The International Journal of Engineering and Science (IJES)
|| Volume || 6 || Issue || 6 || Pages || PP 49-58 || 2017 ||
ISSN (e): 2319 – 1813 ISSN (p): 2319 – 1805

Multimedia Databases: Performance Measure Benchmarking Model (PMBM) Framework

Abdulrehman A. Mohamed and Dr. Cyrus A. Makori, PhD
Mount Kenya University, School of Computing and Informatics, Department of Information Technology, P. O. Box 342-01000 Thika, Kenya
Mount Kenya University, School of Computing and Informatics, Department of Information Technology, P. O. Box 342-01000 Thika, Kenya

----------ABSTRACT----------
Multimedia database consist of media data types such as text, images, sound, and video, which are retrieved by image descriptors such as texture, shape, and color as their primary key. New technologies have numerous challenges, and multimedia database have its share of challenges. Some of these challenges are in Content-based Image Retrieval (CBIR) techniques, such as irregular performance measurements in motion, location and sketches, hence creating gap in multimedia database evaluation techniques. As a result of which, this paper was motivated to close this gap of lack of an effective and precise performance evaluation benchmarking measure. To address the irregular performance measure, the paper developed a performance measure benchmarking model (PMBM) framework using image descriptors in query by image content (QBIC). Moreover, the paper adopted the de Groot's empirical research cycle methodology by implementing the five stages methodology of: image preparation, query definition, data collection, data evaluation and framework building for the PMBM framework development. Furthermore, the paper conducted several experiments on texture, color and shape image datasets with respect to performance measures of accuracy, recall, precision and F-Measures on CBIR DB2 database. The results of these experiments were used to develop the PMBM framework and were analyzed by a statistical software SPSS version 20. Even more importantly, the paper developed a software evaluation tool in JAVA programming language from the PMBM framework to qualify and quantify its effectiveness to measure performance of the existing CBIR database systems. The results of this evaluation showed that the open source CBIR database FIRE was ranked as High with baseline score value (BSV) of 0.92 (92%) and IMG(Anaktisi) was ranked as Low, with BSV of 0.87 (87%) from BSV of 0.90 (90%) of IBM DB (benchmark system).

Keywords: *Content Based Image Retrieval (CBIR), Query by Image Content (QBIC), Image Descriptors, Multimedia Databases and Performance Measure Benchmarking Modal (PMBM) Framework*

Date of Submission: 14 May 2017 Date of Accepted: 09 June 2017

I. INTRODUCTION

1.1 Background Information

Multimedia databases store, manage and retrieve media data type such as images, sounds and videos, while traditional database store alphanumeric data types. The traditional database retrieve data through alphanumeric unique keys known as primary keys, while multimedia databases use content based image retrieval (CBIR) such as texture, color and shape as unique keys for retrieving data. CBIR is also known as query by image content (QBIC) and content-based visual information retrieval (CBVIR), is an application of computer vision techniques for searching and retrieving digital images in large databases [4]. In recent times, extensive research works have been done on CBIR systems, where the emergence of the first commercial CBIR system – an International Business Machine (IBM) QBIC system was developed as the benchmarking system. However, it is observed that the many systems use different standards of retrieval techniques such as motions, contours and sketches of various objects to retrieve data [7]. This is evident by the proposed evaluation frameworks by [1] and [4], where their shortcomings to address irregular performance measurement were also elusive. The above discussion infers biasness in multimedia database system evaluation, and hence creating a gap, which is the motivation of this paper to close this gap by developing a performance measure benchmarking model framework (PMBM) using image descriptors in QBIC.

II. PMBM FRAMEWORK DEVELOPMENT METHODOLOGY

2.1 Introduction

The PMBM framework development methodology was grounded on A.D. de Groot's empirical research cycle. The de Groot cycle was implemented as five stages methodology of the proposed PMBM model using image descriptors in QBIC, which is shown in Figure 1 below and defined as follows:

1. Image Preparation
2. Query Definition
3. Data Collection
4. Data Evaluation
5. Framework Building

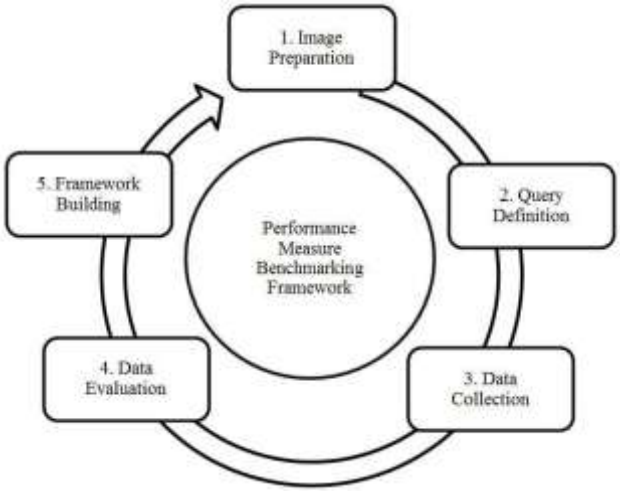

Figure 1: PMBF Model Development Methodology

2.2 Image Preparation

This paper assumed open source datasets for shape, color and texture images to conduct experiments for performance measure model development. It used the ETHZ open curve data for shape images, Uncompressed Color Image Database (UCID) for color images, and USC-SIPI (University of South California, Signal and Image Processing Institute) Image Database for texture images [9]

2.2.1 ETHZ Open Curve Dataset

It was reported by [2], that ETHZ open curve dataset is a set for testing entity class recognition algorithms. It comprises of 255 test images attributes and five diverse shapes constructed on apple logos, bottles, giraffes, mugs and swans. The shapes were generated by expert human judges consisting of 44 open curves for apple logos, 55 open curves for bottles, 91 open curves for giraffes, 48 open curves for mugs, and 32 open curves for swans. The dataset was used for evaluating the open curve matching.

2.2.2 Uncompressed Color Image Database - UCID

It was elaborated by [6], that the objective of the Uncompressed Color Image Database (UCID) was to proposal a benchmark dataset for image retrieval in an uncompressed format. Currently the database has over 1300 images which are predefined with corresponding models by human judge experts. The dataset is used for evaluation of compressed image retrieval in the 4th criterion algorithms and to explore the effect image compression on the performance of CBIR methods.

2.2.3 USC-SIPI Image Database

It was described by [8], that the USC-SIPI Image Database (University of South California, Signal and Image Processing Institute) is a suit of digital images comprising of brodatz textures, texture mosaics, etc. It is sustained by assisting researches in image processing, image analysis, and machine vision. The database consists of basic attributes of pictures such as 256x256 pixels, 512x512 pixels, or 1024x1024 pixels. All images are 8 bits/pixel for black and white images; while 24 bits/pixel for color images.

2.3 Query Definition

A suit of 15 queries were established to be used by both the experts' human judges and novice human judges. The experiments were conducted by DB2 (QBIC), a multimedia database engine as the baseline software, where 5 sets of queries for each descriptor of shape, color and texture were defined.

2.3.1 Shape Query

The ETHZ open curve dataset was used for shape query definition. A set of 5 queries consisting of 10 shape images of apple logos, bottles, giraffes, mugs and swans were defined for possible positive or negative retrieval by DB2 database.

2.3.2 Color Query

The Uncompressed Color Image Database (UCID) image format was used for color query definition. A set of 5 queries each consisting of 10 predefined images from the database with corresponding models by human judge experts were defined for possible positive or negative retrieval by DB2 database.

2.3.3 Texture Query

The USC-SIPI Image Database (University of South California, Signal and Image Processing Institute) was used for texture query definition. A set of 5 queries consisting of 10 digital images of brodatz textures and texture mosaics were defined for possible positive or negative retrieval by the DB2 database.

2.4 Data Collection

The proposed methodology for data collection of the paper was experimental model with experimental questionnaire forms to be filled**.** The paper used sets of different images from open source dataset for possible positive or negative retrieval by the DB2 (QBIC) multimedia database. The data collected was then subjected to SPSS for descriptive statistics analysis and correction measures tests of precision, recall, fall-out and f-measure.

2.5 Data Evaluation

It was elaborated by [5], that the evaluation performance measures of the quality of classification are built from a confusion matrix which records positively and negatively recognized examples for each class. The Table 1 below presents a confusion matrix for binary classification, where tp are true positive, fp – false positive, fn – false negative, and tn – true negative counts.

Table 1: Confusion matrix for binary classification [3]

Class/Recognized	**As Positive**	**As Negative**
Positive	Tp	Fn
Negative	Fp	tn

The following equations were used to calculate the four performance metrics of accuracy, precision, recall, and F-Measure during the experiments [10].

1.	Accuracy = {tp+tn}/{tp+fp+fn+tn}	Equation 1
2.	Precision (P) = {tp}/{tp+fp}	Equation 2
3.	Recall (r) = {tp}/{tp +fn}	Equation 3
4.	F-Measure = 2[{P*r}/{P+r}]	Equation 4

These performance measurements metrics were used for evaluation and validation of the proposed PMBM Framework development.

2.6 Framework Building

Data collected from human judge from both experimental form questionnaire and performance measures metrics were computed and the results were used for PMBM Framework development. Furthermore, the paper assumed the following techniques to implement the PMBM framework.

i. Established domain of images categorized as texture images, color images, and shape images, each with a set of 150 images.

ii. Established ground-truth human judges by randomly selecting 15 experts and novices from various university departments who have no prior knowledge of CBIR technique to perform visual retrieval of image performance measure
iii. Established set of 15 queries to be used by the 5 experts human judges during validation
iv. Computed and analyzed the results to build the PMBM framework.

III. PMBM DEVELOPMENT EXPERIMENTS

3.1 Introduction

The paper conducted three experiments using three datasets of: Color Image Dataset, Texture Image Dataset and Shape Image Dataset and were evaluated against four performance metrics of: Accuracy, Precision, Recall and F-Measure in order to develop the PMBM Framework

3.2 Experiment 1: Color Descriptors

The experiment was conducted by 30 respondents on the color image dataset to explore various retrieval scores. The scores were recorded in the questionnaire, and thereafter computed in SPSS to find the mean score for various retrieval scores as shown in Table 2. The results showed that out of 30 retrievals queries conducted; 7.40 (74.0%) were True Positive, 0.70 (7.0%) were False Negative, 1.00 (10.0%) were False Positive, and 0.90 (9.0%) were True Negative.

Table 2: Color Image Dataset Retrieval Mean

Average Retrival	Color Image Dataset
True Positive	7.40
False Negative	0.70
False Positive	1.00
True Negativie	0.90

These results inferred that the DB2 CBIR database using color descriptors correctly retrieved about three quarters of the query and only one quarter was incorrectly retrieved as shown in Figure 2. However, these results are not conclusive since there are other factors contributing to the performance of the retrieval process. Therefore, the paper computed four performance metrics of: Accuracy, Precision, Recall and F-Measure to justify these results. The results showed that; the accuracy = 83%, precision = 88%, recall = 91% and F-Measure = 89%. Finally, the computed mean score for color descriptor was 87.75%.

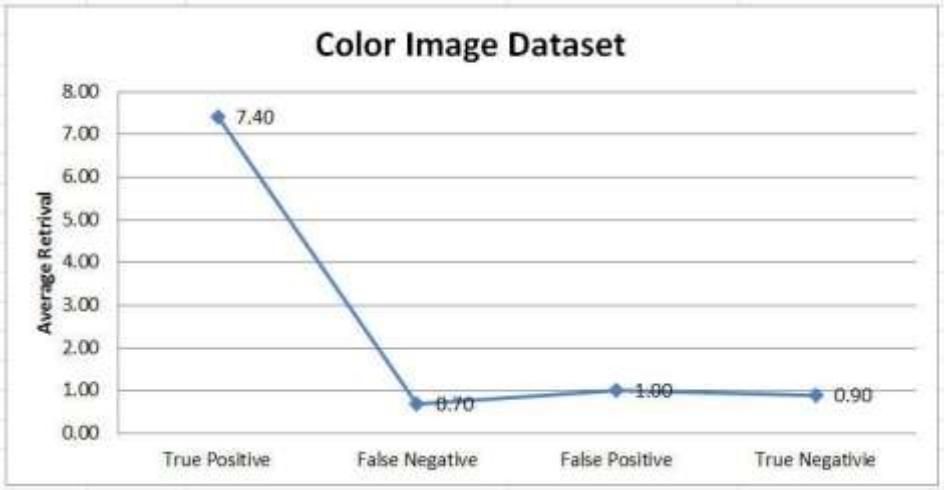

Figure 2: Color Image Dataset

3.3 Experiment 2: Texture Descriptors

The experiment was conducted by 30 respondents on the texture image dataset to explore various retrieval scores. The scores were recorded in the questionnaire, and thereafter computed in SPSS to find the mean score for various retrieval scores as shown in Table 3. The results showed that out of 30 retrievals queries conducted; 7.90 (79.0%) were True Positive, 0.50 (5.0%) were False Negative, 0.80 (8.0%) were False Positive, and 0.80 (8.0%) were True Negative.

Table 3: Texture Image Dataset Retrieval Mean

Average Retrival	Texture Image Dataset
True Positive	7.90
False Negative	0.50
False Positive	0.80
True Negativie	0.80

These results inferred that the DB2 CBIR database using texture descriptors correctly retrieved more than three quarters of the query and only less than a quarter was incorrectly retrieved as shown in Figure 3. However, these results are not conclusive since there are other factors contributing to the performance of the retrieval process. Therefore, the paper computed four performance metrics of: Accuracy, Precision, Recall and F-Measure to justify these results. The results showed that; the accuracy = 87%, precision = 90%, recall = 94% and F-Measure = 92%. Finally, the computed mean score for texture descriptor was 90.75%.

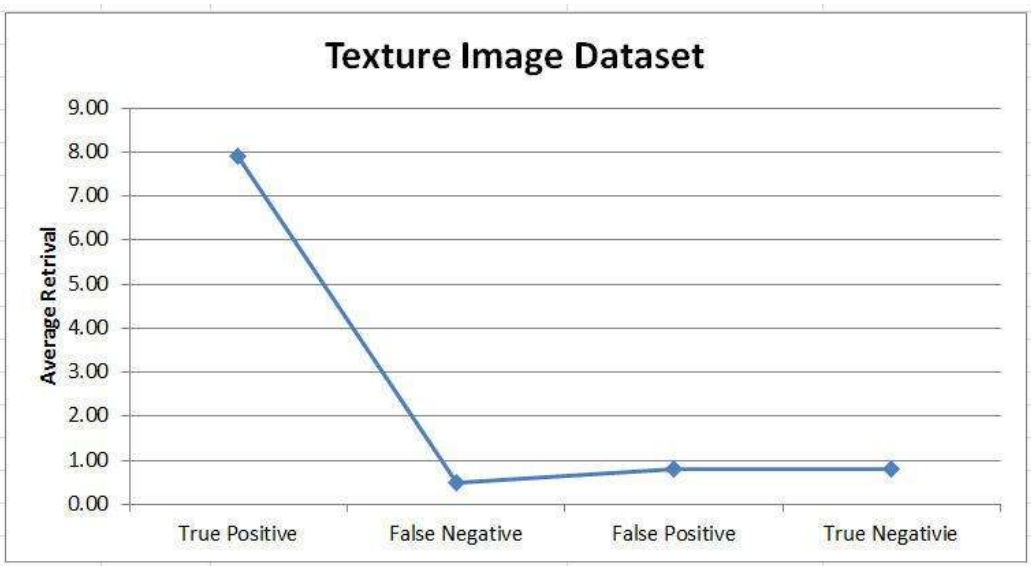

Figure 3: Texture Image Retrieval

3.4 Experiment 3: Shape Descriptors

The experiment was conducted by 30 respondents on the shape image dataset to explore various retrieval scores. The scores were recorded in the questionnaire, and thereafter computed in SPSS to find the mean score for various retrieval scores as shown in Table 4. The results showed that out of 30 retrievals queries conducted; 7.80 (78.0%) were True Positive, 0.40 (4.0%) were False Negative, 0.90 (90.0%) were False Positive, and 0.90 (90.0%) were True Negative.

Table 4: Shape Image Dataset Retrieval Mean

Average Retrival	Shape Image Dataset
True Positive	7.80
False Negative	0.40
False Positive	0.90
True Negativie	0.90

These results inferred that the DB2 CBIR database using shape descriptors correctly retrieved more than three quarters of the query and only less than a quarter was incorrectly retrieved as shown in Figure 4. However, these results are not conclusive since there are other factors contributing to the performance of the retrieval process. Therefore, the paper computed four performance metrics of: Accuracy, Precision, Recall and F-Measure to justify these results. The results showed that; the accuracy = 87%, precision = 90%, recall = 95% and F-Measure = 93%. Finally, the computed mean score for shape descriptor was 91.25%.

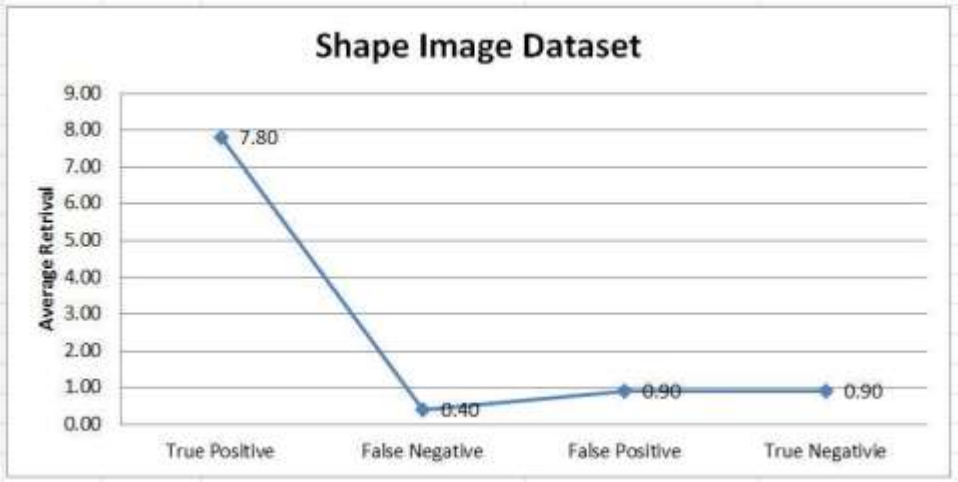

Figure 4: Shape Image Retrieval

IV. PMBM FRAMEWORK DEVELOPMENT

4.1 Introduction

The proposed PMBM Framework was generated from the previous three experiments on color descriptors, texture descriptors and shape descriptors with respect to the four performance metrics of: accuracy, precision, recall and f-measure. The proposed PMBM Framework was developed consisting of three phases: Content Descriptors, PMBM Evaluator Engine and Benchmarking Ranker

4.2 Content Descriptors

The proposed PMBM Framework defined the scope of the type of descriptors to be used as color descriptors, texture descriptors and shape descriptors. The content descriptors phase was generated by the outcome of PMBM development's stages 1 and stage 2 of the image preparation and query definition respectively as discussed at the previous section in Figure 1: PMBM development methodology.

4.3 PMBM Evaluator Engine

The PMBM evaluator engine was generated by the outcome of PMBM development methodology's stages 3 and 4 of data collection and data evaluation respectively as discussed at the previous section of Figure 1: PMBM development methodology. Moreover, the outcome of three experiments on color, texture, and shape descriptors with respect to four performance metrics of: accuracy, precision, recall and f-measure were used. Furthermore, questionnaires were used to collect data from various defined queries using the DB2 database.

4.4 Benchmarking Ranker

The benchmarking ranker defined a single mean score value called Benchmarking Score Value (BSV). This value was calculated by taking the mean score values of performance metrics of accuracy, precision, recall and f-measure for each dataset type of color image, texture image and shape image as shown in Table 5: Benchmarking Score Value.

Table 5: Benchmarking Score value

Dataset / Performance Metrics	Accuracy	Precision	Recall	F-Measure	Mean Score
Color Image Dataset	0.83	0.88	0.91	0.89	**0.88**
Texture Image Dataset	0.87	0.90	0.94	0.94	**0.91**
Shape Image Dataset	0.87	0.90	0.95	0.95	**0.92**
Mean Score	**0.86**	**0.89**	**0.93**	**0.93**	**0.90**

Banchmarking Score Value = 0.90

4.5 Proposed PMBM Framework

The proposed PMBM framework was generated by the previous framework development phases of Content Descriptors, PMBM Evaluator Engine, and Benchmarking Ranker. The proposed PMBM framework was defined visually as shown in Figure 5. The Content Descriptors is presented by the input of Color Descriptors, Texture Descriptors and Shape Descriptors. The PMBM Evaluator Engine is where the descriptors are process, and the Benchmarking Ranker, scores the output against the baseline Benchmarking Score Value (BSV) of 0.90 or 90% mean of the four performance metrics of: accuracy, precision, recall and f-measure.

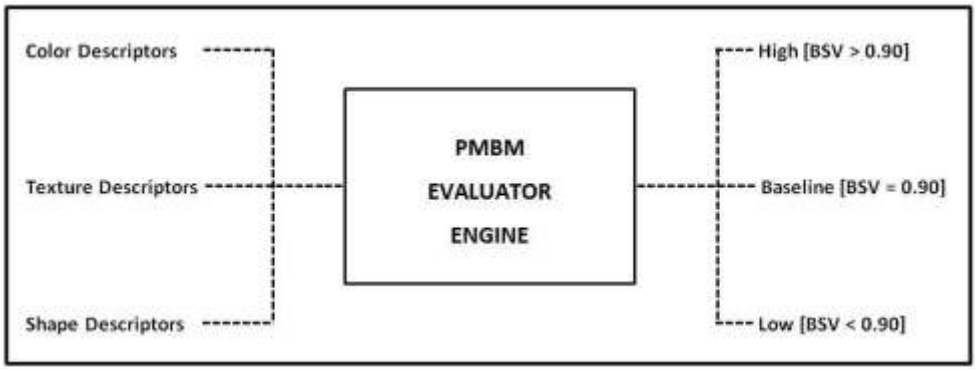

Figure 5: Proposed PMBM Framework

V. PMBM FRAMEWORK EVALUATION

5.1 PMBM Framework Software Evaluator Tool Development

In order to assess the effectiveness of the PMBM framework, the paper developed a PMBM framework software evaluator tool. The evaluator tool was used to conduct experiments to measure effectiveness of the retrieval performance of color, texture and shape descriptors of the existing CBIR databases. The PMBM Framework software evaluator tool was developed in object oriented programming of java in NetBeans environment. The software tool was generated from the proposed PMBM framework of figure 5. The tool had the following four interfaces of: Texture Evaluation, Color Evaluation, Shape Evaluation and Evaluator Engine.

5.2 Texture Evaluation Interface

The Texture Evaluation Interface's function is to assess the texture dataset. It allows the human evaluator to enter four confusion matrix values of: True Positive (tp), False Negative (fn), True Negative (tn) and False Positive (fp). These values are then used to automatically calculate the four performance measures of: Accuracy, Recall, Precision and F-Measure by a click of a button. Finally, a mean score of the four performance measure is derive and rank the results as high or low from the baseline of 0.90 or 90%. The screen shot of the interface is shown in Figure 6: Texture Evaluation Interface.

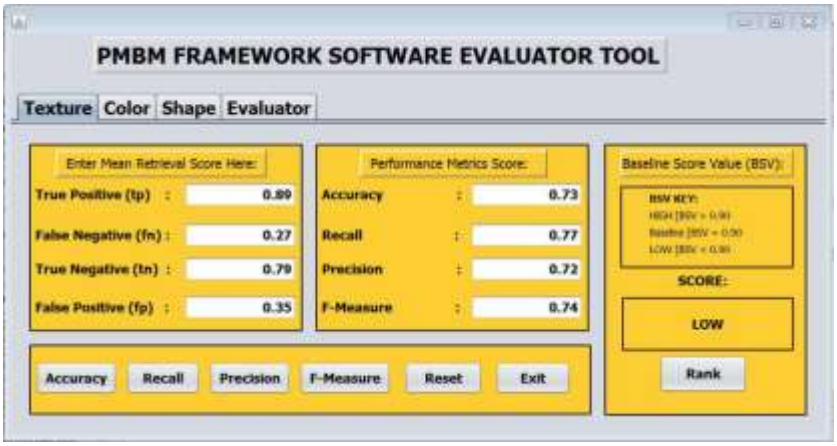

Figure 6: Texture Evaluation Interface

5.3 Color Evaluation Interface

The Color Evaluation Interface's function is to assess the color dataset. It allows the human evaluator to enter four confusion matrix values of: True Positive (tp), False Negative (fn), True Negative (tn) and False Positive (fp). These values are then used to automatically calculate the four performance measures of: Accuracy, Recall, Precision and F-Measure by a click of a button. Finally, a mean score of the four performance measure is derive and rank the results as high or low from the baseline of 0.90 or 90%. The screen shot of the interface is similar to texture evaluation interface of Figure 6 except the inputs are of color descriptors.

5.4 Shape Evaluation Interface

The Shape Evaluation Interface's function is to assess the shape dataset. It allows the human evaluator to enter four confusion matrix values of: True Positive (tp), False Negative (fn), True Negative (tn) and False Positive (fp). These values are then used to automatically calculate the four performance measures of: Accuracy, Recall, Precision and F-Measure by a click of a button. Finally, a mean score of the four performance measure is derive and rank the results as high or low from the baseline of 0.90 or 90%. The screen shot of the interface is similar to texture evaluation interface of Figure 6 except the inputs are of shape descriptors.

5.5 Evaluator Engine

The Evaluation Engine Interface's function is to assess the performance mean scores of texture, color and shape dataset. The evaluation engine calculates the three performances mean scores of: texture, color and shape by a click of a button. Finally, a mean score of the mean performance scores is derive and rank the results as high or low from the baseline of 0.90 or 90%. The screen shot of the interface is shown in Figure 7: Evaluation Engine Interface.

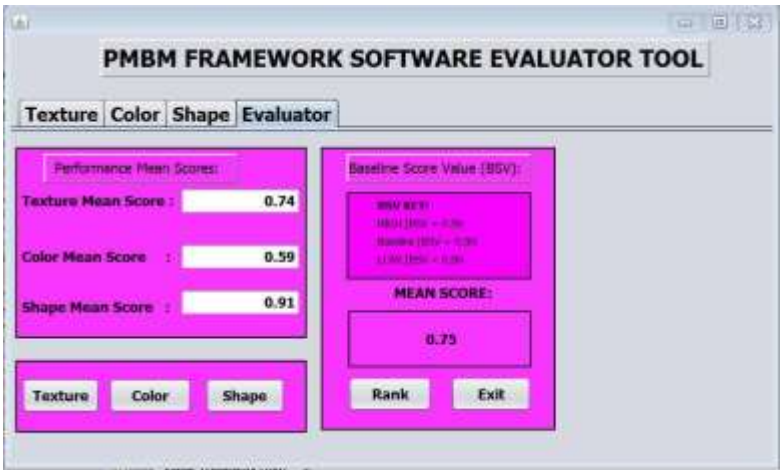

Figure 7: Evaluation Engine Interface

5.6 PMBM Framework Evaluation Experiments

In order to assess the effectiveness of the PMBM framework software tool, the paper conduct experiments to measure effectiveness of the retrieval performance of color, texture and shape descriptors of existing open source CBIR databases of: Img(Anaktisi) and FIRE. The experiments were conducted by 5 expert's respondents on the texture, color and shape image datasets to explore various performance mean-scores. The scores were recorded in the evaluation experiment form, and thereafter computed in PMBM Framework Evaluator Software Tool to find the mean-scores for various retrieval mean- scores

5.7 Evaluation Results

The results showed that out of 30 retrievals queries conducted for color image dataset; 0.88 (88.0%) mean-score for IBM DB2 (Baseline), 0.93 (93.0%) for FIRE and 0.85 (85.0%) for IMG(Anaktisi). The values for texture image dataset showed that; 0.91 (91.0%) mean-score for IBM DB2 (Baseline), 0.91 (91.0%) for FIRE and 0.89 (89.0%) for IMG(Anaktisi). Finally, the values for shape image dataset showed that; 0.82 (82.0%) mean-score for IBM DB2 (Baseline), 0.90 (90.0%) for FIRE and 0.79 (79.0%) for IMG(Anaktisi) as shown in Table 6.

Table 6: CBIR Systems Performance Evaluation

Dataset / CBIR Systems	IBM DB2	FIRE	IMG(Anaktisi)
Color Image Dataset	0.88	0.93	0.85
Texture Image Dataset	0.91	0.91	0.89
Shape Image Dataset	0.82	0.90	0.79
Mean Score	**0.90**	**0.92**	**0.87**
Rank Score	**Baseline**	**High**	**Low**

The mean-scores for the three CBIR were computed and results showed that; 0.90 (90.0%) mean-score for IBM DB2 ranked as Baseline, 0.92 (92.0%) for FIRE ranked as High from the Baseline and 0.87 (87.0%) for IMG(Anaktisi) ranked as Low from the Baseline. These results are shown in Figure 8: CBIR Systems Performance Evaluation. These results inferred that the PMBM Framework Software Tool evaluated the performance of FIRE as 92.0% and IMG(Anaktisi) as 87.% and ranked them as High and Low respectively from the Baseline value.

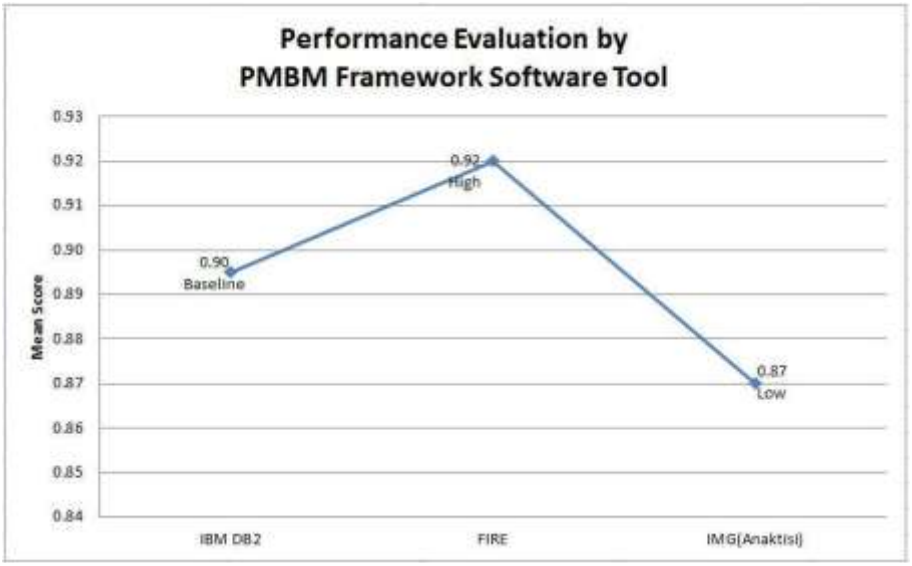

Figure 8: CBIR Systems Performance Evaluation

VI. RECOMMENDATION

Even though the findings of the paper showed that, the PMBM framework implementation was effective to measure the performance of existing CBIR systems using descriptors of: texture, color and shape, an enhanced version is recommended whereby more dynamic descriptors to be incorporated to increase effectiveness of the framework. Finally, the paper also recommended the extension-ability of the PMBM framework using other descriptors classification such as distributed based descriptors, differential descriptors and visual descriptors to give it a wider scope.

REFERENCE

[1]. Christophe, J. (2012). *Next Generation Search Engine: Advanced Models for Information Retrieval*. Hershey, PA: IGI Global. Retrieved from http://www.igi- global.com/book/next-generation-search-engines/59723

[2]. Cong, Y., Oliver, T., & Hassan, K. (2016). *Stripes-based Object Matching In Studies in Computational Intelligence.* Springer. Retrieved from http://congyang.de/downloads.html

[3]. Kohavi, R., & Provost, F. (1998). Applied Research in Machine Learning. In *Application of Machine Learning and the Knowledge Discovery Process* (Vol. 30). New York USA: Columbia University.

[4]. Lewandowski, D. (2012). A Framework for Evaluating the Retriev al Effectivenes s of Search Eng ines. *IGI Global*.

[5]. Powers, D. (2011). Powers, David M W (2011). "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation" (PDF). Journal of Machine Learning Technologies. 2 (1): 37–63. Journal of Machine Learning Technologies, 2(1), 37 –63.

[6]. Schaefer, G., & Stich, M. (2013). UCID - Uncompressed Colour Image Database. Retrieved from http://homepages.lboro.ac.uk/~cogs/datasets/ucid/ucid.html

[7]. Subitha, S., & Suhatha, S. (2013). Survey paper on various methods in content based inforfmation retrieval. *IMPACT: International Journal of Research in Engineering and Technology*, *1*(3), 109–120.

[8]. Weber, A. (2016). SIPI Image Database. Retrieved October 16, 2016, from http://sipi.usc.edu/database/

[9]. Wu, J., & M, J. (2010). CENTRIST: A Visual Descriptor for Scene Categorization. *IEEE Transactions on Pattern Analysis and Ma*, *33*(8), 1489 – 1501.

[10]. Zhu, M. (2004). Recall, Precision and Average Precision. . *IMPACT: International Journal of Research in Engineering and Technology*