Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control
Journal homepage: http://kinetik.umm.ac.id
ISSN: 2503-2267
Vol. 5, No. 2, May 2020, Pp. 87-94

Image retrieval based on texton frequency-inverse image frequency

Yufis Azhar*[1], Agus Eko Minarno[2], Yuda Munarko[3], Zaidah Ibrahim[4]
Universitas Muhammadiyah Malang, Indonesia[1,2,3]
Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia[4]

Article Info

Keywords:
Image Retrieval, Feature Weighting, Texton Co-occurrence

Article history:
Received 27 December 2019
Revised 13 April 2020
Accepted 13 April 2020
Available 20 May 2020

Cite:
Azhar, Y., Minarno, A., Munarko, Y., & Ibrahim, Z. (2020). Image Retrieval Based on Texton Frequency-Inverse Image Frequency. Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control, 5(2). doi:https://doi.org/10.22219/kinetik.v5i2.1026

*Corresponding author.
Yufis Azhar
Email address:
yufis.az@gmail.com

Abstract

In image retrieval, the user hopes to find the desired image by entering another image as a query. In this paper, the approach used to find similarities between images is feature weighting, where between one feature with another feature has a different weight. Likewise, the same features in different images may have different weights. This approach is similar to the term weighting model that usually implemented in document retrieval, where the system will search for keywords from each document and then give different weights to each keyword. In this research, the method of weighting the TF-IIF (Texton Frequency-Inverse Image Frequency) method proposed, this method will extract critical features in an image based on the frequency of the appearance of texton in an image, and the appearance of the texton in another image. That is, the more often a texton appears in an image, and the less texton appears in another image, the higher the weight. The results obtained indicate that the proposed method can increase the value of precision by 7% compared to the previous method.

1. Introduction

Currently, the development of image retrieval methods can be classified into two categories, namely CBIR (Content-Based Image Retrieval) and TBIR (Text-Based Image Retrieval) [1]. In CBIR, an image search technique that is similar to a query done by recognizing the patterns contained in the two images being compared. This pattern can be formed through the extraction of color features, textures, margins, and so on [2][3][4][5][6][7]. However, this method is more suitable if implemented on object detection. Because in CBIR, the similarity of patterns is an absolute requirement that must be met to be able to say that the two images have similarities. Terms are not suitable if implemented in image retrieval, because each user can use a different query image to find the same image. Likewise, a user can use the same query to find a different image. Figure 1 shows an example of the disadvantages of using CBIR techniques in image retrieval.

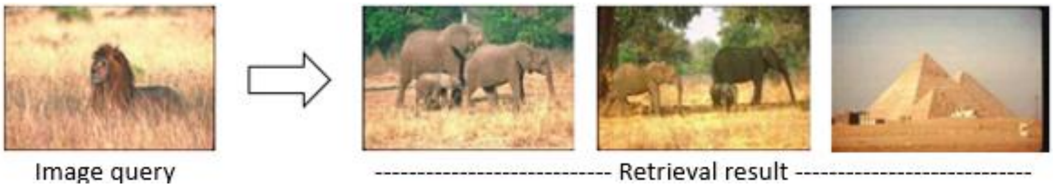


Image query
---------------------------- Retrieval result ----------------------------

Figure 1. Errors that Often Occur in CBIR Technique

In the example above, the user uses the image of a lion sitting on a dry grass field as a query to find an image containing another lion. However, because what is more prominent in the image is the golden yellow pasture, CBIR will display another image that has a similar pattern to the query image, i.e. the image is dominated by golden yellow. This is one of the disadvantages of using CBIR in image retrieval.

Some features that are often used to perform image retrieval in CBIR methods include color, shape, texture, and depth. The color feature is the most popular and most accessible to use to describe an image. J. Huang uses this feature to index an image dataset. He uses a 3-dimensional table based on the color and distance between pixels in an image to compile the index table [8]. The use of color to look for similarities between the two images is done by comparing the color channels of the pixels owned by the first image and the second image. If many pixels have the same color, then the image is said to be similar.

For specific images, batik, for example, color features are less reliable. This is because the colors used in batik images tend to be uniform. Moreover, what distinguishes between 1 type of batik with other types of batik is not the color, but the style or motif. Therefore, besides color, texture features also have an essential role in this case. Julesz researched the use of texture features in image retrieval. He analyzed the interaction of texton to recognize patterns in an image. A texton can be composed of several pixels. The results showed that the use of simple statistical methods in texton could be beneficial for extracting textures in an image [9]. Other studies on the use of textures and image retrieval were carried out by Liu. In his paper, Liu proposed a matrix based on the appearance of a texton in an image. Liu used 4 Textons in his research. The matrix produced from the four Textons then arranged into a histogram.

The histogram is then given the name MTH (Multi Texton Histogram) [10]. The weakness of the MTH method proposed by Liu is that it cannot wholly describe the texture in an image. This is because MTH only uses local features, and when searching for pixel pairs that appear together using texton, there is missing information, which may be useful for describing an image. The weakness of MTH was later refined by Agus Eko in his research. Agus Eko proposed the Multi Texton Co-occurrence Descriptor (MTCD) method based on MTH. MTCD uses 6 Texton to extract texture features in an image. With the use of more texton, MTCD managed to overcome the weaknesses of MTH [11].

Whereas in TBIR, each image is given a label or caption in the form of text that contains an explanation of the image. With the help of this text, the system can determine the right image to display to the user [12][13][14][15][16]. Although this technique can provide an exact retrieval result, the necessity to give a caption to each text is a weakness of the TBIR technique. Attempts to adapt the method that is usually done to text retrieval in image retrieval have been made several times. Yu Suzuki took a very similar approach to this research, namely adapting the term weighting model in text retrieval. However, Yu only paid attention to the color histogram that had gone through the previous quantification process. The model proposed by Yu proved to be a failure because the resulting precision value was shallow [1].

In this research, a method of merging between CBIR and TBIR is proposed to conduct image retrieval. Where an image will be identified through a feature extraction approach with the appearance of texton (texton co-occurrence), then, the extracted feature will be weighted using the term weighting model, which is often used in text retrieval. However, because the feature used is the appearance of texton in the image, the weighting method proposed in this research is given the name TF-IIF (Texton Frequency-Inverse Image Frequency). This method will extract critical features in an image based on the frequency of the appearance of texton in an image, as well as the appearance of the texton in another image. That is, the more often a texton appears in an image, and the less texton appears in another image, the higher the weight.

2. Research Method

In general, the system workflow starts with processing the image of the dataset using the proposed method. This processing will produce indexing tables which will be saved in CSV format. The same thing is done for the query image. Processing the query image using the proposed method will produce a vector, or commonly called the Vector Space Model (VSM). Later, the vector will look for similarities with other vectors in the indexing table using cosine similarity. Figure 2 and Figure 3 will show the system workflow in more detail.

Figure 3 shows in more detail about the method proposed in this research. As discussed in the previous chapter, in terms of feature extraction, this method is inspired by the MTCD method [11]. In MTCD, there are three features extracted, namely color, edge and texture. However, in the method proposed in this research, only two features are used (namely color and edge). Texture feature is not used because, in MTCD, the texton detection process is only done on these two features.

There are four main stages in the method proposed in this research, namely quantification, identification of texton, texton weighting, and calculating similarity. The following will be explained in more detail about these stages.

2.1 Quantization

The quantization is the process of simplifying or grouping the values of a pixel in an image. This process is carried out to reduce the variation in values in an image as is known, that a color image is composed of 3 color layers, namely red layer, green layer, and blue layer. Each pixel in each of these layers has a range of values between 0 to 255, where 0 means black and 255 means red, green, or blue.

A color image will have very different pixel values (0-255). This will be quite difficult when the image is compared in similarity with other images. Therefore, this quantification process is needed. In the process of quantification, the value of each pixel in an image will be grouped into several bins. So that the variation in value will decrease. Figure 4 below will show the image quantification process that has been described previously.

In Figure 4, for example, bin number 8. That is to say, the value of each pixel will be grouped into eight sections. Because each pixel has values ranging from 0 to 255, this means: values from 0 to 31 will be changed to 1, values from 32 to 63 will be changed to 2, values from 64 to 95 will be changed to 3, values from 96 to 127 will be changed the value is 4, the value 128 - 159 will be changed to 5, the value 160-191 will be changed to 6, the value 192 - 223 will be changed to 7, and the value 224 - 255 will be changed to 8. Referring to research [11], the number of bin that will be used in this paper is 4 for color features, and 18 for edge features.

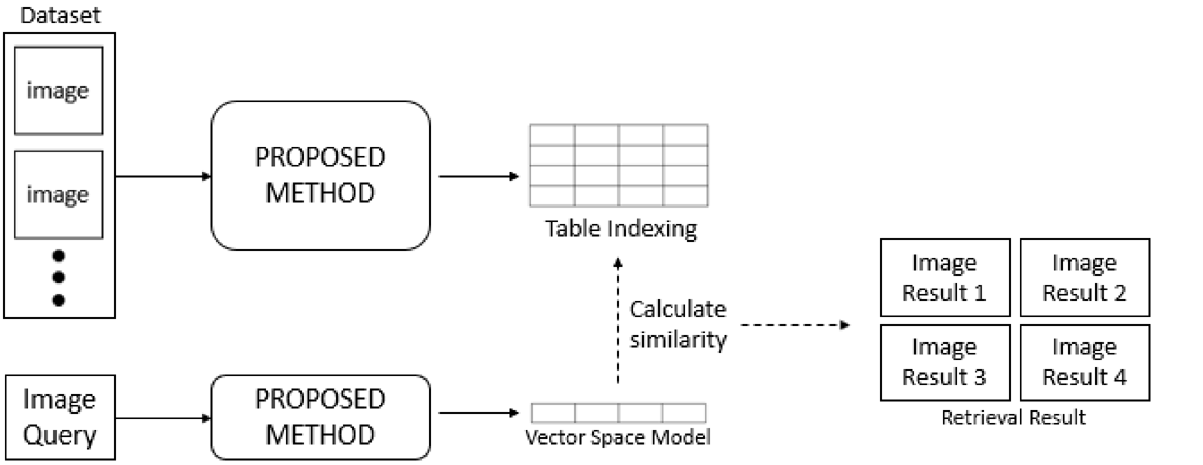


Figure 2. Proposed Method Position in Research Diagram

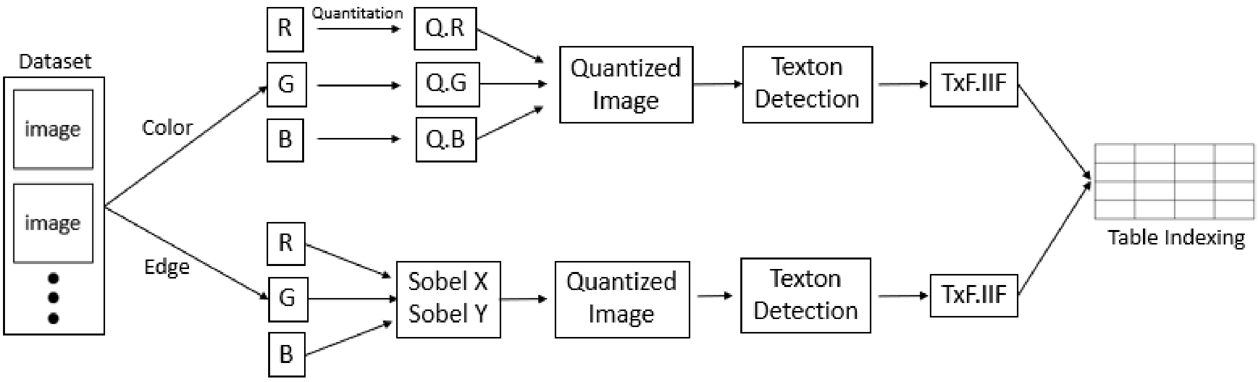
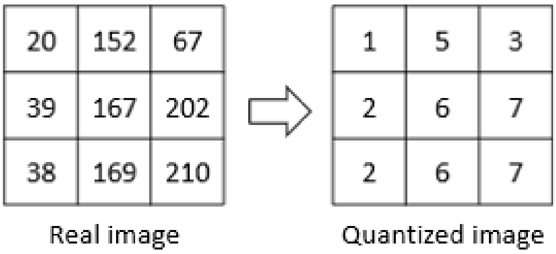


Figure 3. Proposed Method Diagram



20	152	67
39	167	202
38	169	210

Real image

1	5	3
2	6	7
2	6	7

Quantized image

Figure 4. The Example of the Quantization Method

This quantification process will be carried out on the two features used in this research, namely the color features (color) and the edge (edge) by first splitting the image into three layers (red, green, blue). In this research, the Sobel method is used to extract the edge features. After the three layers have been quantified, the three will be combined again until they form 1 whole image again. This merging process is done using Equation 1. This merged image will later be used as input parameters in the texton identification stage.

$$c = (b \times g \times r) + (b \times g) + b \qquad (1)$$

Where c is the combined pixel value, b is the pixel value on the blue layer, g is the pixel value on the green layer, and r is the pixel value on the red layer.

Cite: Azhar, Y., Minarno, A., Munarko, Y., & Ibrahim, Z. (2020). Image Retrieval Based on Texton Frequency-Inverse Image Frequency. Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control, 5(2). doi: https://doi.org/10.22219/kinetik.v5i2.1026

2.2 Texton Identification

The next stage is texton identification. Texton is a pattern that will later be convoluted in the image to detect the appearance of a similar value on neighbouring pixels. Six types of Texton will be used in this research, as can be seen in Figure 5. While the process of Texton identification can be seen in Figure 6.

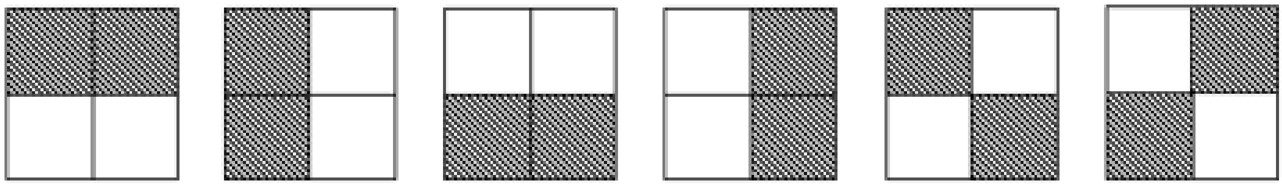


Figure 5. Textons that Used in this Research

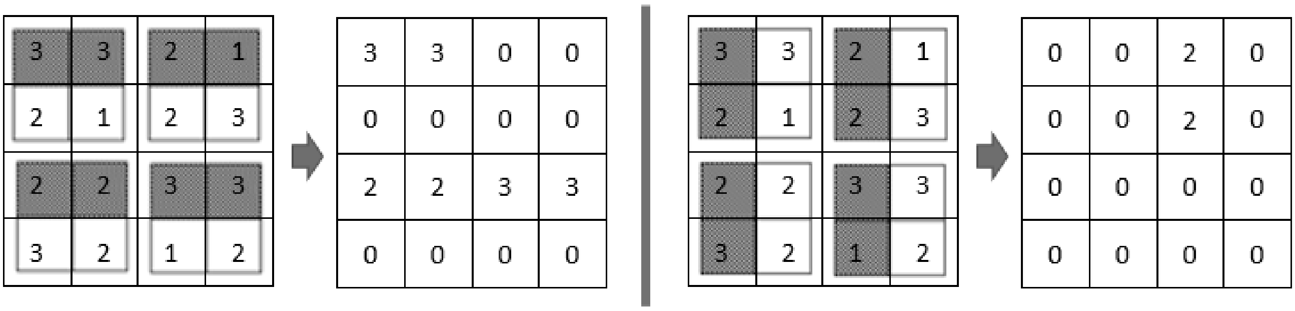


Figure 6. The Convolution Process Uses Texton Type 1 and 2

In Figure 6, it is shown the process of identifying texton used in this research. A zero matrix will be created first with the same dimensions as the original image. Then each type of Texton will be convoluted from the top left end of an image to the bottom right end. If there are pixels that have the same value according to the texton pattern used, then that value will be moved to the zero matrices that have been prepared. The final result of this process is six matrices that already contain the values resulting from the identification of texton. These matrices will be used at a later stage.

2.3 Texton Weighting

This stage was inspired by Yu's research on the adaptation of term weighting to image retrieval [1]. Each texton that has been identified in the previous stage will act as a feature that will be weighted using Equation 2.

$$W_i = TF_i \times log\frac{N}{IF_i} \quad (2)$$

where, W_i is the weight of texton *i*, TF_i is the frequency of appearance of texton *i* in an image, N is the sum of all images in the dataset, and IF_i is the number of images containing texton *i*. Thus, the number of bin used at the quantization stage, as well as the number of textons used at the texton identification stage will affect the number of features *i* weighted. In this paper, a bin number of 4 is used for color features (after combined, the 64 different values in a pixel is created), and 18 for edge features. While the number of textons used to identify co-occurrence is 6. Therefore, the total number of features to be weighted is (64 x 6) + (18 x 6) = 492 features. After calculating the weights of all identified feature, the weights will be arranged in an indexing table to facilitate the process of calculating the similarity at a later stage.

2.4 Calculate Similarity

The final step is to calculate the similarity between the query image and the image in the dataset. To calculate it, the Cosine Similarity method is used in Equation 3 below.

$$sim_{(d,q)} = \frac{\sum_{k=1}^{i}\left(W_{dk} \times W_{qk}\right)}{\sqrt{\sum_{k=1}^{i}\left({W_{dk}}^2 \times {W_{qk}}^2\right)}} \quad (3)$$

where, $sim_{(d,q)}$ is the similarity value between the image of the query *q* with the dataset image *d*, W_{dk} is the weight of texton *k* in image *d*, and W_{qk} is the weight of texton *k* in image *q.*

3. Results and Discussion

3.1 Dataset

There are two datasets used in this research. The first dataset is Batik300 which has 300 batik images divided into 50 categories (each category has six images). This dataset was chosen because batik image has quite complex and repetitive details so that it is considered suitable for the application of the proposed method. Examples of images contained in this dataset can be seen in Figure 7.

The second dataset is the Corel-10k dataset. The Corel-10k dataset has 100 categories. In total there are 10,000 images from diverse content such as beach, elephant, sunset, flower, car, fruit, food, lion, wolf, fish, mountain, door, etc. Each category consists of 100 images with a size of 192 x 128 or 128 x 192 pixels. All images in this dataset are in JPEG format. Figure 8 shows some examples of images in the Corel-10k dataset.

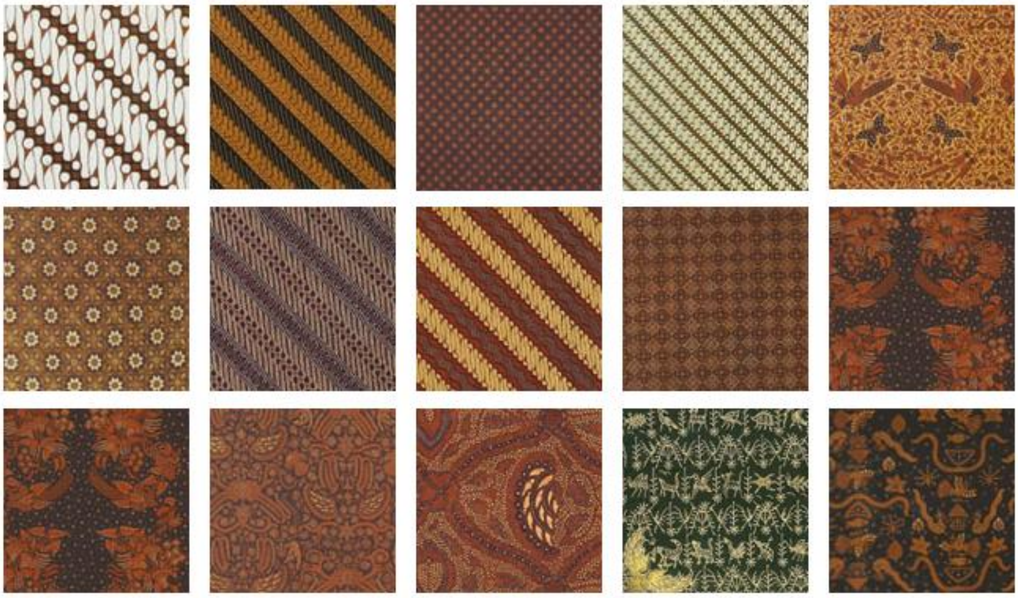


Figure 7. The Example of Images in the Batik 300 Dataset

Figure 8. The Example of Images in the Corel-10k Dataset

3.2 Test Results

There are two test scenarios conducted in this paper. The first scenario is to compare the precision value of the results of the TF-IIF method with MTH-4 (Multi Texton Histogram - 4 textons) [17][18][19] and MTH-6 (Multi Texton Histogram - 6 textons) [20][21][22]. The MTH method extracts color and edge features from an image, then uses the help of Texton to detect the appearance of the same value from neighboring pixels. After that, the similarity between the query image and image in the dataset will be measured using Canberra distance. The difference between MTH-4 and MTH-6 only lies in the number of textons used.

Table 1. The Measurement Results of the Precision Value of TF-IIF vs MTH

Method	Precision	
	Batik300	Corel10k
MTH-4	0.85	0.44
TF-IIF (4 textons)	0.86	0.49
MTH-6	0.88	0.47
TF-IIF (6 textons)	0.90	0.53

This first test uses two different datasets, namely Batik300 and Corel-10k. For Batik300, the query image is arranged by taking one image from each category. So that it will be composed of a set of query images totaling 50 pieces. As for Corel-10k, ten images were taken from each category to create a 1000 image query set. The precision value is calculated by observing the 5 top-ranked images for Batik300 dataset, and 12 top-ranked images for Corel10k dataset that were successfully retrieved by the method used.

Table 1 shows the results of the tests carried out. It can be seen that the TF-IIF method is able to outperform the MTH method, both when paired with 4 Texton, and 6 Texton. In fact, in the Corel-10k case, the TF-IIF method was able to outperform the MTH method even though the number of textons used was smaller.

The second test carried out was to pair the feature extracted by the TF-IIF method with the feature extracted from the GLCM (Gray Level Co-occurrence Matrix) method. GLCM is a first and second-order statistical method used by Haralick to extract texture features from an image [23]. The GLCM method used in this second test refers to Minarno's research which produces four values, namely entropy, correlation, energy, and contrast [24][25].

Table 2. The Test Results when the TF-IIF is Paired with the GLCM Method

Method	Precision	
	Batik300	Corel10k
GLCM	0.88	0.51
MTCD (MTH-6 + GLCM)	0.92	0.55
TF-IIF + GLCM	0.93	0.62

Just like in the first test, in this second test, Batik300 and Corel-10k datasets were re-used. Table 2 shows the results of the comparison performed. It can be seen that the TF-IIF method is also able to outperform both the GLCM method and the MTCD method. To note, MTCD method is a feature extraction method that combines the MTH-6 method with the GLCM method [11]. The total features extracted by the GLCM method are 16 features (4 features for entropy, 4 for correlation, 4 for energy, and 4 for contrast). While on MTCD, the total features extracted amounted to 98 (64 color features, 18 edge features, and 16 features from GLCM). While for TF-IIF + GLCM, 508 extracted features (384 colour features, 108 edge features, and 16 features from GLCM) were produced.

4. Conclusion

The TF-IIF method proposed in this paper is a combination of two approaches in image retrieval, namely CBIR and TBIR. The TF-IIF method still retains the characteristic of CBIR, namely the extraction of features from the content of an image, then combined with the method of weighting the features usually performed on the TBIR approach. The results obtained from the test show that the TF-IIF method is able to outperform several previous methods, both for images that have a repetitive pattern (such as batik images), as well as for varied images (such as Corel-10k).

References

[1] K. K. Yu Suzuki, Masahiro Mitsukawa, “A Image Retrieval Method Using TFIDF Based Weighting Scheme,” in *19th International Conference on Database and Expert Systems Application*, Pp. 112–116, 2008. https://doi.org/10.1109/DEXA.2008.106

[2] S. G. Sanu and P. S. Tamase, “Satellite image mining using content based image retrieval,” *Int. J. Eng. Sci*, 2017.

[3] Z. Xia, N. N. Xiong, A. V. Vasilakos, and X. Sun, “EPCBIR: An efficient and privacy-preserving content-based image retrieval scheme in cloud computing,” *Inf. Sci. (Ny).*, Vol. 387, Pp. 195–204, 2017. https://doi.org/10.1016/j.ins.2016.12.030

[4] S. R. Dubey, S. K. Singh, and R. K. Singh, “Multichannel decoded local binary patterns for content-based image retrieval,” *IEEE Trans. image Process.*, Vol. 25, No. 9, Pp. 4018–4032, 2016. https://doi.org/10.1109/TIP.2016.2577887

[5] N. Pasumarthi and L. Malleswari, “An empirical study and comparative analysis of Content Based Image Retrieval (CBIR) techniques with various similarity measures.,” 2016. https://doi.org/10.1049/cp.2016.1529

[6] Z. Liu, Z. Zhao, and M. Larson, “Who’s Afraid of Adversarial Queries?: The Impact of Image Modifications on Content-based Image Retrieval,” in *Proceedings of the 2019 on International Conference on Multimedia Retrieval*, Pp. 306–314, 2019. https://doi.org/10.1145/3323873.3325052

[7] A. Nazir, R. Ashraf, T. Hamdani, and N. Ali, “Content based image retrieval system by using HSV color histogram, discrete wavelet transform and edge histogram descriptor,” in *2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, Pp. 1–6, 2018. https://doi.org/10.1109/ICOMET.2018.8346343

[8] J. Huang, S. R. Kumar, M. Mitra, and W.-J. Zhu, “Image Indexing Using Color Correlograms,” *U.S. Pat.*, Vol. 2, No. 12, Pp. 12–15, 2011. https://doi.org/10.1109/CVPR.1997.609412

[9] B. Julesz, “Textons, the Elements of Texture Perception, and Their Interactions,” *Nat. 290*, No. 5802, Pp. 91, 1981. https://doi.org/10.1038/290091a0

[10] G.-H. Liu and J.-Y. Yang, “Image Retrieval Based on the Texton Co-Occurrence Matrix,” in *Pattern Recognition 41*, pp. 3521–3527, 2008. https://doi.org/10.1016/j.patcog.2008.06.010

[11] A.-E. Minarno and S. Nanik, “Image retrieval using multi texton co-occurrence descriptor,” *J. Theor. Appl. Inf. Technol.*, Vol. 67, No. 1, Pp. 103–110, 2014.

[12] S. A. J. Zaidi, A. Buriro, M. Riaz, A. Mahboob, and M. N. Riaz, “Implementation and comparison of text-based image retrieval schemes,” *Int. J. Adv. Comput. Sci. Appl.*, Vol. 10, No. 1, Pp. 611–618, 2019. https://dx.doi.org/10.14569/IJACSA.2019.0100177

[13] M. Kiruthika and S. Sukumaran, “A Comparative Study Of Image Retrieval Techniques,” *IJRAR-International J. Res. Anal. Rev.*, Vol. 6, No. 2, Pp. 252–255, 2019.

[14] S. Dey, A. Dutta, S. K. Ghosh, E. Valveny, J. Lladós, and U. Pal, “Learning cross-modal deep embeddings for multi-object image retrieval using text and sketch,” in *2018 24th International Conference on Pattern Recognition (ICPR)*, Pp. 916–921, 2018.

[15] K. H. Tan, *Text-based Image Retrieval Using Image Captioning*. 2019.

[16] A. Li, J. Sun, J. Yue-Hei Ng, R. Yu, V. I. Morariu, and L. S. Davis, “Generating holistic 3d scene abstractions for text-based image retrieval,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Pp. 193–201, 2017.

[17] G.-H. Liu, L. Zhang, Y.-K. Hou, Z.-Y. Li, and J.-Y. Yang, “Image Retrieval Based on Multi-Texton Histogram,” in *Pattern Recognition 43*, Pp. 2380–2389, 2010. https://doi.org/10.1016/j.patcog.2010.02.012

[18] A. Jayachandran and R. Dhanasekaran, “Automatic detection of brain tumor in magnetic resonance images using multi-texton histogram and support vector machine.,” *Int. J. Imaging Syst. Technol.*, Vol. 23, No. 2, Pp. 97–103, 2013. https://doi.org/10.1002/ima.22041

[19] A. Jayachandran and R. Dhanasekaran, “Brain tumor severity analysis using modified multi-texton histogram and hybrid kernel SVM,” *Int. J. Imaging Syst. Technol.*, Vol. 24, No. 1, Pp. 72–82, 2014. https://doi.org/10.1002/ima.22081

[20] M. Ambarawati and R. D. Agustin, “Two Dimensional Object in square and rectangles: Batik artwork approach,” *J. Phys. Conf. Ser.*, Vol. 1339, No. 1, 2019. https://doi.org/10.1088/1742-6596/1339/1/012128

[21] R. E. Caraka and T. W. C. B. Pardamean, “Batik Parang Rusak Detection Using Geometric Invariant Moment,” in *2018 Indonesian Association for Pattern Recognition International Conference (INAPR)*, Pp. 71–74, 2018. https://doi.org/10.1109/INAPR.2018.8627000

[22] A. E. Minarno, A. S. Maulani, A. Kurniawardhani, and F. Bimantoro, “Comparison of Methods for Batik Classification Using Multi Texton Histogram,” *TELKOMNIKA*, Vol. 16, No. 3, Pp. 1358–1367, 2018. http://dx.doi.org/10.12928/telkomnika.v16i3.7376

[23] R. M. Haralick, “Statistical and structural approach to texture,” *Proceeding IEEE* Vol. 67, No. 5, Pp. 786–804, 1979. https://doi.org/10.1109/PROC.1979.11328

[24] A. E. Minarno and N. Suciati, “Batik Image Retrieval Based on Color Difference Histogram and Gray Level Co-Occurrence Matrix,” *TELKOMNIKA (Telecommunication Comput. Electron. Control.*, Vol. 12, No. 3, Pp. 597–604, 2014. http://dx.doi.org/10.12928/telkomnika.v12i3.80

[25] A. E. Minarno, A. Kurniawardhani, and F. Bimantoro, “Image Retrieval Based on Multi Structure Co-occurrence Descriptor,” *TELKOMNIKA (Telecommunication Comput. Electron. Control.*, Vol. 14, No. 3, Pp. 1175–1182, 2016. http://dx.doi.org/10.12928/telkomnika.v14i3.3292

Cite: Azhar, Y., Minarno, A., Munarko, Y., & Ibrahim, Z. (2020). Image Retrieval Based on Texton Frequency-Inverse Image Frequency. Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control, 5(2). doi:https://doi.org/10.22219/kinetik.v5i2.1026