Image Retrieval Evaluation Metric for Songket Motif

Nadiah Yusof[1], Amirah Ismail[2], Nazatul Aini Abd Majid[3], Zurina Muda[4]
Faculty of Computer Science and Multimedia, University Poly-Tech Malaysia, Cheras Malaysia[1]
Faculty of Islamic Technology, Sultan Sharif Ali Islamic University, Bandar Seri Begawan, Brunei[2]
Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Bangi, Malaysia[3, 4]

Abstract*—Songket is a fine art heritage specializing in promoting the unique features of Malay identity. Past studies have shown that hundreds of Songket motifs had been produced, but unfortunately, most were not stored digitally. However, the digital collection of image data and determining its ground truth data should be given attention. This paper focuses on an evaluation metric to retrieve image Songket motifs. The initial label for each class of images in the database is ground truth data. The activity of determining the ground truth data involves two research objectives that have been discussed, namely identifying the ground truth data set of Songket Motifs involving Activity One, to obtain two ground truth data sets, precisely the training data set and the test data set involving six categories, to be specific 'Flora', 'Fauna', 'Nature', 'Cosmos', 'Food' and 'Calligraphy'. This phase test was carried out through a survey using a qualitative method, which is participatory by 15 respondents who have classified 413 specific motif images into 56 Songket motifs categories that refer to the six prominent motifs. Meanwhile, Activity Two is a validation-classification test of ground truth data sets by three experts to equate the selection of general and expert respondents to obtain training data sets for testing purposes involving six categories. After rearranging, only 50 ground truth Specific Motifs have been selected. Accordingly, the relationship coefficient correlation method is also implemented to see the relationship between two data through a statistical evaluation angle. In addition, precision and recall methods are also used to obtain precision and recall values for each ground truth data, and the F-measurement method is used to make a single evaluation. The F-measurement result for each category 'Flora': 26.7 – 100 (20 ID-Category), 'Fauna': 35.3 – 100 (6 ID-Category), 'Nature': 30.8 – 100 (5 ID-Category), 'Cosmos': 53.3 – 100 (7 ID-Category), and 'Motif': 47.6 – 100 (9 ID-Category). Using ground truth data enables image retrieval research to conduct unbiased system testing and evaluation.

Keywords*—Heritage; songket motifs; songket motifs retrieval; ground truth data

I. INTRODUCTION

Malaysia has a rich cultural and artistic heritage that can be divided into two categories: tangible cultural heritage and intangible cultural heritage. Tangible cultural heritage refers to physical objects that can be seen and touched, such as tombs and tombstones. On the other hand, intangible cultural heritage is based on knowledge and expertise that is passed down through oral traditions, cultural values, language, literature, and textiles, such as the traditional Songket [1]. Researchers have explored the use of the Content-Based Image Retrieval (CBIR) technique to study the Songket in intangible cultural heritage. Some of the research focuses on determining the basic geometric structure [2], decomposing the Songket Motif texture to match motif similarity [3], and refining the boundaries of Songket Motif images using canny edge detection techniques [4].

The study conducted by [2] and [3] focuses on the image retrieval model, which includes three main stages: pre-processing, processing, and post-processing. In this model, the pre-processing stage is particularly important as it involves the preparation of datasets, whether they are original or synthetic. Obtaining original data is a more challenging task as it requires the collection of images using a camera lens and converting them into digital format through editing. Therefore, researchers must put in extra effort to locate individuals or groups who actively use the data, such as Songket entrepreneurs and researchers.

According to the viewpoint put forth by [5], most of the information regarding Songket is still reliant on traditional methods, such as documentation and bookkeeping. Therefore, to conduct this study, it is imperative to undertake measures to digitize the information related to Songket. In the meantime, synthetic data can be obtained by downloading it from websites or producing it oneself, such as through paintings or anime, among others [6].

The image pre-processing stage involves two main parts: the determination of the ground truth data and the development of the image segmentation system [7]. Ground truth data is a set of master images that represent each image class in the database [8]. Determining the ground truth data helps in testing the image retrieval system more effectively. This process is carried out either based on human observation or by the system itself [9, 10]. The initial label for each class of images in the database is the ground truth data [8]. The use of ground truth data enables unbiased system testing and evaluation in image retrieval research [9]. Ground truth data can be established either by the system or through human observation [2, 3].

The image pre-processing stage was a significant focus of the study when Songket Motif's image retrieval model was introduced in references [2] and [3]. Image pre-processing involves determining the ground truth data and image segmentation. However, the determination of Songket Motif's ground truth data has not received enough attention, and most researchers perform testing by randomly selecting the Songket Motif. The goal of determining the ground truth data is to help in the fair processing of the image retrieval system's searching and browsing functionality.

The following text discusses the concept of ground truth data sets in research and their importance in producing accurate results. Specifically, the paper aims to determine ground truth data sets for Songket Motifs. To achieve this goal, both

qualitative and quantitative methodologies are employed. Earlier studies, including [3, 8, 2], have also identified ground truth data sets for their research. However, this paper focuses solely on the Songket Motifs and their ground truth data sets. It is worth noting that the ground truth data set is a collection of data that has been established as authentic by either the user or the system. Its purpose is to ensure that the results obtained are reliable and trustworthy.

This paper is divided into five sections. The first section is an introduction and overview that discusses ground truth data determination along with any issues, objectives, and research methodologies. In the literature review section, key points from the literature are covered in detail and with a critical discussion. The third section focuses on the research methodology used to gather precise information about Songket Motifs. The fourth section deals with selecting experts and general users for the research. The fifth section contains the results and discussion, and the conclusion and acknowledgment follow afterward.

II. Literature Review

The main images in computer vision typically serve as a group representation for each image. They contain ground truth data, which is meant to show the essential components of each object in the image collection [9]. The primary label outlines the critical component of the object and its location in the image. As stated in [13], humans or systems must decide how to classify objects. Some commonly asked questions regarding identifying ground truth data include:

1) How ground truth data has been detected?
2) If using the system, is the first image ground truth data?
3) How has the system detected ground truth data?
4) If humans have detected ground truth data? What is the characteristic of expertise to be able to confirm it?

A study conducted by [9] utilized 80 general and eight expert participants to gather ground data based on human perception. The general participants were asked to identify relevant and commonly used English labels while searching and browsing landscape drawings. Meanwhile, the expert participants were tasked to provide the appropriate keyword for the ground truth data that could accurately describe each color notion evaluated in the landscape painting.

Furthermore, according to [10], establishing ground truth data by human observation is a challenging endeavor that takes much time. Different definitions and meanings exist of human perception and response to an object in an image [9]. To ensure that the ground truth data used is balanced to represent each category of variable data, however, human observation and the ability to determine the ground truth data are necessary [12].

Accordingly, geometrical data faces issues describing complex data structures and is even challenging to detail geometrically by natural data systems [10]. Thus, human observation of the shape structure is found to help facilitate the process of determining the ground truth data. In addition, the process of determining ground truth data can also determine the selection of labels or words of ground truth data required to implement tests at the testing stage of the proposed model and system [14]. This aims to enable the implementation of the evaluation and comparison of image retrieval results to take place somewhat and equitably for testing the image retrieval system.

III. Method

This phase involves two levels of activity. Activity One aims to identify the ground truth data set of Songket Motifs. Meanwhile, Activity Two is a ground truth data set classification process involving Songket Motif's images.

A. Identify Songket Motif's Ground Truth Data Set

This activity takes the form of a participatory test to determine the Songket Motif's data set in obtaining the ground truth data set, which is a participatory survey of human observation of form, where respondents must classify 413 images into six Songket Motif's categories, namely 'Flora', 'Fauna', 'Nature', 'Cosmos', 'Food' and 'Calligraphy'.

The respondents consisted of 15 users from two different backgrounds; Culture (directly involved in Songket research) and Information Technology. Respondents from the Cultural field are needed because they are more adept at identifying the structure of the ground truth data involving Songket Motifs. Meanwhile, respondents in the Information Technology field are more sensitive to the similarity of the shapes they see. Table I shows the background of the respondents.

TABLE I. Background of the Respondents

Occupations	Gender	Field
Student (1)	Female (8)	Culture (Songket Weaving) (11)
Work (14)	Male (7)	Information Technology (4)

Respondents were selected based on their knowledge of the form and motifs of Songket to conduct the test. Eight out of 15 respondents (53.3%) were female, while seven (46.6%) were male. All respondents were between 25 and 60 years old. The study aims to obtain a collective initial result to continue the system testing by classifying Songket Motif's data sets to ground truth data sets.

1) Research activity design: Based on Table II, this section is a further detail involving two Activities.

TABLE II. Aim and Method

Label	Research	Aim	Method
Activity One	Identifying the ground truth data set of Songket Motif's	Gets two sets of ground truth data, namely the set training data and test data involving six Songket Motif's categories ('Flora', 'Fauna', 'Nature', 'Cosmos', 'Food', and 'Calligraphy')	Qualitative method survey: Participatory. 15 respondents categorized 413 images involving six Songket Motif categories
Activity Two	Validation of ground truth data set classification	Validating a training data set for testing and evaluation purposes involves five categories ('Flora', 'Fauna', 'Nature', 'Cosmos', and 'Motifs')	Validation and use of training data sets (50 Songket Motif's). Comparison of F-measurement for five categories.

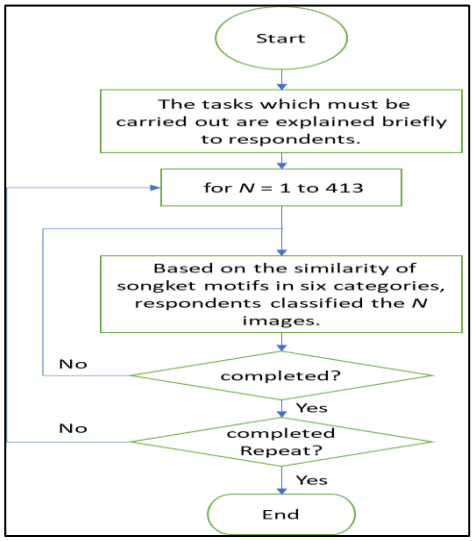


Fig. 1. Flow chart of activities identifying the ground truth data set of Songket Motif's.

Identifying the ground truth data set of Songket Motifs refers to the flow chart displayed in Fig. 1. Each respondent was given a brief description of the task that needed to be carried out. After that, respondents were required to classify 413 Songket Motif images into six Songket Motif categories.

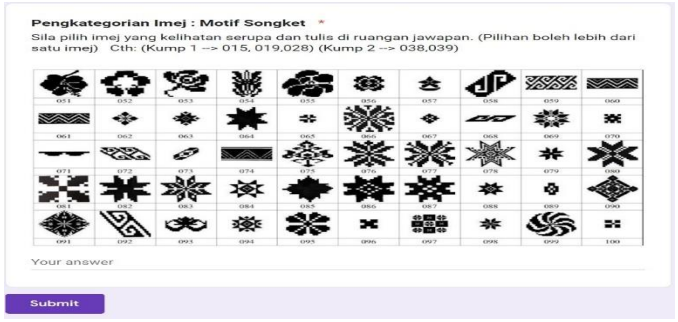
Activities are conducted using the same computer for each respondent in each session under the researcher's supervision. This aims to make it easier for respondents if there are questions on unclear matters. The specific Motif's image sizes are varied, and all specific motifs are labeled in numerical order from 001 to 413.

Respondents were free to choose shapes that were considered similar from a selection of specific Motif's image categories. Respondents must provide feedback for each page before the system allows the display of the next screen. This aims to ensure respondents provide feedback for all screen displays of the questions.

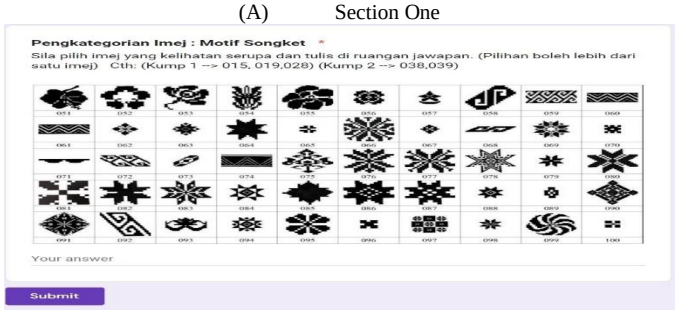
The test results show that the respondent's time to classify the similarity of 413 specific Motif images was more than 60 minutes, and the respondent's concentration decreased over time. Therefore, the number of specific Motif images was reduced to 200, and this test was divided into two sessions to ensure that the respondent's focus could be maintained as well as possible to obtain valid results.

All responses from respondents are recorded directly into an online file. Updates to this file are done automatically whenever there is feedback from respondents.

Fig. 2 displays an example of a participatory survey question for the two earliest examples of specific Motifs for sections One and Two with image IDs labeled 001-100. Respondents must choose to match the similarity of images that are seen to have similar shapes. If the respondent does not provide feedback, the respondent cannot press the "Submit" button because each question is marked * as a "Required question." Respondents' choice of matching answers can exceed more than one image for each group.



(A) Section One



(B) Section Two

Fig. 2. An example of a participatory survey question on the classification of Songket Motif's.

2) Validation of classification of ground truth data sets: Validation technique activities are carried out to validate or cross-check the ground truth data set of Songket Motifs identified in the previous phase. Table III is the background of the respondents involved in verifying the ground truth data arrangement of Songket Motifs according to different categories. The respondents consisted of three experts in the field of Culture (studying Songket and Batik) and Information Technology (conducting research involving image retrieval technology and Songket motifs), aged between 35 and 46 years. Because the respondents are experts, it is used as purposive sampling in Activity Two, that is, verification - classification of ground truth data sets.

This phase involves human evaluation of the appearance of specific motifs. The activities of this section are labeled as Activity Two and are carried out to achieve the objective of this study. Activity Two involved validation in identifying similar features of specific Motif shapes and evaluation of the confusion matrix. The two research questions for this phase are listed as follows to ensure that this activity achieves results.

TABLE III. RESPONDENT BACKGROUND

Respondent	Gender	Age	Field
Respondent #01	F	46 Years Old	Senior Lecturer in Information Technology & Songket
Respondent #02	F	37 Years Old	Lecturer in Heritage (Songket Researcher)
Respondent #03	F	35 Years Old	Researcher (Information Technology & Songket)

1) Is there a difference in the observation of specific Motifs between expert users and general users?

2) What are the results of the validation?

a) Validation test design – songket Motif's classification: Songket motifs that general respondents have classified are 56 specific Songket motifs rearranged as in Table IV.

TABLE IV. SPECIFIC MOTIFS THAT THE GENERAL RESPONDENT HAS REARRANGED

				
Biji Tamar	Cermai	Nanas	Nona	Buah Delima
				
Cendawan	Pucuk	Pucuk Rebung	Sulur Kacang	Bunga Bintang
				
Bunga Bebaling	Bunga Ros Meletak	Kembang Semangkuk	Kemunting Cina	Melur
				
Pecah Empat	Tampuk Kecupu	Tampuk Kesemak	Tampuk Manggis	Tanjung
				
Bunga Kekwa	Bunga Flora	Bunga Kemuncup	Daun	Daun Keladi
				
Daun Lalang	Cengkeh	Peria	Burung Merak	Kupu-kupu
				
Siku Keluang	Gigi Yu	Rantai Unduk Laut	Unduk Laut	Sinar Matahari
				
Sirik	Bogan	Bunga Mahkota	Bunga Mangga	Pecah Lapan
Bunga Tudung Saji	Petak Catur	Pitis	Semangat	Mahkota Raja
Keris	Cabit	Awan Larat	Kendik Tali	Ombak
Kaligrafi	Madu Manis	Potong Wajik	Tepung Talam	Seri Kaya
Motif				

The results of the selection of general respondents have been rearranged and make the total number of Songket Motif's used 413 involving 56 specific Songket Motif's, and each has been classified similarly to 413 specific Motif, and the number of specific motifs classified is between two to 52 specific Songket Motif's images.

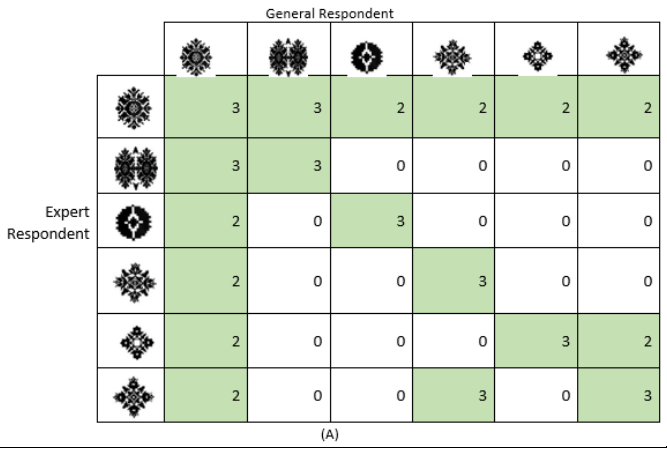
All specific Motif are printed on A4 size paper, and the size measurement for each specific Motif's image is 2.7cm, intended to help expert respondents see each specific Motif's image well. Meanwhile, activities are done individually in the same or different rooms. Before undergoing Activity Two, respondent information such as name, age, and gender is recorded. Then, the purpose of Activity Two and what the respondent needs to do are explained to the respondent. After the description related to the task was clarified, the expert respondents were shown 56 images of specific motifs, and the classification results that the general respondents had carried out are shown in Fig. 3.

Fig. 3. Examples of specific Motif's image categories and classification results that general respondents have implemented.

After expert respondents carry out the reclassification test, the ground truth data classification of the Songket Motifs is evaluated by calculating the confusion matrix formula, that is, using the technique formula of precision, recall, and F-measurement. The result of the selection of general respondents is the image in the database and all images achieved. Meanwhile, the selection of expert respondents is a relevant image. Accordingly, Fig. 4 shows that the column marked with the number two or three is a Songket Motif considered relevant by expert respondents. At the same time, the space marked with a blank (0) is a Songket Motif that is considered irrelevant. In Fig. 4(A), there are expert respondents (actual class) and general respondents (predicted class).

Next, Fig. 4(B) refers to the specific motifs in the first column displaying the highest F-measurement value, while the F-measurement value for the other specific Motif is lower. Therefore, for the specific Motif of 'Kembang Semangkuk' only one specific motif was selected as the ground truth data of the Songket Motifs for testing purposes.

A detailed description of this section is explained in the next section, which is the section on verifying the ground truth data analysis, and precisely in the next section.



(A)

Total	Precision %	Recall %	F Measurement %
6	100	100	100
2	33.3	100	50
2	33.3	100	50.
2	28.6	100	44.4
3	50	100	66.7
3	50	100	66.7

(B)

Fig. 4. Evaluation of the confusion matrix for the ground truth data Songket Motif's class 'Kembang Semangkuk' (A) Mark the confusion matrix (B) Calculation.

IV. Results and Discussion

Through the reclassification activity of ground truth data selection by expert respondents, 364 (88.14%) Songket Motifs were classified into 50 categories. The 50 classified categories refer to the structural similarity of specific motifs considered relevant by respondents. At the same time, another 49 (11.9%) motifs could not be classified because there was no matching of similar characteristics for different Motifs the respondent could implement. This is because the specific motifs stand alone, for example, the specific Motif 'Ayam Jantan', although there are other specific motifs that can be classified in the same class (birds), namely the specific Motif 'Itik Serati'. However, the similarity of the specific motifs is seen as the difference is too far. Therefore, such images cannot be considered in the same class. Table V. shows the name and amount for each specific Motif reclassified through expert respondent verification and confusion matrix evaluation. The results of this evaluation show that 50 categories of ground truth Songket Motif's data were obtained.

The ground truth data analysis process of Songket motifs was carried out after the participatory survey was completed; 56 categories of Songket Motifs classified by general respondents were rearranged by expert respondents into 50 categories of actual Songket Motifs data. This is because general respondents randomly arranged each specific Motif into 56 categories. Meantime, the expert respondents rearranged the motives based on the appropriateness of the actual structure of the Motifs according to the category. Therefore, the critical analysis required for this activity is the frequency distribution of 50 similar appearance categories for each of 413 specific Motifs and the percentage of agreement between respondents who agree to set a similar appearance for each specific Motif. Table VI is an example of the frequency distribution for 20 Songket Motif images taken from the test involving the first and last 10 results.

The Songket Motifs involved are the first ten Songket Motifs labeled ID-Image between 001 to 010 and the last 10 Songket Motifs labeled ID-Image between 404 to 413. ID-Image represents the identification of each specific Motif in the database. In contrast, the ID-Category represents the identification of the identity of each image of the ground truth data of the Songket Motifs that has been determined. Based on the frequency distribution, some information can be deciphered, such as ID-Image 002, categorized as a similar match for ID-Category 001 by 15 respondents. While referring to ID-Image 412-413, eight respondents stated that it was similar to ID-Category 013. This test found that the higher the number of respondents who made a mark, the more similar a form of ID-Category was to the marked ID-Image.

TABLE V. A Group of Songket motifs that Expert Respondents have Classified

Bil	Motif's	Total		Bil	Motif's	Total
001	Anggur	8		026	Motif	6
002	Awan Larat	5		027	Motif	5
003	Bunga Bintang	12		028	Motif	8
004	Bunga Bintang	11		029	Motif	9
005	Bogan	5		030	Motif	8
006	Cabit	6		031	Patah	3
007	Cengkeh	15		032	Pecah Empat	4
008	Cengkeh	9		033	Pecah Empat	5
009	Cengkeh	19		034	Pecah Empat	5
010	Cengkeh	10		035	Pagar	3
011	Cengkeh	8		036	Peria	8
012	Daun	7		037	Pitis	3
013	Kepala Lalat	10		038	Rantai Pecah Lapan	8
014	Motif	9		039	Pucuk Rebung Gigi Yu	10
015	Motif	7		040	Rantai Unduk Laut	6
016	Pergunungan	4		041	Semangat	9
017	Pergunungan	6		042	Siku Keluang	3
018	Pergunungan	5		043	Siku Keluang	11
019	Kembang semangkuk	6		044	Sinar Matahari	6
020	Keris	2		045	Kendik Tali	4
021	Melur	11		046	Kendik Sirik	5
022	Melur	16		047	Tanjung	7
023	Motif	2		048	Tanjung	7
024	Motif	3		049	Kendik Tali	8
025	Melur	7		050	Unduk Laut	10

TABLE VI. FREQUENCY DISTRIBUTION OF RESPONDENTS ACCORDING TO ID-IMAGE AND ID-CATEGORY FOR THE FIRST 10 SONGKET MOTIFS (ID-IMAGE 001-010) AND THE LAST 10 SONGKET MOTIFS (ID-IMAGE 404-413)

ID-Image / ID-Category	001	002	003	004	005	006	007	008	009	010	...	404	405	406	407	408	409	410	411	412	413
001		15																			
002			15	15	15	15	15														
003																					
004																					
005										10							15				
006	10																				
007																		15			
008																					
009																					
010																					
011																					
012		15																			
013											.....									8	8
014																					
015																		15			
038																			15		
039												8	8	8							
040																			11		
041																					
042																					
043																					
044																					
045																				15	
046																					15
047																					
048																					
049																					
050																					

The respondent records no information for 20 ID-Image (016-037)

The minimum amount considered for the similarity percentage is eight respondents, equal to 53.3%, more than 50% of the respondents' agreement. The yellow column in Table VI refers to the ID-Category between 016 and 037, which means there is no matching information of specific Motif's image similarity done by the respondents for the ID-Category.

The highest frequency for each specific Motif shows the percentage of respondents who agree to classify the specific Motifs into a particular category. Some specific motifs get a higher or lower percentage of respondents' approval than others. Table VII shows the average value of the respondent's agreement evaluated through the results of the frequency distribution test, so the overall average value of the respondent's agreement amounts to 79.2%, categorized as ID-Category 001. However, the percentage of agreement of each ID-Category is different. For example, ID-Category 002, 020, and 037 have the highest agreement, 100%. All 15 respondents agreed that the specific Motif's match was similar. As for ID-image 041, the percentage of respondents' approval was only recorded at 63% for nine images, which is the lowest average value compared to other ID categories. The division of ID-Category 041 involves several parts, namely one image agreed upon by 15 respondents that is similar. In addition, two other images were agreed by 11 respondents that are like ID-Category 041. Simultaneously, there are six more Songket Motif images, with only eight respondents agreeing that they are similar, and the total number of respondents who agree is 85 out of 135. The number of similar images for ID-Category 041 is as many as nine specific motifs.

TABLE VII. SUMMARY OF THE AVERAGE VALUE OF THE RESPONDENTS' AGREEMENT EVALUATED THROUGH THE RESULTS OF THE FREQUENCY DISTRIBUTION TEST

ID-Category	Similar Image Matches for each ID-Category	Total Number of Respondents Consent	Actual Total Mark	Average Value of Respondents' Agreement (%)
001	8	95	120	79.2 %
002	5	75	75	100 %
003	12	130	180	72.2 %
004	11	156	165	94.5 %
005	5	57	75	76 %
006	6	69	90	76.7 %
007	15	189	225	84 %
008	9	106	135	78.5 %
009	19	231	285	81 %
010	10	128	150	85.3 %
011	8	91	120	75.8 %
012	7	80	105	76.2 %
013	10	109	150	72.7 %
014	9	107	135	79.3 %
015	7	94	105	89.5 %
016	4	46	60	76.7 %
017	6	69	90	76.7 %
018	5	61	75	81.3 %
019	6	70	90	77.8 %
020	2	30	30	100 %
021	11	130	165	78.8 %
022	16	161	240	67.1 %
023	2	28	30	93.3 %
024	3	41	45	91 %
025	7	92	105	87.6 %
026	6	62	90	68.9 %
027	5	70	75	93.3 %
028	8	104	120	86.7 %
029	9	101	135	74.8 %
030	8	85	120	70.8 %
031	3	34	45	75.6 %
032	4	49	60	81.7 %
033	5	66	75	88 %
034	5	58	75	77.3 %
035	3	38	45	84.4 %
036	8	101	120	84.2 %
037	3	45	45	100 %
038	8	88	120	73.3 %
039	10	97	150	64.7 %
040	6	65	90	72.2 %
041	9	85	135	63 %
042	3	34	45	75.6 %
043	11	108	165	65.5 %
044	6	70	90	77.8 %
045	4	53	60	88.3 %
046	5	54	75	72 %
047	7	73	105	69.5 %
048	7	71	105	67.6 %
049	8	95	120	79 %
050	10	125	150	83.3 %

Accordingly, the correlation between general and expert respondents must be assessed using statistical methods. This aims to see the relationship between the two data that have been used.

A. Correlation between the Selection of General Respondents and Expert Respondents

To identify the relationship between the selection of general and expert respondents, the implementation of the relationship correlation test is one of the statistical methods often used [15]. The correlation test of the relationship coefficient using the 'Pearson Correlation method can be carried out, but it is necessary to go through a standard data determination test to see the P-Value *(Asymp. Sig(2 Tailed))* between the data of general respondents and expert respondents [16]. Accordingly, Table VIII shows the results of the selection of general respondents and expert respondents.

Based on Table VIII, the frequency distribution data test was carried out, and the results are displayed in Figure 5. According to [11], the P-value *(Asymp. Sig(2 Tailed))* obtained through the Kolmogorov Smirnov test needs to exceed 0.5% to enable the 'Pearson Correlation' method to be implemented. Thus, the evaluation results show that the data distribution for general and expert respondents is above 0.5%, which is general .450 and expert .528. Therefore, the results of the frequency distribution for this study are in a normal state. Therefore, the 'Pearson Correlation' method can be used for this study.

```
NPAR TEST
        /KOLMOGOROV-SMIRNOV (NORMAL) =  Umum Pakar.
```

One-Sample Kolmogorov-Smirnov Test

		Umum	Pakar
N		50	50
Normal Parameters	Mean	85.52	18.68
	Std. Deviation	40.64	8.68
Most Extreme Differences	Absolute	.12	.11
	Positive	.12	.11
	Negative	-.08	-.07
Kolmogorov-Smirnov Z		.86	.81
Asymp. Sig. (2-tailed)		.450	.528

Fig. 5. The test results used one sample Kolmogorov – Smirnov for the entire general election, and respondents showed a P-value exceeding 0.05%.

Referring to Fig. 5, the implementation of the relationship coefficient correlation evaluation has used the 'Pearson Correlation formula, which considers the correlation coefficient value *(r)*, the *N* value, the statistical *T* value, the *DF* value, and the *P* value. The *DF* value is the *'degree of freedom'* that indicates the number of independent values that can change in the analysis without violating any constraint 11]. Whereas the value of N is paired ranks' value the value of *N* I is required to calculate the statistical T formula. The T-Statistic formula is like1).

$$t = \frac{r * \sqrt{n-2}}{\sqrt{1-r^2}} \quad (1)$$

The correlation value of a good relationship should be between *-1 < r < +1* [12]. Accordingly, Fig. 6 shows the results of the evaluation that was carried out, obtaining a value of 0.991a. Thus, the results of the relationship coefficient correlation for this study show that there is a positive relationship coefficient correlation between general *(Umum)* respondents and expert *(Pakar)* respondents.

TABLE VIII. RESULTS OF THE SELECTION OF GENERAL AND EXPERT RESPONDENTS

ID-Category	General *(Umum)*	Expert *(Pakar)*
001	95	20
002	75	15
003	130	29
004	156	32
005	57	13
006	69	15
007	189	38
008	106	24
009	231	48
010	128	27
011	91	20
012	80	17
013	109	25
014	107	24
015	94	19
016	46	10
017	69	16
018	61	13
019	70	14
020	30	6
021	130	28
022	161	38
023	28	6
024	41	9
025	92	19
026	62	14
027	70	14
028	104	21
029	101	23
030	85	20
031	34	8
032	49	11
033	66	14
034	58	13
035	38	8
036	101	21
037	45	9
038	88	19
039	97	23
040	65	15
041	85	21
042	34	8
043	108	26
044	70	16
045	53	11
046	54	12
047	73	17
048	71	16
049	95	21
050	125	28

```
CORRELATION
        /VARIABLES =  Umum Pakar
        /PRINT = TWOTAIL NOSIG.
```

Correlations

		Umum	Pakar
Umum	Pearson Correlation	1.000	.991[a]
	Sig. (2-tailed)		.000
	N	50	50
Pakar	Pearson Correlation	.991[a]	1.000
	Sig. (2-tailed)	.000	
	N	50	50

a. Significant at .05 level

Fig. 6. The result of the Pearson Correlation value for general respondents is 1.00. while expert respondents are 0.991a.

Based on a comprehensive correlation coefficient test of fifty data between general and expert respondents, it was found that ID-Category 016, 018, 026, 035, 045, and 046 obtained the highest correlation coefficient result, which is one (1), the ID-Category representing specific Motif's 'Pergunungan1', 'Pergunungan2', ' Motif's ', 'Pagar', 'Kendik Tali', and 'Kendik Sirik'. Meanwhile, the other ID-Category obtained lower correlation coefficient results ranging from 0.65 to 0.95.

Furthermore, the ground truth data analysis part of the Songket Motif is a statistical information extract on matching similar images carried out by 15 general respondents and has been rearranged by expert respondents. After that, the similarity matching results were evaluated using confusion matrix analysis.

B. Confusion Matrix Analysis of the Ground Truth Data of the Songket Motif's

Songket motifs comprise six main categories: 'Flora', 'Fauna', 'Nature', 'Cosmos', 'Food', and 'Calligraphy' Motifs. Three prominent motifs have split motifs; the main motif 'Flora' is divided into six split motifs, namely 'Fruits', 'Trees', 'Flowers', 'Leaves', 'Spices', and 'Vegetables'. The main motif of the fauna is divided into two fractional motifs, namely 'Land' and 'Sea'. As for the main motif, 'Nature' is divided into two Motifs: ' Natural' and 'Things'. The main Motif of 'Cosmos' consists of three specific motifs; ' Awan Larat', 'Kendik Tali', and 'Ombak'; the total number of Songket Motif images for this category is 15 Songket Motif's images. However, the categories 'Food' and 'Calligraphy' were not chosen as the ground truth data of Songket Motifs because these motifs are rarely operated in any pattern commercially.

1) Songket Motif's category: The selection of Songket Motif's classification was done by respondents through the Activity One test, which identified the ground truth data set of Songket Motif. The re-determination of the ground truth data classification of Songket motifs by expert respondents was evaluated through a comprehensive confusion matrix calculation and is detailed in this section.

The confusion matrix technique's evaluation process is based on the criteria chosen by the respondents and categorized as images in the prediction class. Concurrently, images in the ground truth group of Songket Motifs are categorized as accurate class data. Once the confusion matrix evaluation table is completed, the matching process that the respondent has selected is carried out to enable the calculation process using the confusion matrix technique to be carried out. According to [9, 17], the advantage of the confusion matrix is that it can analyze the performance of each classifier in detail, even if the data set is unbalanced.

The performance measures often used are precision, recall, and F-measurement techniques. F-measurement evaluation is performed using True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) matrix values. The value of this matrix is evaluated for each class based on the obtained confusion matrix table [2, 13].

The definitions of TP, TN, FP, FN, precision, recall, and F-measurement are as follows:

TP: The number of accurate data correctly classified into its class by the classifier.

TN: Number of actual data classified into other classes by the classifier.

FP: The number of non-real data classified into its class by the classifier.

FN: Number of non-real data classified into other classes by the classifier.

Precision: The fraction of the number of accurate data classified exactly into its class divided by the total number of data classified in its class as in Eq. (2).

$$TP / (TP + FP) \quad (2)$$

Recall: The fraction of real data correctly classified into its class as in Eq. (3).

$$TP / (TP + FN) \quad (3)$$

F-Measurement: The combination of precision and recall values produces a value called the harmonic mean to ensure that the measure of effectiveness correlates with precision. The F-measure formula is shown in Eq. (4),

$$F - Measurement = 2 * \frac{Precision * Recall}{Precision + Recall} = \frac{2TP}{2TP+FP+FN} \quad (4)$$

Table IX shows the confusion matrix table involving the fractional motifs of 'Alam Benda', and the specific motifs of 'Keris'. Part A is a matching selection of respondent number three, representing three expert respondents in Part (A) who agree on the similarity of the specific Motif. When Part B results from a complete evaluation using the formula of precision, recall, and F-measurement.

TABLE IX. EVALUATION OF THE CONFUSION MATRIX FOR THE GROUND TRUTH DATA CLASS OF THE SONGKET MOTIF 'KERIS' PART (A) MARK THE CONFUSION MATRIX, PART (B) EVALUATION RESULTS FOR PRECISION, RECALL, AND F-MEASUREMENT

		
	3	3
	3	3

Total	Precision	Recall	F-Measurement
2	100	100	100
2	100	100	100

Next, Table X to Table XIV displays the performance evaluation results of the main motifs 'Flora', 'Fauna', 'Motif', 'Nature', and 'Cosmos' which involves 50 ground truth data of Songket Motifs that have been selected for the continuity of the image retrieval system testing process implemented in the testing and evaluation phase. Table X is the performance evaluation result of the 'Flora' ground truth Motif's image. The ID-Category involved is 004, 019, 036, 001, 012, 047, 031, 003, 022, 048, 033, 034, 032, 008, 021, 007, 025, 010, 009 and 011. Each obtained the result of calculating the F-measurement value between 26.7% to 100%.

TABLE X. RESULTS OF THE EVALUATION OF PERFORMANCE MEASUREMENT OF SONGKET MOTIF'S IMAGES IN THE 'FLORA' CATEGORY

Flora			
ID-Category	***Precision***	***Recall***	***F-Measurement***
004	100	100	100
019	100	100	100
036	100	100	100
001	72.7	100	84.2
012	63.6	100	77.8
047	61.5	100	76.2
031	60	100	75
003	52.2	100	68.6
022	50	100	66.7
048	46.2	100	63.2
033	45.5	100	62.5
034	45.5	100	62.5
032	36.4	100	53.4
008	36.5	100	53.5
021	31.3	100	47.6
007	25	100	40
025	21.9	100	35.9
010	19.2	100	32.3
009	17.3	100	29.5
011	15.4	100	26.7

Table XI shows the performance evaluation results of Songket Motif's images for the 'Fauna' category. Based on Table XI, six actual Songket Motif data are classified in the 'Fauna' category. The six ground truth data involve ID-Category 013, 043, 040, 039, 050, and 042. Meanwhile, the results of the F-measure evaluation are 100%, 92.3%, 92.3%, 89%, 72%, and 35.3%.

TABLE XI. RESULTS OF THE EVALUATION OF PERFORMANCE MEASUREMENT OF SONGKET MOTIF IMAGES IN THE 'FAUNA' CATEGORY

Fauna			
ID-Category	***Precision***	***Recall***	***F-Measurement***
013	100	100	100
043	85.7	100	92.3
040	85.7	100	92.3
039	80	100	89
050	56.3	100	72
042	21.4	100	35.3

Table XII. evaluates the performance measurement of specific Motif images for the 'Motif' category. This category is labeled as 'Motif' because specific Motif images obtained through different sources from [18], which is titled 'Symbolism in Terengganu Melayu songket motifs' and specific Motif in this class, cannot be mapped in any category that was introduced in the writing of the book. Therefore, in Table XII, nine ground truth Songket Motif's data are classified under the 'Motif' category. The nine ground truth data involve ID-Category 014, 015, 024, 029, 023, 026, 028, 030 and 027. Meanwhile, the results of the F-measurement evaluation are 100%, 100%, 75%, 58.1%, 57.1%, 54.5%, 53.3%, 53.3% and 47.6%.

TABLE XII. RESULTS OF THE EVALUATION OF PERFORMANCE MEASUREMENT OF SONGKET MOTIF'S IMAGES IN THE 'MOTIF' CATEGORY

Motif			
ID-Category	***Precision***	***Recall***	***F-Measurement***
014	100	100	100
015	100	100	100
024	60	100	75
029	40.9	100	58.1
023	40	100	57.1
026	37.5	100	54.5
028	36.4	100	53.3
030	36.4	100	53.3
027	31.3	100	47.6

Further, Table XIII shows the performance evaluation results of the ground truth Songket Motifs of the 'Nature' category. Referring to Table XIII, eight ground truth data of specific motifs are classified in the 'Nature' category. The eight ground truth data involve ID-Category 006, 020, 038, 046, 037, 005, 044, and 35. Meanwhile, the results of the F-measurement evaluation are 100%, 100%, 100%, 84.2%, 79.8%, 76.9 %, 63.2%, and 30.8%.

TABLE XIII. RESULTS OF THE EVALUATION OF PERFORMANCE MEASUREMENT OF SONGKET MOTIF'S IMAGES IN THE 'NATURE' CATEGORY

Nature			
ID-Category	*Precision*	*Recall*	*F-Measurement*
006	100	100	100
020	100	100	100
038	100	100	100
046	72.7	100	84.2
037	66.7	100	79.8
005	62.5	100	76.9
044	46.1	100	63.2
035	18.2	100	30.8

Table XIV is the ground truth data of Songket Motif's classified under the 'Cosmos' category. The seven ground truth data are ID-Category 002, 041, 045, 017, 049, 018, and 016. Meanwhile, the evaluation results of the F-measurement for ID-Category are 100%, 94.1%, 84.2%, 70.6%, and 70, respectively. %, 62.5%, and 53.3%.

TABLE XIV. RESULTS OF THE EVALUATION OF PERFORMANCE MEASUREMENT OF SONGKET MOTIF'S IMAGES IN THE 'COSMOS' CATEGORY

Cosmos			
ID-Category	*Precision*	*Recall*	*F-Measurement*
002	100	100	100
041	88.889	100	94.1
045	72.727	100	84.2
017	54.5	100	70.6
049	53.846	100	70
018	45.5	100	62.5
016	36.4	100	53.3

Referring to Table X to Table XIV, there is a low F-measurement value due to the calculation process using the confusion matrix technique; the selection for each category of similar Songket Motif types has the sum of several fractions of different Songket Motif types. For example, 'Daun' and 'Anggur' specific motifs were chosen by respondents as similar but in different categories. This is because the motif has a similar structure.

V. CONCLUSION

Two research objectives were discussed in determining the ground truth data for Songket Motif. The first objective was to identify the ground truth data set involving Activity One, which aimed to obtain two ground truth data sets: training data set and test data set. These data sets involved six categories: ' Flora', 'Fauna', 'Nature', 'Cosmos', 'Food', and 'Calligraphy'. The survey was conducted through a qualitative method involving 15 respondents who classified 413 specific motif images into 56 Songket Motif categories referring to the six prominent motifs.

The second objective was a validation-classification test of ground truth data sets by three experts. The aim was to equate the selection of general respondents with expert respondents to obtain training data sets for testing purposes involving six categories. After rearranging, only 50 ground truth Songket Motif's data were selected referring to five categories: 'Flora', 'Fauna', 'Nature', 'Cosmos', and 'Motif'. The relationship coefficient correlation method was implemented to evaluate the relationship between two data through a statistical evaluation angle.

In addition, precision and recall methods were also used to obtain precision and recall values for each ground truth data that had been selected. The F-measurement method was used to make a single evaluation that was the average value of each ground truth data. In Phase Two, the analysis of the study focused on human observation of the appearance of Songket Motifs.

VI. FUTURE WORK

The limitation of the study of determining ground truth data through human observation is that the detailed geometric characteristics are less effective because every human observation evaluates shape characteristics according to the individual's perception.

Thus, the technology determination technique is also able to help in the process of determining the ground truth data for the image retrieval research. Accordingly, this research can be developed by providing data determination techniques in computer vision.

REFERENCES

[1] Kementerian Komunikasi dan Multimedia Malaysia. 2018. Jenis Warisan. Kementerian komunikasi dan multimedia Malaysia. http://www.kkmm.gov.my/index.php?option=com_content&view=article&id=594:jenis-warisan&catid=65:dokumen&lang=en [17 December 2020].

[2] J. Nursuriati, A. B. Zainab, & T. S. Tengku Mohd, Image retrieval of songket motifs using simple shape descriptors. GMAI '06 Proceedings of the conference on Geometric Modeling and Imaging: New Trends, Vol. (2006), pp.171–176.

[3] Yuhandri, S. Madenda, E. P. Wibowo, & Karmilasari, Pattern recognition and classification using backpropagation neural network algorithm for songket motifs image retrieval. International Journal on Advanced Science, Engineering and Information Technology. vol. 7(6): pp.2343–2349, 2017.

[4] Hasan, M. A. & Liliana, D. Y. 2020. Pengenalan Motif Songket Palembang Menggunakan Deteksi Tepi Canny, PCA dan KNN. Multinetics 6(1): 1–7. doi:10.32722/multinetics.v6i1.2700.

[5] Jusam, A., Yu, W. C., Rafee, Y. M., Awang, A., Md.Yusof, S. Z., Jussem, S. W. & Abol Hassan, M. Z. 2021. Pengaplikasian Teknologi Visual dalam Penghasilan Inovasi berkaitan Proses Penghasilan Songket Rajang di Sarawak. Jurnal Dunia … 3(3): 105–116. Retrieved from https://myjms.mohe.gov.my/index.php/jdpg/article/view/16072%0Ahttps://myjms.mohe.gov.my/index.php/jdpg/article/download/16072/8351.

[6] K. Scott, Computer Vision Metrics Survey, Taxonomy, and Analysis. Apress Open, 2014.

[7] Feng, D., Siu, W. C. & Zhang, H. 2003. Multimedia Information Retrieval and Management. (H. J. (Eds.) Feng, David, Siu, W.C., Zhang, Ed.), 1st (2003). New York: Springer.

[8] J. Gbolahan Adigun, Ground Truth Data for Object Detection in Autonomous Vehicle from a Driving Simulator. Tallinn University of Technology. Retrieved from https://www.researchgate.net/publication/352573147, 2020.

[9] O. Aniza, Ciri Pembeza Pengelas Penampilan Warna Pemandangan Lukisan Landskap Berdasarkan Pengamatan Manusia Terhadap Warna. Universiti Kebangsaan Malaysia, 2021.

[10] M. W. A. Kesiman, S. Prum, J. C. Burie, & J. M. Ogier, An Initial Study on the Construction of Ground Truth Binarized Images of Ancient Palm Leaf Manuscripts. Proceedings of the International Conference on Document Analysis and Recognition, ICDAR. IEEE Computer Society, 2015, pp.656–660.

[11] Y. Nadiah, Model Capaian Imej Motif Songket Berasaskan Teknik Analisis Komponen Utama dan Jarak Kuadratik Geometri. Universiti Kebangsaan Malaysia, 2023.

[12] J. Mccormac, A. Handa, S. Leutenegger, & A. J. Davison, SceneNet RGB-D: Can 5M Synthetic Images Beat Generic ImageNet Pre-training on Indoor Segmentation?. International Conference on Computer Vision (ICCV) IEEE, 2017, 2380-7504.

[13] A. Antonacopoulos, D. Karatzas, & D. Bridson, Ground Truth for Layout Analysis Performance Evaluation. International Workshop on Document Analysis Systems, SpringerLink. vol. 3872, 302-311, 2006.

[14] S. Ibrahimi, A. Sors, S. D. R. Rafael, & C. Stéphane, Learning with Label Noise for Image Retrieval by Selecting Interactions. IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) IEEE, pp. 2642-9381, 2022.

[15] S. R. Charan, Statistical Method in Medical Research. Springer Singapore, 2018.

[16] B. Sarah, Statistic in a Nutshell A Quick Desktop Reference. O'Reilly Media, 2012.

[17] A. Bhandari, Everything you Should Know about Confusion Matrix for Machine Learning. https://www.analyticsvidhya.com/blog/2020/04/confusion-matrix-machine-learning/ [14 November 2022].

[18] A. A. Arba'iyah, Simbolisme dalam motif songket Melayu Terengganu. Dewan bahasa dan Pustaka, 2018.