

Visualising Image Data through Image Retrieval Concept using a Hybrid Technique: Songket Motif's

Nadiah Yusof¹, Amirah Ismail², Nazatul Aini Abd Majid³

Faculty of Information Science and Technology
Universiti Kebangsaan Malaysia (UKM) Bangi Selangor Darul Ehsan Malaysia

Abstract—It has been proven that the massive dataset is strictly complex in Content Based Image Retrieval (CBIR) because the present strategies in CBIR might have faced difficulties in feature extraction of the images. Moreover, technological constraints encountered in the analysis and extraction of the image arrays are how the system customizes the primitive geometric structures known as polygonal approximations structure. Hence, this study has discovered that image feature extraction is utilized by applying the Principal Component Analysis (PCA) technique, which is primarily based on the matrix of image representation that will enlarge the similarity of detection. The PCA approach needs to be enhanced resulting from the lack of the extraction of features in songket motives images. Therefore, this study proposes a new hybrid model that will integrate PCA with geometric techniques for image feature extraction to increase the recall and precision result. This paper employs the use of a qualitative experimental design model that involves three phases of activities. First, the analysis and design phase, secondly is a development phase, and lastly is the testing and evaluation phase. This paper focuses on those two phases in terms of design and development phases. The outcome process of the empirical phase is followed by designing the algorithm and model based on the result of literature review. This study has found that the hybrid between the principal component analysis model and the geometry technique will help to reduce the problems faced by the basic engineering technique model, which is the constraint in analysing and extracting the image features to customize the geometric primitive structure.

Keywords—Multimedia; image; content-based image retrieval (CBIR); image retrieval; near-duplicate; principal component analysis (PCA); geometric

I. INTRODUCTION

Malays are rich in fine art heritage, inherited since immemorial time. Weaving, embroidery, engraving and fine arts are so synonymous with Malay custom. Creativity and fine arts are reflected in each of these traditional handicrafts. The fine arts of Malay cultural heritage include weaving, batik art, fireplace, stick, dastard, beads, engraving, brass, and songket. [1],[2]. Songket is a fine art heritage with many privileges. The important part of the songket fabric is the structure of songket motif's image that shows the different philosophy and meaning [3]. Table I shows that preservation aspects of songket motif's from many websites are not solely the main purpose but focuses towards the commercialization aspects and concerns [4]. Researchers play an important role in preserving the continuity of the Malay cultural heritage from extinction in this modern era [2]. Songket preservation and conservation efforts

can be undertaken by digitizing the principal component of songket fabric; the songket motif's arrangement.

The arrangement of songket motif's is an important element of songket fabric which provides an avenue for the songket motives image digitization research and studies. The digitization of songket motif's will help to preserve the songket motif's from extinction and allow the continuation of heritage for future generations. While defining features and structure of images will help to facilitate a database of songket motif's proceeds with the clustering process according to a specific category. This process eases the users to store and retrieve the image of songket motif's without having to categorize the songket motif's image manually. Thus, automatic clustering of songket motif's can be performed by the implementation of near-duplicate image retrieval technique. Various image retrieval techniques have exploited and included those traditional image retrieval techniques which have been implemented through the content-based image retrieval (CBIR).

Content-based image retrieval [6] field has been rather popular among researchers through a rapid study along with other popular research areas such as Database Management, Natural Language Processing, Signal Processing, Computer Vision, Network Systems and Human-Computer Interaction [7]. Since studies on the field of image retrieval have proliferated [8] to the result of the image obtained through the image retrieval system [9],[10],[11]. Various models of image feature extraction were introduced through the implementation of the development on a variety of retrieval systems. The developed system uses multiple models of processing query image features to represent the information space contained in the image, whether in the query or the database space.

The implementation of a model description of the image feature in the field of image retrieval has been extended through the customization of the model in the field of pattern recognition to help analyse the query image features and images contained in the database to fit the image retrieval field. Implementation of the study mentioned above is part of the detect and matching process to retrieve near-duplicate images. The precision of the image can almost be seen from the point of similarity found in image features (such as; shapes, textures, text, sketches, and colours). The features mentioned are part of the basic content-based image retrieval technique. All these techniques can assist the process of near-duplicate image entirely, as evidenced by various studies that have been carried out [10],[12],[13],[14]. However, the techniques involved in

CBIR face obstacles in processing image datasets in high-dimensional clusters [15],[16],[17]. Technological constraints encountered in the analysis and extraction of image arrays indicate how to customize primitive geometric structures and are better known as polygonal approximations.

TABLE. I. SURVEY ON QUERY TECHNIQUE ON RELATED SONGKET WEBSITE : [5])

Survey on Query Techniques	
Website	Explanation/Content
Institut Kraf Negara (www.ikn.gov.my)	Craft institution website
Visit Terengganu (www.visit-terengganu.net)	Tourist and advertisement website
Songket Moden (songketmoden.com)	Product commercialize
Warisan Budaya Melayu (malaysiana.pnm.my)	Malaysian culture heritage website
Azizah Songket Terengganu (azizahsongket.wordpress.com)	Product commercialize
Songket Restaurant (www.songketrestaurant.com)	Product commercialize
Bibah Songket (www.bibahsongket.com)	Product commercialize
Atikah Songket TTDI (www.atikahsongket.com)	Product commercialize
Aura Batik (aura-batique.blogspot.com)	Product commercialize
Kain Songket.com (kainsongket.com)	Product commercialize

Therefore, this study has found that the hybrid between the Principal Component Analysis [18] model and the geometry implemented help to reduce the problems faced by the basic Engineering technique model, which is the constraint in analysing and extracting the image outline for the purpose of customizing the geometric primitive structure where the problem can solved by emphasizing aspects image representations of the primitive data structure angles within the image [19],[20] which include image size, image depth, angle and compression of images. These are the components found in the Principal Component Analysis model. Also, the principal component analysis model will help to decrease each dimension of the image features and structure [16], while geometric techniques are identified to help detect the global space within the image and continued with scale, rotation, small and big size of images.

The paper is organized as follows. In Section 2, a brief review of relevant works is presented, and labelled as Research Background; Section 3 presents the research flow described as Method. Meanwhile, Section 4 discusses the structure of the system and development; The experimental results of the studies on algorithm evaluation are presented in Section 5 and the paper is concluded in Section 6.

II. RESEARCH BACKGROUND

Multimedia data (text, images, audio, and video) have always inundated websites [21], including YouTube, Google videos, Facebook, Instagram, Twitter, and Flickr [22]. In addition to providing information and connecting to the world without boundaries, it can also harm database information

management in managing the flow of multimedia data, especially images. Uploading images again will increase a similar image and is known as a near-duplicate image. Currently, almost all images on various websites show over 80 percent similarity [23],[24]. This issue necessitates the study of near-duplicate image retrieval in order to match and cluster the near-duplicate image in unique characteristic.

A. Challenges in Near-Duplicate Image Retrieval

There are two challenges in detecting the near-duplicate image. The first challenge is that the selection of features extraction technique aims to compute the percentage of similarity of image features in large-scale databases [25]. The second challenge is the process of image clustering with an existing template in a database. Assignments in the clustering process involve time, angles, illumination and resolution of the image that need to be detailed and compared for similarity properties, which makes the process involved in clustering process complicated than the implementation of near-duplicate image retrieval tasks [17], [26]. However, the two tasks are mutually required to enable image processing in the search system.

B. Image Features Extraction

The essential element in near-duplicate image retrieval is through the technique of matching and comparison between query images and stored images in the database [26]. Furthermore, several characteristics are related to near-duplicate image retrieval whereby detection for near-duplicate image involves two-part of images; the first part is a query images and the second part is images inside the database. Category of images divided into two main sections, the general image, and the specific image category. Commonly, general images are images of various activities provided by users through the social media sites, image management centres, personal blogs, online drivers such as Google Drive, Dropbox and others [17] whereas the specific image is a group of images that come from different areas such as medical image scan, cultural heritage, criminal image, chemical structure, and building structure. Usually, specific images database is provided only for a specific user and private databases. Besides, to extract the image features, the image needs to be extracted based on the structural primitive contained within the image.

Image features are analyzed through the calculation of structural primitives and the technique of placement within the image because that technique is very effective against complex image structures. While the other technique method of calculating similar structure of images is based on statistics that include Fourier Power Spectrum Statistics [27], Co-Occurrence Matrices [28], [29], Fixed-Invariant Principal Component Analysis (SPCA) [19],[30], Tamura Feature [31], Wold Decomposition [32], Random Markov Fields [33], Fractal Models [34], and Multi-Resolution Filtering Techniques such as Gabor and Wavelet Transform [11]. The features found in texture techniques are statistically performed through the distribution of image intensity [11],[29],[31], [26],[35],[36]. This technique is very effective in measuring the shape of the images matrix to obtain the different surfaces of the objects [37],[38]. However, the constraints encountered are that the

boundaries between objects that have a similar tone of objects that are difficult to be identified, thus, the process to decipher the structure of each object in the image being executed is hindered.

Furthermore, research on near-duplicate image retrieval also includes Sketch-Based Image Retrieval (SBIR), and it has begun as early as the 1990s [39],[4]. This study was sparked by the domain of cartoon images. The primary purpose of this technique is to help those who are not able to read especially children. This technique allows the user to retrieve the cartoon image by sketching the imaged bone into the sketch canvas, and the system initiates the sketch structure in order to match a similar image by sketch bone. This technique has been proven to be more effective in assisting users than textual techniques and is more user-friendly, especially for children [5],[4],[40]. In general, almost all researchers in the field of image retrieval support and agree that content-based image retrieval techniques will allow the users retrieve similar image much more efficiently and effectively [7],[4]. However, the technique in CBIR encounters barriers with image issues as more complex and challenging images are required to determine the similarity of queries that users aim to reach, and this is a complex task for the CBIR system to perform [4],[41]. Therefore, various technological innovations have been proposed by subsequent researchers who aim to assist in the process of features extraction of the near-duplicate image retrieval structure.

Furthermore, the process of extraction of image features needs a calculation on the precision value of the similarities between the query image and the stored image in the database. Additionally, the value of similarity in image retrieval is measured based on the estimated empirical values contained in the near-duplicate image characteristics that have been identified by the system and is shown in Fig. 1. The similarities between query images and stored images may impede the measurement of the retrieval value and image accuracy in determining the effectiveness of the developed image retrieval system [43].

Moreover, various studies and techniques involving near-duplicate image retrieval have been conducted recently in order to solve some issues and problems pertaining the related technique, which is the bag of visual word and Min-Hash [44],[16], This technique focuses on clustering the near-duplicate images through the matching of space and position within the image. Meanwhile, the Singular Value Decomposition-Scale Invariant Feature Transform (SVD-SIFT) technique [45] implements a catalyst method to speed up the image detection process. Subsequent Salient Riemannian Visual technique [22] aims to identify the prominent space within the processed image. Furthermore, the data duplication technique [46],[47],[48],[49] optimizes database usage by reducing subsequent data match and comparing to bits of help reduce electricity consumption which can increase heat production. Another technique is Similarity Join Operator Technique [42],[50] that evaluate image similarity based on absolute ratios, whereas, Fourier Mellin Transform technique [51] helps image detection via image rotation, image scaling and invariant changes found in the image, followed by the Haar-wavelet technique [52],[53] that assist to extract features

vector that is included in the image in order to find the Manhattan distance of the object in the image. Furthermore, kernel hashing technique [54] helps to detect the near-duplicate image by examining the various features contained within the image to detect the differences of each image and convert it to binary of images to place within the kernel space. Although most of the techniques mentioned above focus more on image features extraction for low dimensionality, but when applied to high dimensional images, the system has to deal with the time-consuming problems in image feature extraction and can cause the delay of the process [17]. In consequence, this problem has suggested several solutions.

Therefore, among other techniques that have been introduced in order to solve image features extraction problems for a massive dataset, the locality-constrained linear coding (LLC) & max-idf [17] technique has been proposed to improve features extraction structure of the image through the fragmentation of the matrix contained in the image and continue with clustering the near-duplicate image by dividing the image into several baskets following by the essential features. Furthermore, the next step is to apply to a matching technique within the space for the image clustering task based on the image feature extraction. The hashing technique [16],[48] also will help to extract the features on the image in large-scale images dataset, the process following by detecting Jaccard similarity structure where the images have high precision included in the same cluster. Another technique used is the Min-Hash & tf-idf weighting technique [55],[56] whereas those techniques are more focused on detecting near-duplicate images in local space of images. As shown in Table II, almost all of the researchers focused only on local space while only one researcher focused on the two spaces (local & global) in the image [26].

A study conducted by Hassanian [23] agreed whereby focusing on the two spaces within the image helps to improve the image accuracy results. This combination, therefore, enhances the effectiveness and overall efficiency of image features extraction almost immediately [57]. Also, this statement was supported by [58], which stated that a combination of global and local features helped to increase accuracy by up to 65.5 percent. The global space is the basic structure in the image, while the local space is the space that meets the basic structure.



Fig. 1. Framework of Near-Duplicate Image Proposed by [42].

TABLE. II. SUMMARY OF PREVIOUS TECHNIQUES HAVE BEEN USED TO DETECT NEAR-DUPLICATE IMAGE RETRIEVAL PROBLEM

Previous Techniques in Near-Duplicate Image Retrieval		
References	Technique	Local /Global
[22]	Rotation, Scale, Translation (RST) Invariance features + Salient Covariance Matrix (SCOV) + ICA Independent component analysis	Global
[59]	Scale In variance Features Transform (SIFT)	Local
[60]	Scale In variance Features Transform (SIFT)	Local
[61]	Scale In variance Features Transform (SIFT) + Histograms of oriented Gradient (HoG) + BoF + KMean Clustering	Local
[62]	Locality Sensitive Hashing + K-Nearest Neighbor SIFT + K means + BoVW	Local
[63]	Colour Texture Moment (CTM)	Local
[64]	Strong Geometry Consistency (SGC) + Scale Weighting	Local
[65]	Bag of Visual Word (BoV)	Local
[26]	Color Moment + Wavelet Transform + SIFT	Local & Global
[66]	Bag of Visual Word (BoVW)	Local
[67]	K-Nearest Neighbor	Local
[16]	Min-Hashing + Jaccard Similarity	Local
[68]	Bag of Word	Local
[69]	K-Mean Clustering + Bag of Word	Local
[42]	K-Mean Clustering	Local
[17]	Locality Linear Coding + MaxIDF-cut + K-Means clustering	Local
[55]	Canny Edge Detection+ Great Deluge Algorithm	Local
[23]	Min-Hash + Locality Sensitive Hashing	Local

III. METHOD

This section focuses on three main phase's iteration; first, is the analysis and design phase, secondly is a development

phase, and lastly is the testing and evaluation phase. Fig. 2 shows the interconnection between the three phases implemented in the research, but only two phases which are coloured has been applied in this paper.

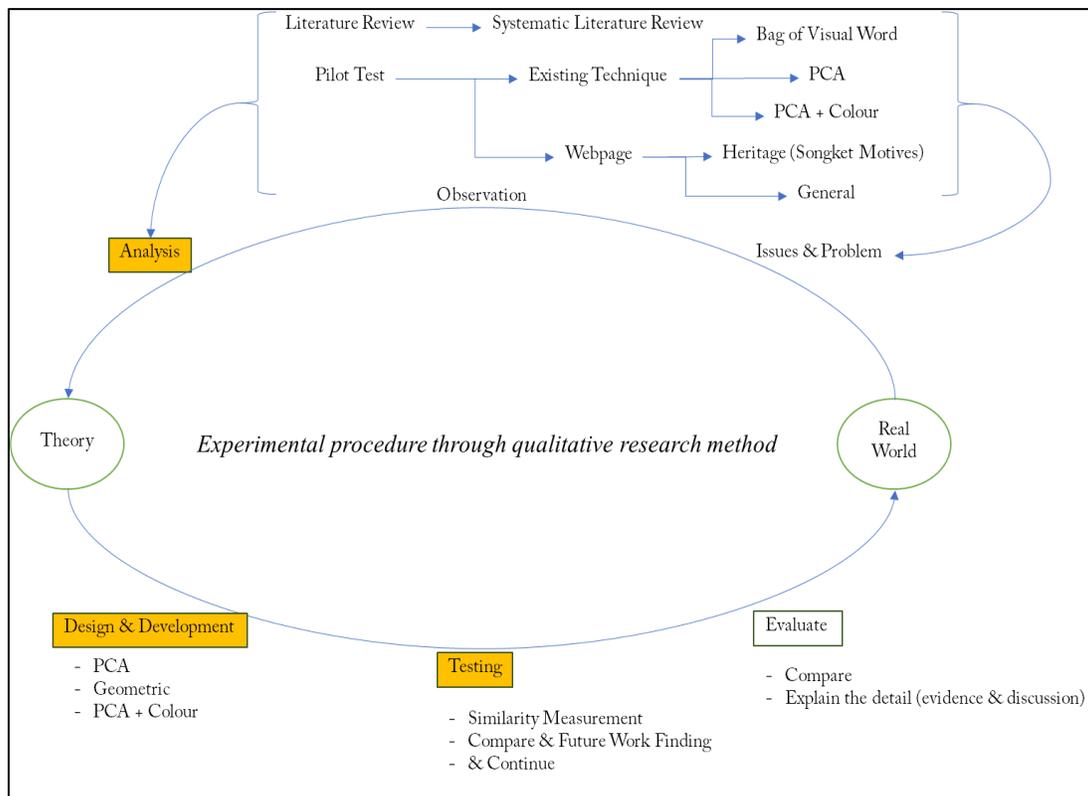


Fig. 2. The Interconnection between the Three Phases has Implemented in the Paper[70].

The analysis phase divided is into two main categories of the process. The process is followed by analysing more than 400 articles, that include books, in order to structure and prepare a systematic literature review on a variety of techniques that have been applied and proposed in near-duplicate image retrieval and clustering. The advantages of a systematic literature review will help the readers to find a new significant contribution that can contribute to the body of knowledge in image retrieval research field. Also, it is important to look at the advantages and disadvantages of each proposed technique and, which techniques are appropriate to be applied to the cultural heritage domain and various other domains in general. As can be seen, the songket motif's image structure is more geometric, so the hybrid of the two techniques is assumed to coincide with the image structure described earlier after obtaining technique suitable for near-duplicate image retrieval. Subsequent design and algorithm development were performed using MATLAB 2019a software. Finally, after image detection algorithms were almost fully utilized, the PCA technique was developed. Technical testing is performed by applying Local Binary Pattern and similarity measurement technique to see image results and is a guide to improvements to the development of hybrid image detection of hybrid algorithms.

IV. SYSTEM IMPLEMENTATION

System implementation starts with applying principal component analysis technique with eigenvalues on images (query/databases) in order to proceed on comparison of image size following by converting the images into double-precision value using sin & cos algorithm $(\% (s_j \sin h_j - s_i \sin h_i)^2)$ calculation, then proceeding with the features extraction process.

Features extraction process is followed by computing, selecting, and normalizing the eigenvalue of images. The number of eigenvalues is always lower or equal than the number of the original image (i.e., $K < M$), and the process in normalization is to remove all the standard features in the images in order to get the lower dimension of images. A method on normalizing shape value using principal component analysis following by calculation of average shape value then calculating the mean (average) shape value to urge the normalizing of the shape value. Consequently, to calculate eigenvalue in images, the system must be used and calculate the covariance matrix first and the algorithm is shown in (1) and Fig. 3 below is an example direction and calculation of features extraction using principal component analysis on the images.

$$\% cov(X, Y) = \frac{1}{n} \sum^n (X_1 \bar{X})(Y_1 - \bar{Y}) \quad (1)$$

Principal component analysis eigenvalues will be considering each pixel in an image as a separate dimension.

For example, $N \times N$ images has N^2 pixels; therefore, the value in N^2 has a dimension (e.g., $236 \times 236 = 55696$), that values are showing a vast amount number, and the calculation process will make the system slower, and run out of memory, and computation calculation on that amount requires a vast process. The system will process to a reduced dimension of images by calculating the eigenvalue from covariance matrix using principal component analysis into lower dimension ($100 \times 100 = 10000$) of size images in order to recognize the vital image structure and known as a principal component. Fig. 4. shows the process of a reduced dimension of geometric images by computing, selecting, and normalizing the dimension inside the image's matrix.

Rotation changes the spatial relationships between the grid cells and the object boundaries, which leads to a completely different matrix number representation for a similar object. Therefore, objects should be through the normalizing process for a geometric shape rotation. The most important axis of the shape will be accomplished by connecting the two points on the shape boundary furthest faraway from one another.

The shape is then rotated to create a significant parallel to the coordinate axis. An example of rotation normalization in Fig. 4 shows an object before rotation normalization and scale normalization of an object [72] and the whole process in the system is also shown in Fig. 5. The principal component analysis is improved by enhancing the process of rotation, scale, and normalization using a geometric algorithm calculation. By describing the direction of rotation, an example is shown in Fig. 6 from 0 to 315 degree [71].

```
Features Extraction
% compute, select & normalize eigenvector
[V, D] = eigs(cov([M1(:) M2(:)]));
% matrix D containing the eigenvalue on the main
diagonal
% matrix V whose columns are the corresponding
eigenvalue

.....Direction of image scan


|         | Rows (:,n); |
|---------|-------------|
| C       |             |
| O       |             |
| L       |             |
| U       |             |
| M       |             |
| N       |             |
| S       |             |
| (m, :); |             |


if (D(1,1) > D(2,2))
    a = V(:,1) ./ sum(V(:,1));
else
    a = V(:,2) ./ sum(V(:,2));
end;
```

Fig. 3. Image Features Extraction Process using Covarians Matrix.

A training set consisting of total M images

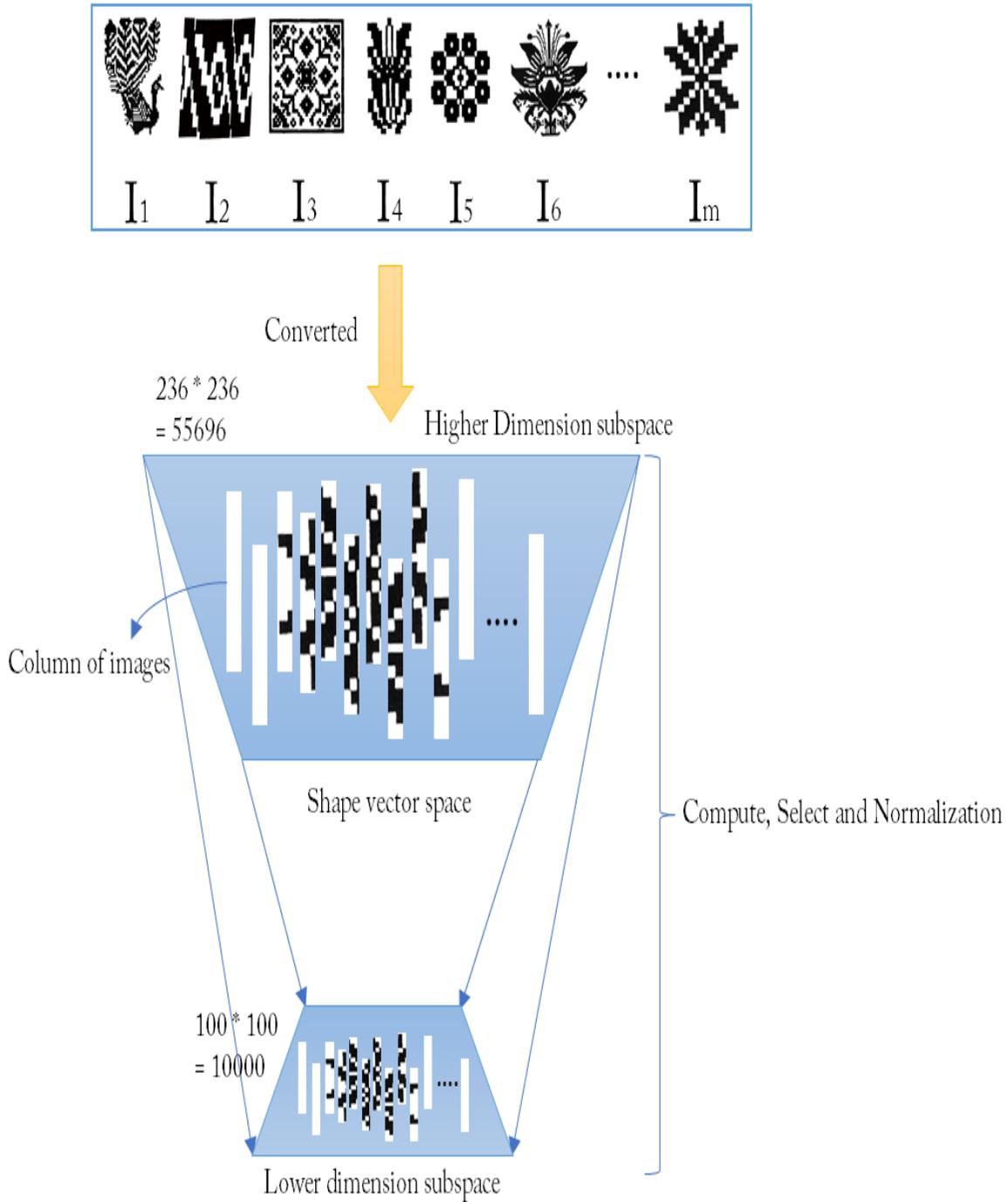


Fig. 4. A Step of Compute, Select and Normalize the Image Vector Space.

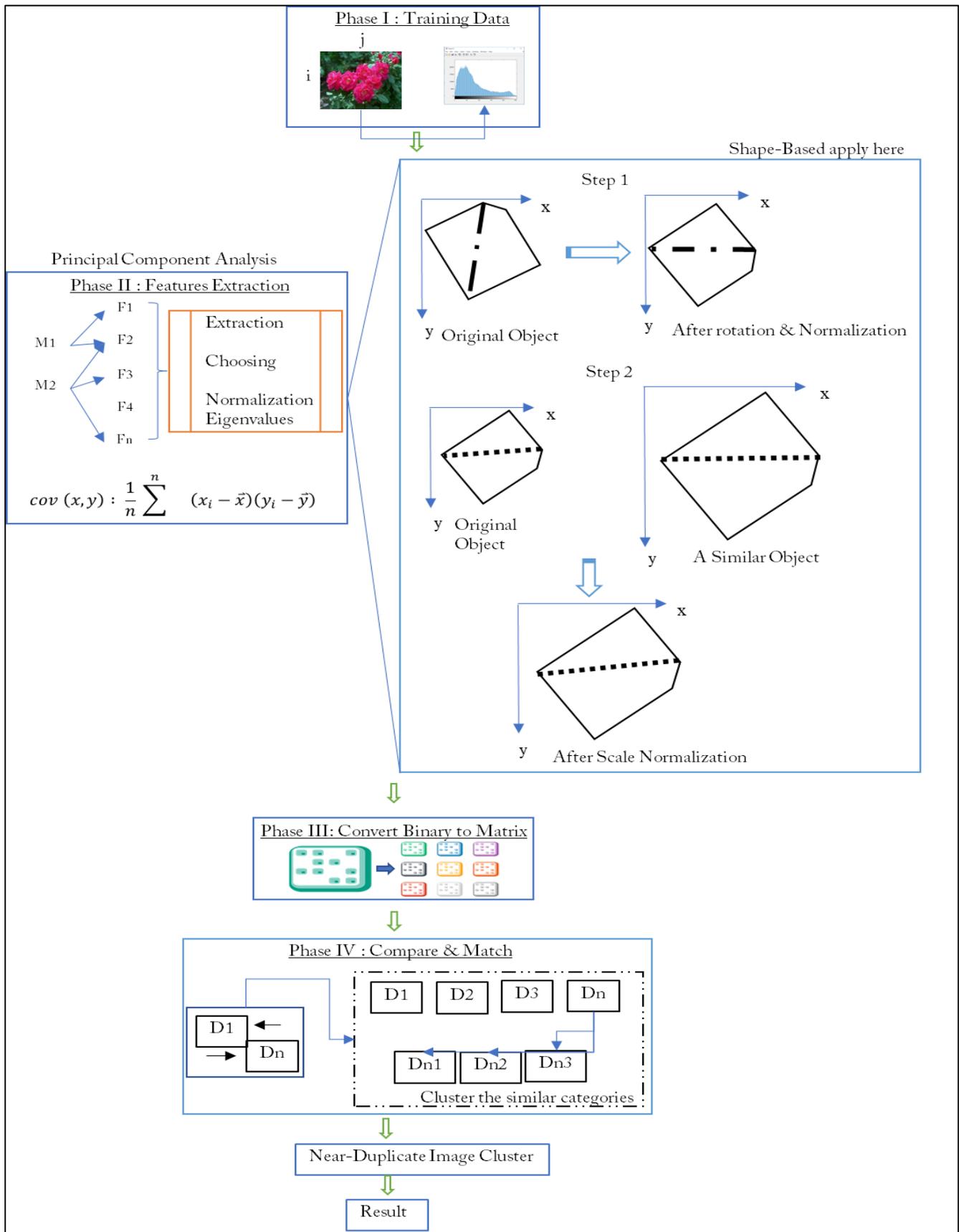


Fig. 5. The System Flows for Near-Duplicate Image Retrieval.

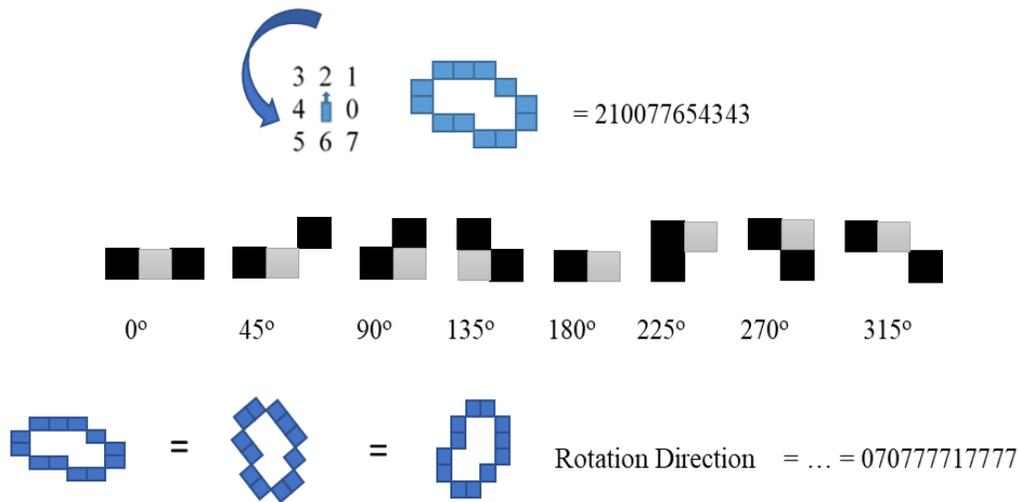


Fig. 6. The Degree of Direction Applying in the System in Order to Help Improve the Detection of Near-Duplicate Images Redrawn from [71].

V. RESULT AND DISCUSSION

The testing recall results using songket motives images data that has been deployed in two different techniques in order to prove the theory conception of this paper. Table III could be a half part of the recall result using principal component analysis technique. Table V shows a half part of a recall result using a bag of visual word technique, while Table IV (PCA) and Table VI (BoVW) show a full result of recall and precision for the two technique that has been tested in this paper.

Based on the preliminary tests conducted on the principal component analysis technique and the bag of visual words technique, the percentage of image precision value is calculated by local binary pattern and similarity measurement technique showed that the principal component analysis technique result is higher than the bag of visual word technique as in Fig. 7. The bag of visual words technique analysed the

image through the chunk of the image, and it is difficult to finalize the real shape structure of image template, and thus, the results obtained are less accurate than the principal component analysis techniques. Whereas the principal component analysis technique is seen to produce higher precision results in image retrieval but lacking with less recall of images. It is seen as limiting the choice of the user; based on the suggestion by Prof. Dr. Nursuriati Jamil during the interview (May 2019) session, result of retrieve images can be further enhanced by improving image rotation, image scaling, improvising the image size by enlarging and reducing the size of images. According to her suggestions, the suitable technique is the implementation of a hybrid technique together with the principal component analysis and quadratic distance algorithm to be applied in the near future in order to improve a recall result and precision result of image songket motives.

TABLE III. A PARTIALLY OF PRINCIPAL COMPONENT ANALYSIS TECHNIQUE ON SONGKET MOTIVES IMAGES RETRIEVAL RESULT

A part of Principal Component Analysis						
Query	Result					

TABLE IV. A FULL AVERAGE OF PRINCIPAL COMPONENT ANALYSIS TECHNIQUE ON SONGKET MOTIVES IMAGES RETRIEVAL RESULT

Query images										
	1	2	3	4	5	6	7	8	9	10
Principal component analysis Precision Value of Songket Motives	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
	0.41	0.37	0.38	0.86	0.79	0.71	0.70	0.71	0.66	0.77
	0.41	0.38	0.46	0.81	0.77	0.72	0.91	0.73	0.69	0.78
	-	0.48	0.48	0.92	-	0.63	0.81	-	0.71	0.74
	-	0.59	-	0.71	-	-	-	-	-	-
	-	0.59	-	0.62	-	-	-	-	-	-
Average result	0.61	0.57	0.58	0.82	0.85	0.76	0.85	0.81	0.76	0.82

TABLE V. A PARTIALLY OF BAG OF VISUAL WORD TECHNIQUE ON SONGKET MOTIVES IMAGES RETRIEVAL RESULT

A part of Bag of Visual Word									
Query	Result								

TABLE VI. A FULL AVERAGE OF BAG OF VISUAL WORD TECHNIQUE ON SONGKET MOTIVES IMAGES RETRIEVAL RESULT

	Query Image									
	1	2	3	4	5	6	7	8	9	10
Bag of Visual Word Precision Value of Songket Motives	0.99	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
	0.99	0.37	0.49	0.37	0.55	0.77	0.83	0.69	0.76	0.62
	0.41	0.55	0.75	0.82	0.45	0.43	0.85	0.72	0.57	0.78
	0.33	0.47	0.45	0.82	0.55	0.72	0.91	0.75	0.55	0.85
	0.41	0.60	0.58	0.81	0.58	0.73	0.95	0.67	0.72	0.80
	0.40	0.53	0.78	0.83	0.57	0.76	0.78	0.72	0.67	0.86
	0.47	0.59	0.56	0.83	0.47	0.77	0.77	0.77	0.42	0.50
	0.45	0.61	0.75	0.84	0.48	0.73	0.92	0.66	0.32	0.92
	-	0.61	0.77	0.74	-	0.39	0.35	0.49	0.76	0.67
	0.43	0.59	0.77	0.83	0.55	0.76	0.82	0.68	0.59	0.72
	0.42	0.63	0.76	0.81	0.33	0.74	0.82	0.68	0.64	0.87
	0.33	0.63	0.75	0.79	0.53	0.81	0.41	0.76	0.64	0.48
	0.41	0.59	0.76	0.49	0.46	0.64	0.57	0.73	0.58	0.79
	0.37	0.54	0.55	0.82	0.35	0.50	0.89	0.76	0.69	0.79
	0.44	0.61	0.77	0.54	-	-	0.31	0.78	0.61	0.65
	0.46	0.62	-	0.78	0.48	0.31	-	0.76	0.57	0.66
	0.43	0.47	0.75	0.30	0.57	0.75	0.66	0.76	0.77	0.41
	0.43	0.59	0.64	0.75	0.57	0.85	0.69	0.61	0.74	0.94
	0.42	0.45	0.44	0.80	0.33	0.86	0.78	0.71	0.30	0.82
-	0.62	0.73	0.48	0.37	0.89	0.86	0.36	0.74	0.42	
Average result	0.43	0.58	0.65	0.72	0.46	0.67	0.71	0.70	0.63	0.73

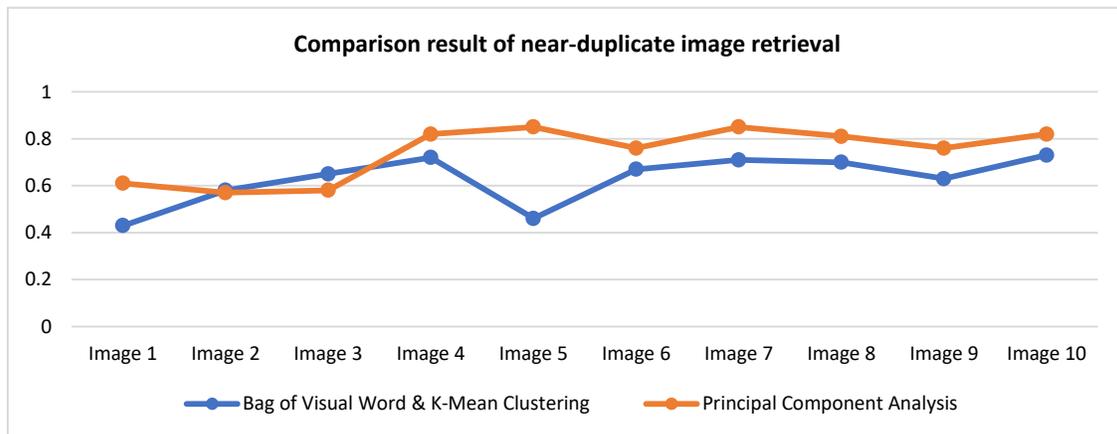


Fig. 7. Calculation by Local Binary Pattern and Similarity Measurement Technique.

VI. CONCLUSION

Similarity calculation in near-duplicate images retrieval faces several issues primarily based on unique images dataset and cluster. Furthermore, the image is in line with their category, which has comparable features that may want to assist the system and as a result, it will enlarge the percentage of recall and precision incomparable image retrieval and detection. The motive is that the way to increase the percentage of similar image detection is to center the attention on more detail of two areas - local and global image representation. This can be executed by utilizing hybrid PCA and geometric techniques. This solution is predicted to cater to the near-duplicate image retrieval issues that contain mostly local and global feature extraction.

ACKNOWLEDGMENT

This research was supported by GGP-2017-077, MyPHD MyBrain and Universiti Kebangsaan Malaysia. Special appreciation to our associates who provided the skill and knowledge that enormously assisted in the exploration of this research.

REFERENCES

- [1] N. Nawawi, A. Aziz, N. Legino, RafeahAhmad, and N. Ismail, "the Nature of Malay Songket Textile," *Bus. Manag. Q. Rev.*, vol. 6, no. (2/3), pp. 41–56, 2015.
- [2] N. bt Yusof, T. S. M. T. Wook, and S. F. M. Noor, "Comparison Result of Songket Motives Retrieval through Sketching Technique with Keyword Technique," *Int. J. Comput. Sci. Netw.*, vol. 3, no. 2, pp. 70–76, 2014.
- [3] Tenas Effendy., "Falsafah Dalam Motif's Songket Melayu," *Seminar Antarabangsa Tenun Nusantara*. pp. 1–17, 2009.
- [4] N. Yusof, T. S. M. T. Wook, and S. F. M. Noor, "Songket Motives Retrieval through Sketching Technique," *Procedia Technol.*, vol. 11, no. Iceei, pp. 263–271, 2013.
- [5] N. Yusof, "Pencarian Imej Motif's Songket Menggunakan Teknik Lakaran," *Universiti Kebangsaan Malaysia*, 2014.
- [6] M. A. Hannan, M. Arebey, R. A. Begum, H. Basri, and M. A. Al Mamun, "Content-based image retrieval system for solid waste bin level detection and performance evaluation," *Waste Manag.*, vol. 50, pp. 10–19, 2016.
- [7] S. Noah and S. Sabtu, "Binding Semantic to a Sketch Based Query Specification Tool," *Int. Arab Inf. Technol.*, vol. 6, no. 2, pp. 116–123, 2006.
- [8] A. Othman, T. S. M. T. Wook, and S. M. Arif, "Quantization selection of colour histogram bins to categorize the colour appearance of landscape paintings for image retrieval," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 6, no. 6, pp. 930–936, 2016.
- [9] Q. Dao Thi Thuy, Q. Nguyen Huu, C. Phuong Van, and T. Ngo Quoc, "An efficient semantic – Related image retrieval method," *Expert Syst. Appl.*, vol. 72, pp. 30–41, 2017.
- [10] N. Jamil, Z. A. Bakar, and T. M. T. Sembok, "Image retrieval of songket motif's using simple shape descriptors," in *GMAI '06 Proceedings of the conference on Geometric Modeling and Imaging: New Trends*, 2006, vol. 2006, pp. 171–176.
- [11] A. Verma, "Content Based Image Retrieval Using Color , Texture and Shape Features," vol. 4, no. 5, pp. 383–389, 2014.
- [12] T. Bui, L. Ribeiro, M. Ponti, and J. Collomosse, "Compact descriptors for sketch-based image retrieval using a triplet loss convolutional neural network," *Comput. Vis. Image Underst.*, 2017.
- [13] J. Kalpana and R. Krishnamoorthy, "Color image retrieval technique with local features based on orthogonal polynomials model and SIFT," *Multimed. Tools Appl.*, vol. 75, no. 1, pp. 49–69, 2016.
- [14] G. Lu and A. Sajjanhar, "Region-based shape representation and similarity measure suitable for content-based image retrieval," *Multimed. Syst.*, vol. 7, no. 2, pp. 165–174, 1999.
- [15] E. Coppa, "On Data Skewness , Stragglers , and MapReduce Progress Indicators," pp. 139–152, 2015.
- [16] S. Kim, X. J. Wang, L. Zhang, and S. Choi, "Near duplicate image discovery on one billion images," in *2015 IEEE Winter Conference on Applications of Computer Vision*, 2015, pp. 943–950.
- [17] W. Zhao, H. Luo, J. Peng, and J. Fan, "MapReduce-based clustering for near-duplicate image identification," *Multimed. Tools Appl.*, vol. 76, no. 22, pp. 23291–23307, 2016.
- [18] L. S. Xue, N. A. A. Majid, and E. A. Sundararajan, "Dynamic virtual machine allocation policy for load balancing using principal component analysis and clustering technique in cloud computing," *J. Telecommun. Electron. Comput. Eng.*, vol. 10, no. 3–2, pp. 47–52, 2018.
- [19] P. Sanguansat, *Principal Component Analysis - Multidisciplinary Applications*. Croatia: InTech, 2012.
- [20] X. Zhu, X. Li, S. Zhang, Z. Xu, L. Yu, and C. Wang, "Graph PCA Hashing for Similarity Search," *IEEE Trans. Multimed.*, vol. 19, no. 9, pp. 2033–2044, 2017.
- [21] S. A. M. N. Yanti Idaya Aspura Mohd Khalid, "The electronic library," *Collect. Manag.*, vol. 35, no. 6, pp. 1191–1214, 2017.
- [22] L. Zheng, Y. Lei, G. Qiu, and J. Huang, "Near-duplicate image detection in a visually salient riemannian space," *IEEE Trans. Inf. Forensics Secur.*, vol. 7, no. 5, pp. 1578–1593, 2012.
- [23] R. Hassanian-esfahani and M. javad Kargar, "Sectional MinHash for near-duplicate detection," *Expert Syst. Appl.*, vol. 99, no. 1 June 2018, pp. 203–212, 2018.
- [24] W. Youzhong, D. Zeng, Z. Xiaolong, and W. Feiyue, "Propagation of online news: Dynamic patterns," in *2009 IEEE International Conference on Intelligence and Security Informatics, ISI 2009*, 2009, pp. 257–259.
- [25] J. Huang, R. Zhang, R. Buyya, J. Chen, and Y. Wu, "Heads-Join: Efficient Earth Mover's Distance Similarity Joins on Hadoop," *IEEE Trans. Parallel Distrib. Syst.*, vol. 27, no. 6, pp. 1660–1673, 2016.
- [26] J. Li, X. Qian, Q. Li, Y. Zhao, L. Wang, and Y. Y. Tang, "Mining near duplicate image groups," *Multimed. Tools Appl.*, vol. 74, no. 2, pp. 655–669, 2014.
- [27] E. Sodic and S. Konjicija, "Phase preserving Fourier descriptor for shape-based image retrieval," *Signal Process. Image Commun.*, vol. 40, pp. 82–96, 2016.
- [28] J. M. Patel and N. C. Gamit, "A review on feature extraction techniques in Content Based Image Retrieval," *2016 Int. Conf. Wirel. Commun. Signal Process. Netw.*, pp. 2259–2263, 2016.
- [29] V. I. Patil and S. Kotaly, "Survey on Content Based Image Retrieval Using Color and Texture Features," *Int. J. Adv. Electron. Comput. Sci. ISSN 2393-2835*, vol. 2, no. 10, pp. 1424–1429, 2015.
- [30] I. T. Jolliffe, *Principal Component Analysis*, Second Edition, vol. 98, no. 3. 2002.
- [31] H. Zhu, Z. Shen, L. Shang, and X. Zhang, "Parallel Image Texture Feature Extraction under Hadoop Cloud Platform," *Springer Int. Publ.*, pp. 459–465, 2014.
- [32] J. M. Francos, A. Narasimhan, and J. W. Woods, "Maximum Likelihood Parameter Estimation of Textures Using a Wold-Decomposition Based Model," *IEEE Trans. Image Process.*, vol.4, no. 12, pp. 1655–1666, 1995.
- [33] C. Wang, N. Komodakis, and N. Paragios, "Markov Random Field Modeling, Inference & Learning in Computer Vision & Image Understanding : A Survey," *Comput. Vis. IMAGE Underst.*, vol. 117, no. 11, pp. 1610–1627, 2013.
- [34] H. Jiang, T. Feng, D. Zhao, B. Yang, L. Zhang, and Y. Chen, "Statistical Fractal Models Based on GND-PCA and Its Application on Classification of Liver Diseases," *Biomed Res. Int.*, vol. 2013, pp. 1–8, 2013.
- [35] S. Deb, *Multimedia systems and content-based image retrieval*. 2004.
- [36] J. Yue, Z. Li, L. Liu, and Z. Fu, "Content-based image retrieval using color and texture fused features," *Math. Comput. Model.*, vol. 54, no. 3–4, pp. 1121–1127, 2011.
- [37] M. K. Alsmadi, "Query-sensitive similarity measure for content-based image retrieval using meta-heuristic algorithm," *J. King Saud Univ. - Comput. Inf. Sci.*, 2017.
- [38] T. S. Kumar, V. V. Kumar, and B. E. Reddy, "Image retrieval based on hybrid features," *ARN J. Eng. Appl. Sci.*, vol.12, no.2, pp.591–598, 2017.

- [39] R. Datta, D. Joshi, J. I. A. Li, and J. Z. Wang, "Image Retrieval: Ideas, Influences, and Trends of the New Age," *J. ACM Comput. Surv.*, vol. 40, no. 2, pp. 1–60, 2008.
- [40] C. Wang, J. Zhang, B. Yang, and L. Zhang, "Sketch2Cartoon: Composing Cartoon Images by Sketching," in *Proceeding MM '11 Proceedings of the 19th ACM international conference on Multimedia*, 2011, pp. 789–790.
- [41] H. Wang and C. Wang, "MindFinder: Interactive Sketch-based Image Search," in *Proceeding MM '10 Proceedings of the 18th ACM international conference on Multimedia*, 2010, pp. 1605–1608.
- [42] L. O. Carvalho, L. F. D. Santos, W. D. Oliveira, A. J. M. Traina, and C. Traina, "Self Similarity Wide-Joins for Near-Duplicate Image Detection," in *Proceedings - 2015 IEEE International Symposium on Multimedia, ISM 2015*, 2015, pp. 237–240.
- [43] D. Feng, W. C. Siu, and H. Zhang, *Multimedia Information Retrieval and Management, 1st (2003)*. New York: Springer, 2003.
- [44] J. Feng, Y. Liu, and L. Wu, "Bag of Visual Words Model with Deep Spatial Features for Geographical Scene Classification," *Comput. Intell. Neurosci.*, vol. 2017 (2017), p. 14, 2017.
- [45] H. Liu, H. Lu, and X. Xue, "SVD-SIFT for web near-duplicate image detection," in *Proceedings - International Conference on Image Processing, ICIP, 2010*, pp. 1445–1448.
- [46] Q. He, Z. Li, and X. Zhang, "Data Deduplication Techniques," in *2010 International Conference on Future Information Technology and Management Engineering*, 2010, pp. 430–433.
- [47] N. Mandagere, P. Zhou, and M. A. Smith, "Demystifying Data Deduplication," pp. 12–17, 2008.
- [48] J. Zhang et al., "IM-Dedup: An image management system based on deduplication applied in DWSNs," *Int. J. Distrib. Sens. Networks*, vol. 2013, 2013.
- [49] J. Zhang, T. Li, and Y. Pan, "Parallel rough set based knowledge acquisition using MapReduce from big data," in *BigMine '12 Proceedings of the 1st International Workshop on Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications*, 2012, pp. 20–27.
- [50] L. Chen and F. Stentiford, "Comparison of near-duplicate image matching," *Cvmp06*, 2006.
- [51] S. H. Srinivasan and N. Sawant, "Finding Near-duplicate Images on the Web using Fingerprints," in *Proceeding of the 16th ACM international conference on Multimedia MM 08*, 2008, pp. 881–884.
- [52] M. Chen, Y. Wang, X. Zou, S. Wang, and G. Wu, "A duplicate image deduplication approach via Haar wavelet technology," *Proc. - 2012 IEEE 2nd Int. Conf. Cloud Comput. Intell. Syst. IEEE CCIS 2012*, vol. 2, pp. 624–628, 2013.
- [53] S. G. Lakshmi and N. R. Gayathiri, "a Framework for Hosting Image Compression in Cloud," *Int. J. Comput. Sci. Mob. Comput.*, vol. 3, no. 3, pp. 845–848, 2014.
- [54] F. Zou, Y. Chen, J. Song, K. Zhou, Y. Yang, and N. Sebe, "Multiple Kernel Hashing," *IEEE Trans. Multimed.*, vol. 17, no. 7, pp. 1006–1018, 2015.
- [55] M. K. Alsmadi, "An efficient similarity measure for content based image retrieval using memetic algorithm," *Egypt. J. Basic Appl. Sci.*, vol. 4, no. 2, pp. 112–122, 2017.
- [56] O. Chum, J. Philbin, and A. Zisserman, "Near Duplicate Image Detection: min-Hash and tf-idf Weighting," in *Proceedings of the British Machine Vision Conference*, 2008, vol. 810, pp. 812–815.
- [57] D. P. Tian, "A review on image feature extraction and representation techniques," *Int. J. Multimed. Ubiquitous Eng.*, vol. 8, no. 4, pp. 385–395, 2013.
- [58] D. A. Lusin, M. A. Mattar, M. B. Blaschko, E. G. Learned-Miller, and M. C. Benfield, "Combining Local and Global Image Features for Object Class Recognition," *2005 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. - Work.*, vol. 3, pp. 47–47, 2005.
- [59] L. Bueno, E. Valle, and R. da S. Torres, "Bayesian approach for near-duplicate image detection," in *Proceeding ICMR '12 Proceedings of the 2nd ACM International Conference on Multimedia Retrieval*, 2012, pp. 1–8.
- [60] W. Dong et al., "High-Confidence Near-Duplicate Image Detection," in *Proceeding ICMR '12 Proceedings of the 2nd ACM International Conference on Multimedia Retrieval*, 2012, pp. 3304–3311.
- [61] H. Wang, Y. Shen, L. Wang, K. Zhufeng, W. Wang, and C. Cheng, "Large-Scale Multimedia Data Mining Using MapReduce Framework," in *2012 IEEE 4th International Conference on Cloud Computing Technology and Science*, 2012, pp. 287–292.
- [62] Z. Li and X. Feng, "Near duplicate image detecting algorithm based on bag of visual word model," *J. Multimed.*, vol. 8, no. 5, pp. 557–564, 2013.
- [63] G. Kalaiarasi and K. K. Thyagarajan, "Visual Content Based Clustering of Near-Duplicate Web Search Images," in *2013 International Conference on Green Computing, Communication and Conservation of Energy (ICGCE)*, 2013, vol. 6, pp. 767–771.
- [64] J. Wang, "Strong Geometry Consistency for Large Scale Partial-Duplicate Image Search," in *MM '13 Proceedings of the 21st ACM international conference on Multimedia*, 2013, pp. 633–636.
- [65] L. Li, S. Jiang, Z. J. Zha, Z. Wu, and Q. Huang, "Partial-Duplicate Image Retrieval via Saliency-Guided Visual Matching," *IEEE Multimed.*, vol. 20, no. 3, pp. 13–23, 2013.
- [66] S. Battiato et al., "Aligning codebooks for near duplicate image detection," *Multimed. Tools Appl.*, vol. 72, no. 2, pp. 1483–1506, 2014.
- [67] L. Liu, Y. Lu, and C. Y. Suen, "Variable-length signature for near-duplicate image matching," *IEEE Trans. Image Process.*, vol. 24, no. 4, pp. 1282–1296, 2015.
- [68] F. Nian, T. Li, X. Wu, Q. Gao, and F. Li, "Efficient near-duplicate image detection with a local-based binary representation," *Multimed. Tools Appl.*, vol. 75, no. 5, pp. 2435–2452, 2015.
- [69] H. Wang, F. Zhu, B. Xiao, L. Wang, and Y. Jiang, "GPU-based MapReduce for large-scale near-duplicate video retrieval," *Multimed. Tools Appl. J.*, vol. 74, no. 23, pp. 10515–10534, 2015.
- [70] N. Yusof, A. Ismail, N. Aini, and A. Majid, "A Hybrid Model for Near-Duplicate Image Detection in MapReduce Environment," *TEM J.*, vol. 8, no. 4, pp. 1252–1258, 2019.
- [71] R. Raieli, *Multimedia Information Retrieval (Theory and Techniques)*, vol. 1, no. 1, 2013.
- [72] M. H. Safar and C. Shahabi, *Shape Analysis and Retrieval of Multimedia Objects*, First Edit. USA: Kluwer Academic Publishers, 2003.