International Journal of Electrical and Computer Engineering (IJECE)
Vol. 8, No. 3, June 2018, pp. 1903~1912
ISSN: 2088-8708, DOI: 10.11591/ijece.v8i3.pp1903-1912

Content-based Image Retrieval System for an Image Gallery Search Application

Nicole Tham Ley Mai[1], Syahmi Syahiran Bin Ahmad Ridzuan[2], Zaid Bin Omar[3]
[1,2]Faculty of Electrical Engineering Universiti Teknologi Malaysia, Johor Bahru, Malaysia
[3]Department of Electric and Computer Engineering Faculty of Electrical Engineering Universiti Teknologi Malaysia Johor Bahru, Malaysia

Article Info

Article history:

Received Jan 27, 2018
Revised Apr 20, 2018
Accepted Apr 26, 2018

Keyword:

Auto-tagging
Content-based image retrieval
Mpeg-& powered localized descriptor
Principal component analysis
Text-based image retrieval

ABSTRACT

Content-based image retrieval is a process framework that applies computer vision techniques for searching and managing large image collections more efficiently. With the growth of large digital image collections triggered by rapid advances in electronic storage capacity and computing power, there is a growing need for devices and computer systems to support efficient browsing, searching, and retrieval for image collections. Hence, the aim of this project is to develop a content-based image retrieval system that can be implemented in an image gallery desktop application to allow efficient browsing through three different search modes: retrieval by image query, retrieval by facial recognition, and retrieval by text or tags. In this project, the MPEG-7-like Powered Localized Color and Edge Directivity Descriptor is used to extract the feature vectors of the image database and the facial recognition system is built around the Eigenfaces concept. A graphical user interface with the basic functionality of an image gallery application is also developed to implement the three search modes. Results show that the application is able to retrieve and display images in a collection as thumbnail previews with high retrieval accuracy and medium relevance and the computational requirements for subsequent searches were significantly reduced through the incorporation of text-based image retrieval as one of the search modes. All in all, this study introduces a simple and convenient way of offline image searches on desktop computers and provides a stepping stone to future content-based image retrieval systems built for similar purposes.

Corresponding Author:

Zaid Bin Omar,
Department of Electric and Computer Engineering,
Faculty of Electrical Engineering,
Universiti Teknologi Malaysia,
Johor Bahru, Malaysia.
Email: lsntl@ccu.edu.tw

1. INTRODUCTION

In recent years, rapid advances in electronic storage capacity and computing power have triggered the growth of large digital image collections following the increase of users on the internet. Over the years, we have seen exponential increases in number of digital images and video both over the net and even in our own devices as we attempt to keep more memories through photos and videos. This increased usage may be due to several factors such as education, entertainment, commercial purposes, and etc. and it is now apparent that more and more images are routinely used to convey large amounts of information.

Due to the increasing difficulty in making proper use of the information contained in digital images and videos, advanced information systems are now more important than ever as they are needed to manage

image collections more efficiently. Image searching is one of the most important functions that needs to be supported by devices and computer systems to allow efficient browsing, searching, and retrieval. With the rapid advancement in image capturing devices over the years added with the advent of various social media platforms, it has become a common culture for members of current society to take a lot of photos with their phones every day. According to a survey by personal media startup Magisto, the average smartphone user takes around 150 new photos every month. Despite this, current gallery applications cannot keep up with huge databases of images and the only common method available to retrieve old images is to scroll through many unwanted images before arriving at the desired image which is very tedious and time consuming. Moreover, conventional CBIR systems are generally computationally heavy for offline applications where performance is expected to be fast while still being able to produce relevant results as computing speed may vary from computer to computer.

Content-Based Image Retrieval (CBIR) is one instance of information retrieval that applies computer vision techniques to solve problems related to searching and managing large image databases. However, most CBIR systems that aims to manage digital collections in offline database tend to use image content as query rather than considering user preference in defining the image in question and this may not be convenient especially when it is the only available mode of search as future efforts of searching for the same query image may be redundant. Hence, this project aims to make image galleries more organized by introducing a combination CBIR and TBIR-based system for more convenient offline searches through automatic generation of textual metadata by using information obtained from user input and previous retrieval results.

2. LITERATURE REVIEW

In most of the earlier retrieval systems, video or image contents are managed by keywords or textual metadata [1]. Content-based image retrieval (CBIR) however relies on extracting the appropriate characteristic quantities called ‘descriptors’ or ‘features’ describing the desired contents [2], [3]. A CBIR system consists of an interface for the acquisition of the query image, databases for storing indexing data and distance metrics, and a similarity comparison and retrieval system.

2.1. Feature vector extraction

Some commonly used feature vectors include colour, texture, shape, spatial location, etc. Due to its stability and robustness; application of color features is widely accepted in most CBIR applications. Jalab H.A [4] implemented an image retrieval system based on color layout descriptor (CLD) representing the spatial distribution of colors, Jayamala K.Patil and Raj Kumar [5] suggested a plant leaf disease image retrieval using color moments, Chatzichristofis et al [6] proposed a colour and edge directivity descriptor (CEDD) incorporating both color and texture information in a histogram. Texture feature contain valuable information on the surface structures of objects and their relationship to the surrounding environment [7]. Based on previous research, it is found that the most important texture features are coarseness, contrast, directionality. The Steerable Pyramid Model [8] and Gabor wavelet Transform (GWT) [9] are among the most widely used features. Usually shape feature representations are only useful after image segmentation. Kauppinen et al. have shown that Fourier descriptors used in 2-D shape classification performs better compared to autoregressive modelling based shape descriptor [10].

2.2. Facial recognition

The most common methods of facial recognition are Eigenfaces, Fisherfaces and Local Binary Patterns Histograms (LBPH). The Principal Component Analysis (PCA) proposed by Karl Pearson (1901) and Harold Hotelling is a core component of the Eigen faces method which tries to focus on the most important components of the dataset, however it does not consider classes [11]. In Fisherfaces approach, Linear Discriminant analysis is used to perform dimensionality reduction by classes [12]. LBPH analyses each face in the training set separately and independently [13]. Apart from the traditional methods, there are also some modern research done is this area and the results are promising. DeepFace [14] is used by Facebook to automatically suggest a tag for faces in photos and videos. FaceNet [15] by Google uses a Euclidean space for image representation created images generated through a data-mining method.

2.3. Similar application

Getty Images is an extensive web-based gallery that sells high-quality stock images for use of advertising, marketing, and more. It is based on TBIR by collective tagging where several human indexers look at new image and enter associated keywords and previous user queries are combined to form a thesaurus for future searches. One of the earliest commercial use CBIR systems is the Query By Image Content (QBIC)

developed by IBM. It supports queries based on user sketches, query images, as well as color and texture patterns selected by the user [16] and uses a combination of color, texture, and shape as feature descriptor which includes an improved version of the Tamura texture representation [17], and major axis orientation. It has very fast performance and the results are invariant to small changes in perspectives, however some of its weaknesses include sensitivity to illumination changes and no localization of colour. Like.com, now owned by Google under google.com /shopping uses CBIR to search for products similar to query and return results of similar items with links to retailers such as Amazon.com. It also allows users to select regions of a product image to retrieve products ranked by similar patterns, colours and shapes. Users can determine which of these three criteria are more important to them to improve the search results. The company Google also has another application called Google Photos which is a cloud-based application that uses facial recognition to find photos of people in the gallery. This requires the user to first apply a label to a photo of someone and they will be able to search for that particular individual using that label. Users may also go online and search photos by common keywords without defining them.

3. METHODOLOGY

The system consists of three main components: the gallery interface, the query processing module, and the image database and allows three modes of searches: by a reference image, a name, or a previously defined tag. Figure 1 shows the top-level block diagram for the overall system.

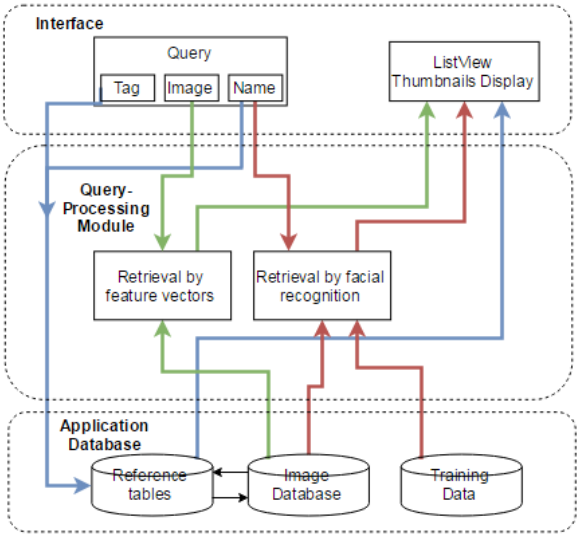


Figure 1. Top-level block diagram of the system

3.1. Retrieval by feature vectors

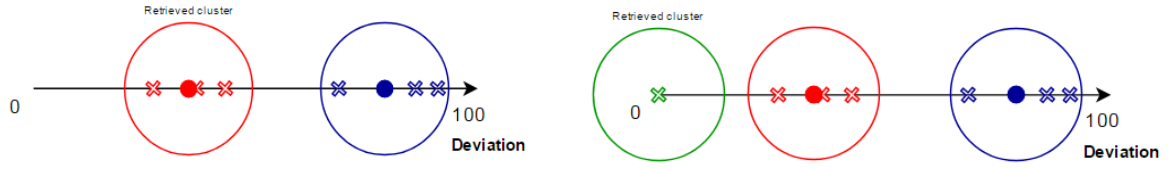
The MPEG-7-like Powered Localized Color and Edge Directivity Descriptor (SIMPLE-CEDD) is used as our feature vector [18]. As the size of a CEDD descriptor is memory efficient (only 54 bytes per image) [6], and requires relatively low computational power to extract, it is suitable to be used for searching large image databases such as a local image gallery in a computer. Furthermore, SIMPLE-CEDD localizes the image features by first locating feature-rich regions and define these patches as regions-of-interest or ROI before extraction. As a result, the feature vector becomes more robust to image transformations and allows faster execution. The extracted feature vectors are compared for similarity using the Tanimoto coefficient is described in Equation (1) where a and b are two separate points.

$$Tanimoto\ Coefficient = \frac{\sum_{j=1}^{k} a_j \times b_j}{\sum_{j=1}^{k} a_j^2 + \sum_{j=1}^{k} b_j^2 - \sum_{j=1}^{k} a_j \times b_j} \quad (1)$$

To filter the results to only show the most similar images, a custom variant of K-means is used instead of a low pass filter as the optimum difference threshold (meaning that ideally, all visually similar images should be associated with difference values below this threshold) may vary with different datasets, hence K-means is used to adapt to the changes of this threshold value. The objective of K-means clustering is

to group n number of elements or data points into k number of clusters by minimizing the total intra-cluster variance as in Equation (2), where J is the objective function, xi is the observation for case i, and cj is the centroid for cluster j.

For this implementation, the number or clusters, k is not defined but gradually added when a point lies outside the boundaries of the cluster radius of all existing centroids during the first iteration. The cluster radius is defined at the beginning of the iteration and serves as the difference threshold for image retrieval. However, the retrieval output will always take the first cluster of images with the least deviation regardless of whether or not they are inherently similar to the query image. (Figure 2(a)) To resolve this issue, an element with a deviation value of 0 (totally similar) is added into the array before clustering. This element does not correspond to any images whatsoever, but added to retrieve images only when similar images exist (Figure 2(b)).



(a) (b)

Figure 2. Retrieval results with no similar images, (a) no dummy element, (b) with dummy element

3.2. Retrieval by facial recognition

Before the system is able to search for images of individuals, the user first has to add a face into the training dataset. For this task, the Haar feature-based cascade classifiers is used for face detection. Image regions that are likely to contain faces by are located by scanning the image several times at different scales by increasing the ROI in subsequent rounds to find the face. All containing regions with detected faces are saved to the training dataset along with the name that the user wishes to identify it as. This information will be useful later during text-based search. To speed up this process, the Canny edge detector is used to ignore image regions with too few or too much edges.

Facial recognition is done using Principal Component Analysis (PCA). In this case, our set of training images is converted into a set of calculated Eigenfaces. For each proceeding Eigenfaces, there are lesser features and more noise, hence only the few first K Eigenfaces are selected. This way, the number of values needed to recognize it is reduced and this helps to speed up the recognition process and reduce error caused by noise. As a result, the total time taken to retrieve images of individuals can be reduced. Once all images are decomposed as Eigen values, the Euclidean Eigen-distance between the query image and every other training image in the database is calculated as in Equation (3).

$$\|array1 - array2\| = \sqrt{\sum_i (array1(I) - array2(I))^2} \quad (3)$$

3.3. Retrieval by text

By incorporating ‘tags’ to the CBIR system, the computation time can be reduced for subsequent searches by first tagging a single or multiple reference images. The query-processing module then proceeds to retrieve the most similar images with respect to the query image and the most similar images will then be classified into the same tag category during automatic tagging. On subsequent searches, the user may simply input a previously defined tag and the system proceeds retrieve the most similar images with the specified tag without performing extraction and similarity computation.

Auto-tagging works almost similarly for retrieval by feature vectors and facial recognition since the output of both process blocks are a series of relevant images. Tagging is done with a reference table, containing information that is updated by the application during runtime and is needed to retrieve images quickly without having to extract and inspect the embedded tags of every image in the database. Names and normal tags are treated as separate entities and hence have their own respective reference tables. For both searches, relevant images retrieved are automatically tagged according to user input. During auto tagging, the EXIF meta information of the file itself is changed to add the input tag, then, this information will be added to the reference table. To retrieve tagged images, the application will scan though the reference table and get the image path. Since the tag may vary according to user preference, the EXIF metadata is checked. If the

information of that particular image no longer contains the corresponding search query or if the image itself no longer exists, the corresponding row of information will be removed.

4. RESULTS

4.1. Performance of CBIR by feature vectors

To remove the subjectivity of human perception in classifying images, the COIL-100 dataset is used to test the reliability of the CBIR system. The result is obtained with cluster radius or difference threshold of 20. 8 objects as shown in Figure 3 are selected from the dataset as query image and are chosen by their varying difficulty to differentiate ranging from a (easiest) to h (most difficult).

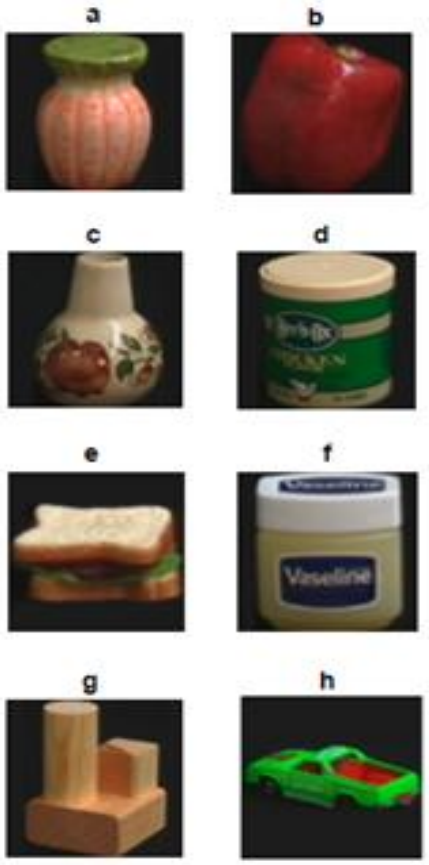


Figure 3. List of query images

Table 1. Retrieval Performance for difference Threshold of 20

	A	b	C	d	E	f	g	h
No. of relevant images retrieved	72	72	72	62	58	72	66	66
Precision (%)	98.6	97.3	64.9	100	100	65.5	26.7	32.7
Recall or Sensitivity (%)	100	100	100	86.1	80.6	100	91.7	91.7
Specificity (%)	99.9	99.9	99.5	100	100	99.5	97.5	98.1
Relevance (%)	98.6	97.3	64.9	116.1	124	65.5	29.1	35.7
Accuracy (%)	99.9	99.9	99.8	93.1	90.3	99.8	94.6	94.9

Figure 4 and Figure 5 are the ROC curve and the Precision-Recall curve for this particular set of query images. The graphs are plotted with multiple sets of precision, recall, and specificity values for different threshold values ranging from 5 to 30.

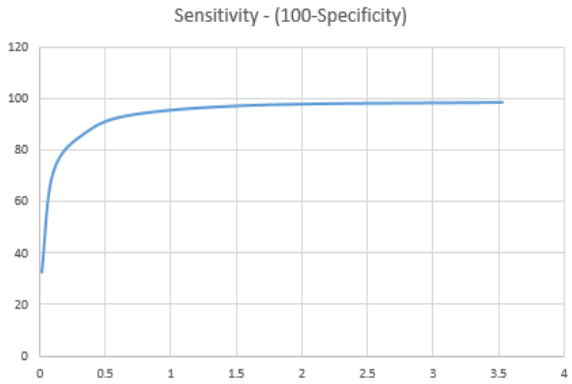


Figure 4. ROC Curve

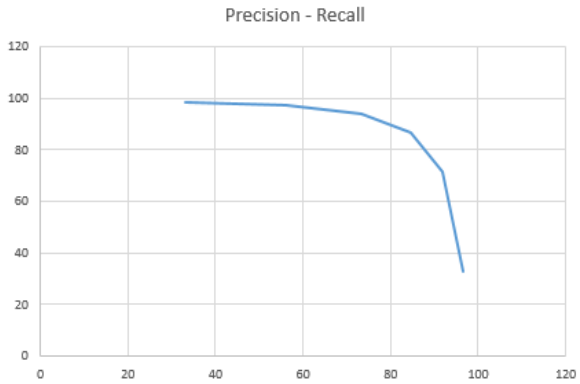


Figure 5. Precision-Recall curve

4.2. Performance of CBIR by facial recognition

Images for this dataset are produced by Dr. Libor Spacek from the Department of Computer Science of University of Essex. The set for this test contain 56 images of 8 different individuals, 4 males and 4 females. Faces of 4 individuals are trained with 4 images and these images are omitted from the gallery database to test the reliability of the facial recognition system. Figure 6 shows the training images for four different individuals.

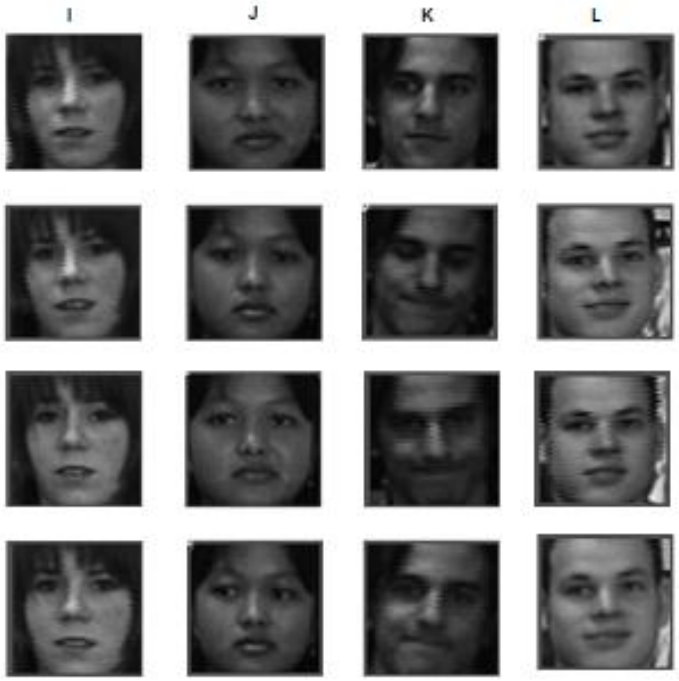


Figure 6. Training images for 4 different individuals

Table 2. Retrieval Performance for Eigen distance Threshold of 2500

	No. of relevant images retrieved	Precision (%)	Recall or Sensitivity (%)	Specificity (%)	Relevance (%)	Accuracy (%)
I	7	100	100	100	100	100
J	7	31.8	100	73.2	31.8	86.6
K	7	35	100	76.8	35	88.4
L	7	100	100	100	100	100

Average relevance for this set = 66.7%
Average accuracy for this set = 93.75%

Figure 7 show some false detections of individuals whose faces are untrained. Note that subject I, J, K, L in this experiment is named as ‘one’, ‘two’, ‘three’, and ‘four’ respectively in the detection labels.

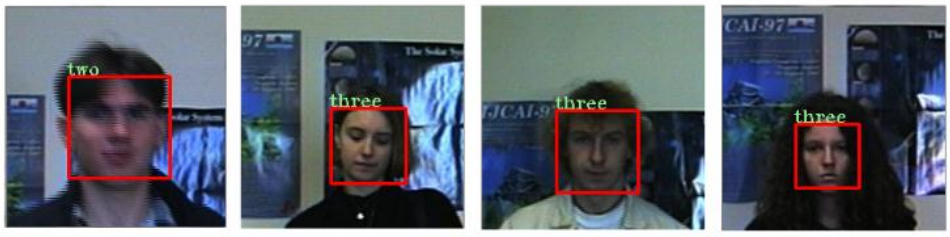


Figure 7. False detections

4.3. Graphical user interface

Shown in Figure 8 is the user interface built to implement the overall system. Table 3 describes its functionality by labeled regions in more detail.

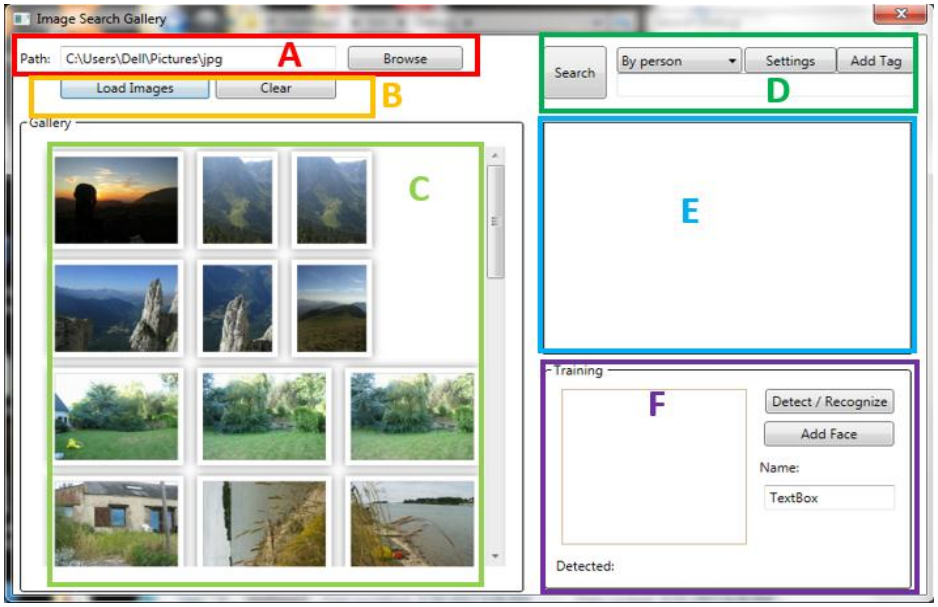


Figure 8. Overview of GUI

Table 3. GUI Functionality by labeled Regions

Region	Function
A	Select gallery database to be searched.
B	Load images without performing CBIR or clear image thumbnails.
C	Listview displays image thumbnails of loaded or retrieved images. All thumbnails are selectable and can be double clicked for a larger preview of the image.
D	Select search mode (by name, image, or tag) and add tag to selected image.
E	A preview of the query image will be shown here when searching by image
F	Allow user to add faces and their corresponding names into the database and see which individuals are detected in the selected image.

5. DISCUSSION

The system consists of three main components: the gallery interface, the query processing module, and the image database and allows three modes of searches: by a reference image, a name, or a previously defined tag. Figure 1 shows the top-level block diagram for the overall system.

5.1. Performance analysis of the feature vector

The descriptor is shown to have very high accuracy as shown in Table 1 as it is able to retrieve almost all instances of the objects (high recall) while weeding out a majority of unrelated images among all the 7128 images (high specificity). However, in some instances it is still easy to confuse the query image with other unrelated objects (medium relevance) in the database when the shape is inconsistent with different perspectives and when there are images with similar colour and texture in the database. The reason for this is because the retrieval system will always rank visually similar images as higher in similarity to help the users filter out images that are very likely unrelated rather than determining the context or definition of the images.

5.2. Analysis of ROC and precision-recall curves and selection of threshold value

Based on Table 1, there are special cases such as object d and e where the precision value is 100% while recall is less than 100%. This is likely because some of the other similar images in the queue did not make it into the first cluster due to a low difference threshold. With higher thresholds, the sensitivity or average recall will be higher, however the specificity will be lower as can be seen in Figure 3. This trade-off is also apparent in the precision-recall curve where the more we try to increase recall to find more instances of relevant images, the more the precision value decreases.

It is observed that a cluster radius or difference threshold between 20 and 25 is preferred as the last relevant image in the dataset will tend have deviation values in between that range. Though for more diverse images having more colours and diverse backgrounds, larger threshold values within that range are more effective. It is possible to assess the precision-recall trade-offs using the weighted harmonic mean in Equation (4) where F is the weighted harmonic mean, P and R is the precision and recall value respectively for that particular threshold and α is the weight to emphasize more on either precision or recall.

$$F = \frac{1}{\alpha\frac{1}{P}+(1-\alpha)\frac{1}{R}} \quad (4)$$

To choose the most suitable threshold, a threshold value corresponding to the pair of precision and recall value that has the highest F score should be chosen. If precision is more important to the user, α should be larger and conversely if recall is more important, α should be lower, although in this context it entirely depends on user preference. A more balanced F1 is described in Equation (5) and is more commonly used by researchers facing the precision-recall trade-off problem.

$$F = \frac{2PR}{P+R} \quad (5)$$

5.3. Performance analysis of the facial recognition system

The retrieval performance shows a high average accuracy however the results are not very consistent for different individuals and this is due to several factors. Notice in Figure 5 that individuals J and K have more variations in facial expressions in training images. In application, it is important to add more variations in expressions for each individual and in different lighting conditions as images of people tend to have different expressions on different sessions. However, in this case it causes some confusion with other individuals whose faces are not trained. This is one of the limitations of the Eigenface approach as it tends to look at the training dataset as a whole rather than analyzing each face in the set individually. However, the recognition can be improved by adding more individual faces into the dataset to reduce the number of unknown individuals or by increasing the Eigen distance threshold to increase precision.

6. CONCLUSION

All the objectives have been met and the three different search modes were successfully implemented on an image gallery software. For retrieval by query image, the feature vectors are extracted using SIMPLE-CEDD and results show a medium average relevance of 78.9% and a high accuracy of 96.5%. This shows that the system is effective in filtering out unwanted images however it is possible to further improve the overall relevance by incorporating classification algorithms to classify visually different images so that the retrieval results can be improved by taking into account the context or definition of the images themselves. Next, retrieval by facial recognition is built around the Eigen faces method. The results show an average relevance of 66.7% and an accuracy of 93.7%, however in some cases the results are not desirable due to false detections for unknown faces and this is an inherent weakness in the Eigenfaces method. As future work, it is recommended to explore other methods of facial recognition such as LBPH to determine if it is possible to increase the precision (and thus, relevance) of the retrieval results. And lastly, the

computational requirements of subsequent searches were reduced by integrating CBIR with TBIR using the concept of reference tables.

REFERENCES

[1] Long, F., Zhang, H., & Feng, D. D., (2003), "Fundamentals of content-based image retrieval", In *Multimedia Information Retrieval and Management*, pp. 1-26, Springer Berlin Heidelberg.

[2] R Zhang, Y. J., (2005), Advanced Techniques for Object-Based Image Retrieval.

[3] Y. K. J. K. Zukuan WEI, Hongyeon KIM, "An efficient content based image retrieval scheme," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 11, no. 11, pp. 6986-6991, November 2013.

[4] Jalab, H. A. (2011, September), "Image retrieval system based on color layout descriptor and Gabor filters".

[5] Jayamala Kumar Patil, Raj Kumar, (2013), "Plant Leaf Disease Image Retrieval Using Color Moments", *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 2, no. 1, pp. 36-42.

[6] Chatzichristofis, S. A., & Boutalis, Y. S., (2008, May), "CEDD: colour and edge directivity descriptor: a compact descriptor for image indexing and retrieval".

[7] Tamura, H., Mori, S., & Yamawaki, T., (1978), "Textural features corresponding to visual perception", *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 8, no. 6, pp. 460-473.

[8] Simoncelli, E. P., & Freeman, W. T., (1995, October), "The steerable pyramid: a flexible architecture for multi-scale derivative computation", In *ICIP*, vol. 3, pp. 444-447.

[9] B. S. Manjunath and W. Y. Ma. "Texture features for browsing and retrieval of large image data" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (Special Issue on Digital Libraries), vol. 18, no. 8, August 1996, pp. 837-842.

[10] Kauppinen, H., Seppanen, T., & Pietikainen, M, (1995), "An experimental comparison of autoregressive and Fourier-based descriptors in 2D shape classification", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 2, pp. 201-207.

[11] Vidya MS, Arul K, "Automated Attendance System Through Eigen Faces Using Image Processing", *International Journal of Informatics and Communication Technology (IJ-ICT)*, 2016 Dec 1, vol. 5, no. 3, pp. 111-118.

[12] Fisher, R. A., (1936), "The use of multiple measurements in taxonomic problems", *Annals of eugenics*, vol. 7, no. 2, pp. 179-188

[13] Ahonen, T., Hadid, A., & Pietikäinen, M, (2004), "Face recognition with local binary patterns", *Computer vision-eccv 2004*, pp. 469-481.

[14] Taigman, Y., Yang, M., Ranzato, M. A., & Wolf, L, (2014), "Deepface: Closing the gap to human-level performance in face verification", In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1701-1708.

[15] Schroff, F., Kalenichenko, D., & Philbin, J, (2015), "Facenet: A unified embedding for face recognition and clustering", In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 815-823).

[16] Niblack, C. W., Barber, R., Equitz, W., Flickner, M. D., Glasman, E. H., Petkovic, D., ... & Taubin, G, (1993, April), "QBIC project: querying images by content, using color, texture, and shape", In *IS&T/SPIE's Symposium on Electronic Imaging: Science and Technology*, pp. 173-187, International Society for Optics and Photonics.

[17] Tamura, H., Mori, S., & Yamawaki, T, (1978), "Textural features corresponding to visual perception", *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 8, no. 6, pp. 460-473.

[18] Iakovidou, C., Anagnostopoulos, N., Kapoutsis, A. C., Boutalis, Y., & Chatzichristofis, S. A., (2014, June), "Searching images with MPEG-7 (& mpeg-7-like) powered localized descriptors: the SIMPLE answer to effective content based image retrieval", In *Content-Based Multimedia Indexing (CBMI), 2014 12th International Workshop on*, pp. 1-6, IEEE.

BIOGRAPHIES OF AUTHORS

Nicole Tham Lay Mei graduated with a Bachelor's degree in Electronic Engineering from Universiti Teknologi Malaysia in 2017. Her work interests are primarily in the area of robotics and autonomous machines. She currently resides in Sabah, Malaysia

Syahmi Syahiran Bin Ahmad Ridzuan is currently doing his PhD at the Faculty of Electrical Engineering, Universiti Teknologi Malaysia. He obtained his Bachelor's degree in Electrical, Electronic and Automation Engineering from Université de Franche-Comté, Besançon and his Master's degree in Network, Telecommunication, Multimedia and Automation from Université de Poitiers. His research interest is primarily in image processing field.

Dr Zaid Omar is a senior lecturer at the Faculty of Electrical Engineering, Universiti Teknologi Malaysia. He obtained his PhD in Electrical Engineering from Imperial College London in 2012. His research interests include image processing, machine learning, and medical imaging.