due to its vast potential in both academic and industry ranging from medical and forensic applications to security applications. However, the major challenge lies in identifying features that adequately represent an image. This area can be further explored and the techniques can be fine-tuned to increase the retrieval efficiency. The neural network used can be trained extensively to fetch images on a broader spectrum and with an enhanced accuracy. Facial and emotion detection is an indivisible part of computer vision. The future scope of the proposed system lies in the ability to detect and recognize more faces with greater accuracy and further categorize them into six or more emotions.

6. **References**

[1] Latif A, Rasheed A, Sajid U, Ahmed J, Ali N, Ratyal NI, Zafar B, Dar SH, Sajid M and Khalil T 2019 Content-based image retrieval and feature extraction: a comprehensive review *Mathematical Problems in Engineering* 2019 1-21

[2] Bhardwaj S, Pandove G and Dahiya PK 2020 A genesis of a meticulous fusion based color descriptor to analyze the supremacy between machine learning and deep learning *I.J. of Intelligent Systems & Applications* 12 21-33

[3] Rahim MA, Azam MS, Hossain N and Islam MR 2013 Face recognition using local binary
patterns (LBP) Global Journal of Computer Science and Technology 13 1-7

[4] He C, Ma M and Wang P 2020 Extract interpretability-accuracy balanced rules from artificial neural networks: a review Neurocomputing 387 346-58

[5] Erkaymaz O 2020 Resilient back-propagation approach in small-world feed-forward neural network topology based on Newman–Watts algorithm Neural Computing and Applications 20 16279-89

[6] Alionte E and Lazar C 2015 A practical implementation of face detection by using Matlab cascade object detector 19th Int. Conf. on System Theory, Control and Computing (ICSTCC) (Romania: IEEE) pp 785-90

[7] Kirana KC, Wibawanto S and Herwanto HW 2018 Facial emotion recognition based on Viola-Jones algorithm in the learning environment Int. Seminar on Application for Technology of Information and Communication (Semarang: IEEE) pp 406-10

[8] Islam B, Mahmud F, Hossain A, Goala PB and Mia MS 2018 A facial region segmentation based approach to recognize human emotion using fusion of HOG & LBP features and artificial neural network 4th Int. Conf. on Electrical Engineering and Information & Communication Technology (ICEEiCT) (Bangladesh: IEEE) pp 642-6

[9] Roncella R, Romeo E, Barazzetti L, Gianinnetto M and Sciaioni M 2012 Comparative analysis of digital image correlation techniques for in-plane displacement measurements 5th Int. Congress on Image and Signal Processing (Chongqing: IEEE) pp 721-6

[10] Ojala T, Pietikäinen M and Harwood D 1996 A comparative study of texture measures with classification based on featured distributions Pattern recognition 29 51-9

[11] Gururaj C, Jayadevappa D and Tunga S 2016 Content based image retrieval system implementation through neural network IOSR Journal of VLSI and Signal Processing (IOSR-JVSP) 6 42-7

[12] Bilbao I and Bilbao J 2017 Overfitting problem and the over-training in the era of data: particularly for artificial neural networks 8th Int. Conf. on Intelligent Computing and Information Systems (ICICIS) (Cairo: IEEE) pp 173-7

[13] Ismael KD and Irina S 2020 Face recognition using Viola-Jones depending on Python Indonesian Journal of Electrical Engineering and Computer Science 20 1513-21