Performance analysis of shape, color and texture features on tracking information face based on CBIR

J Jumi*, A Zaenuddin and T Mulyono
Politeknik Negeri Semarang, Semarang, Indonesia

Abstract. The image has the features of shape, color and texture that are vary. Each feature has a different performance in supporting the accuracy of information retrieval using a process approach to CBIR (Content-Based Image Retrieval). On the image with different objects different performance will be generated on each feature. For example, that the performance features of the form of the more dominant compared features color and texture on the image with the face, while the object on the image with the object of interest feature is more dominant than the features of texture and shape. In this research was conducted on the analysis of the performance features of the shape, color and texture in supporting the accuracy of a search using the approach of CBIR (Content Based Image Retrieval). The method used are invariant moment, color moment and GLCM (Grey Level Co-occurrence Matrix). The results showed that the best search accuracy is 95%, where the features of shape has a performance by 50%, 30% color feature and texture feature by 20% with 600 test image with object database face.

1. Introduction
Image is a data that describes the physical form of an object. Imagery through its feature content has uniqueness or characteristic that can distinguish between object one with other object. This allows the image to be an alternative key to information retrieval other than using a text field key. The search model using the image as a key field can be done through the CBIR (Content Based Image Retrieval) approach using the features owned by each image.

In the development of research on content-based retrieval image or called Content Based Image Retrieval (CBIR) has done some retrieval method based on image features. Research on CBIR using texture features as a similarity analysis method or similarity as a basis for retrieval has been done by producing 54% similarity precision value [1]. The use of combination method of color and shape features has been done that applied to the medical database [2]. CBIR research by combining more than two image features is also done using the features of shapes, colors, textures and edges with a value of retrieval accuracy of more than 75% [3].

2. Related works
Several CBIR studies using a combination of 3 features yielded a precision value of more than 60% [4], subsequently similar studies have resulted in a recall value of test results using Wang data being over 85% [5], followed by similar studies with value results precision and recall more than 90% [6]. Research using the combination and weighting of image features on indoor asset imagery has also been done by
generating retrieval accuracy varies according to the percentage of weighting of the three features, in which the shape features are more dominant than the color features and texture features [7].

From some research above can be concluded that combination of some image feature will yield level of accuracy of resemblance higher than without doing combination. Thus each feature has different performance in its support for improving the accuracy of the search by CBIR. The extraction method used also affects the retrieval accuracy by the CBIR approach.

In this research, the performance analysis of each feature is a feature of shape, color and texture to support the improvement of retrieval accuracy. Each feature has a different percentage of support for CBIR-based information retrieval. In the image of different objects will be different percentage of performance of each feature generated.

3. Research methods
Some research that has been done have not done performance analysis of shape, color and texture in support of CBIR image search accuracy improvement. In this research have done retrieval image using three features with research step shown by Figure 1.

![Figure 1. An overview of CBIR retrieval stages using CBIR approach using texture, color and shape features.](image)

Details of the stages of this study are as follows:

3.1. Preprocessing
Preprocessing is the improvement of image quality prior to feature extraction with the aim of improving image feature performance in support of retrieval accuracy. There is a preprocessing difference in feature form extraction with color and texture feature extraction. The preprocessing difference is to get a quality image before feature extraction.

3.1.1. Preprocessing features of the texture and color features
- Resize: The larger of image size so time extraction is also longer, so it needs resize stages to speed up the computation process. At this stage the image resized to 300 x 300 pixels.
- Grayscale: Preprocessing grayscale is done for texture feature extraction to make texture more visible, whereas in color feature extraction not done because the color value to be calculated.
• Histogram Equalization: At this stage is done histogram alignment so that the image quality becomes more contrast to get the value of color features and texture quality. The result of histogram alignment can be seen in Figure 2.

Figure 2. Comparison of images before and after the histogram process, (a) Original Images, (b) Image of Histogram Equalization (HE).

3.1.2. Preprocessing of the shape feature

• Grayscale: Especially on the calculation of the features of the color element form is not taken into account, so that the change of image to grayscale for faster computation process.
• Resize: Image size will affect the length of feature extraction time, the smaller the size, the faster the computation of its feature extraction. At this stage the image resized to 300 x 300 pixels.
• Edge Enhancement dan Histogram Equalization (HE): The edge enhancement stage will produce a new image with a clearer or sharper edge of the object, so that the shape of the image will be clearer. This stage uses convolution method with Sobel operator [8]. The edge sharpening result can be seen in Figure 3.

Figure 3. Preprocessing stages of shape features, (a) Histogram equalization image, (b) Grayscale, (c) Resize.

3.2. Extraction of the texture feature
Texture is the regularity of certain patterns formed from the arrangement of pixels on an image. Texture value can be used as one of the variables for measuring the similarity of image. There are three algorithms for texture processing in image that is structural, spectral and statistical [8]. The first and second-order stats commonly referred to as Gray Level Co-Occurrence Matrix (GLCM) are one of the
most commonly used feature extraction methods. First-order feature extraction is a feature retrieval method based on the characteristics of an image histogram. The houghrogram shows the probability of occurrence of the pixel gray value of an image. From the value of the resulting histogram then calculated several dimensions of the first order feature i.e. mean, skewness, variance, kurtosis and entropy.

In some cases, the use of first order cannot be used to distinguish between objects one with other objects, so it is necessary to use GLCM on feature extraction feature. Haralick et al have proposed GLCM through extraction using a kookurency matrix which is an intermediate matrix that represents the neighboring relationship between pixels in images in different orientation directions and spacetime spacing as in Figure 4 [9].

![Figure 4. Neighboring relationships between pixels.](image)

One technique that can be used to get a second-order statistical feature is to calculate the probability of an adjacency relationship between two pixels at a certain distance and angle orientation [8]. The pixel connectivity relationship used in this study is using angle orientation in four directions that is $0^\circ, 45^\circ, 90^\circ, 135^\circ$. While the kookurensi matrix is a square matrix with the number of elements as much as the square of the number of pixel intensity levels in the image.

After obtaining the kookurency matrix, it can be calculated second order statistic feature which represents image and subsequently extracted texture feature using kookurensi matrix [9]. The second-order statistical features calculated in this study are energy, homogeneity, contrast and correlation. The energy dimension is calculated using equations 1, 2, 3.

\[
P(i,j|d,\theta) = \frac{P(i,j|d,\theta)}{\sum_i \sum_j P(i,j|d,\theta)}
\]

\[
Energy = \sum_i \sum_j P(i,j)^2
\]

Energy to measure concentration of intensity pairs on the GLCM matrix.

\[
Homogeneity = \sum_i \sum_j \frac{P(i,j)}{1 + |i-j|}
\]

While homogeneity indicates homogeneity of variation in image intensity. The Contrast is calculated using equation 4.

\[
Contrast = \sum_{i,j} P(i,j)|i-j|^2
\]

Contrast shows the size of the spread of matrix elements. If the location is far from the main diagonal, then the value of contrast will be great. Next Correlation is calculated using equation 5.

\[
Correlation = \frac{(1 - \mu_i)(1 - \mu_j)P(i,j)}{\sigma_i \sigma_j}
\]

Correlation shows the size of the gray-level linear dependence of the neighboring pixels in a gray image. The stages of image search based on texture features are:
- Read the query image
- Convert RGB image to Grayscale
- Calculate 4-way GLCM matrix at 0°, 45°, 90°, 135°
- Calculate in each direction for energy value, homogeneity, contrast and correlation.
- Make matching similarity between the image of the query with the image on the database
- Select the minimum distance value as the most similar image.

3.3. Extraction of the shape feature

Feature feature extraction is done using invariant moment method. This method is used because it is not vulnerable to image changes caused by Rotation, Scale and Translation (RST) [10].

Invariant Moment to result seven constant invariant moment values against RST [11]. In the invariant moment can also be analyzed form the set of the moment of a function f(x, y) of two variables defined as follows: [8]

Moment order (p q) for the continuous function f(x, y) with p, q = 0, 1, 2, ... are defined as follows:

$$m_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x, y) \, dx \, dy$$

(6)

For the momentsentral formula is given as below:

$$\mu_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \bar{x})^p (y - \bar{y})^q f(x, y) \, dx \, dy$$

(7)

with

$$\bar{x} = \frac{m_{10}}{m_{00}} \text{ dan } \bar{y} = \frac{m_{01}}{m_{00}}$$

(8)

And for digital image

$$\mu_{pq} = \sum_{x} \sum_{y} (x - \bar{x})^p (y - \bar{y})^q f(x, y)$$

(9)

The region is represented by seven invariant moments which can be searched from the central moments: $\mu_{00}, \mu_{10}, \mu_{01}, \mu_{11}, \mu_{02}, \mu_{12}, \mu_{21}, \mu_{21}, \mu_{03}$, and by defining the normalized central moments i.e. $\eta_{pq} = \mu_{pq} / \mu_{00}$, where $\gamma = 0.5$ (p q) 1 for p q = 2, 3, and so on, which finally obtained seven invariant moments, the seven invariant moments values are shown in equations 10 to equations 16.

$$M_1 = \eta_{20} + \eta_{02}$$

(10)

$$M_2 = (\eta_{20} + \eta_{02})^2 + 4\eta_{11}^2$$

(11)

$$M_3 = (\eta_{30} - \eta_{12})^2 + (3\eta_{21} - \eta_{03})^2$$

(12)

$$M_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2$$

(13)

$$M_5 = (\eta_{30} + 3\eta_{12})(\eta_{30} + \eta_{12})[3(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} - \eta_{03})^2]$$

(14)

$$M_6 = (\eta_{30} - \eta_{02})(\eta_{30} + \eta_{12})[3(\eta_{30} + \eta_{12})^2 + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})$$

(15)
\[ M_7 = (3\eta_{21} - \eta_{30})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{12} - \eta_{30})(\eta_{21} - \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \] (16)

The values of these invariant moments do not change with respect to rotation, translation and scale [11].

The stage of feature extraction features are as follows:

- Read the image
- Convert the image to Gray Scale
- Calculate seven invariant moment values
- Calculate the distance value through the seven values of the form feature between the query image and the database image
- The most similar image distance value is the searched image

3.4. Extraction of color feature

Preprocessing with histogram equalization in Figure 2 produces a more contrasting image. The image is an input image on the color feature extraction. HSV color space (Hue, Saturation and Value) can sort colors in images like human vision so as to recognize images well [8]. HSV is one way to define colors based on color wheel. Hue to measure the angle around the color wheel (0 degrees in red, 120 degrees in green and 240 degrees in blue). Saturation shows the color wheel radius so that the proportion between dark colors to light or pure white or declared the degree of color purity that indicates how much white color is given on the color. Value attribute that states the amount of light received by the eye regardless of the color or declared the brightness of color or brightness.

The HSV model is a derivative of RGB so it needs conversion from RGB to HSV. RGB or Red, Green and Blue are the colors formed by the color model which is a mixture of primary colors of red with green and blue compositions. Stages of image search based on the color features are as follows:

- Read the image
- Convert RGB color image to HSV
- Quantize each pixel in 256 histogram bins
- Perform normalization of the histogram by dividing the number of pixels
- Save 256 values of color features in the database vector
- Calculate the image similarity using euclidean distance
- The image that has the smallest resemblance distance is the image sought.

3.5. Similarity measurement

Measurement of similarity is used Euclidean distance, where in images that have a close similarity distance zero can be categorized the most similar. Calculating the value of the similarity distance using the equation 17.

\[ D(Q, M) = \sqrt{\sum_{n=1}^{k} (Q_n - M_n)^2} \] (17)

Where Q and M refers to the features of query images and images in the database in the dimension number n.

Results of the similarity calculation were sorted in such a way that the similarity whose value is the closest to zero has the highest level of similarity.

3.6. Performance measurement of each feature in supporting retrieval accuracy

Image data search is done by using CBIR (Content Based Image Retrieval) method by using three image features to match the similarity of image that will be the key of image data search. Each feature has different performance in favor of improving retrieval or search results. In this research, K-Fold Cross
Validation and Confusion matrix methods are used to measure feature performance in support of retrieval accuracy.

The data sharing of training data and test data for accuracy, precision and recall testing is varied from \( K = 3 \) to \( K = 10 \). For example, if \( K = 3 \), then the first part starts from the 1st data up to the 180th data as training data, the second part starts from the 181th data up to the 360th data as the validation data and the third part starts from the data to -361 up to 480th data as training data and 481th to 600th data are test data.

4. Results and analysis

Tests in this study using 5 types of face images with the number of images as much as 600 measuring 300 x 300 pixels.

4.1. Retrieval results

Retrieval testing is done by varying the number of clusters and the percentage variation of feature weights. Test results with weighted variations of 50% for feature weight, 30% color feature weight and 20% texture feature weight with variation of 5 clusters with number of displayed images shown in Figure 5.

\[ \text{Figure 5. Image of search results.} \]

In Figure 5. The retrieval result is shown with a similarity ranking of 1 to 24 from the face image database. Tests with weighted variations and cluster number variations will result in different retrieval outputs in the same data.

4.2. Discussion

At this stage, the analysis of facial image akurasiretrieval based on the features of shape, color and texture using face image database before and after clustering. In Table 1 it shows the percentage of retrieval accuracy using different feature weights on the image database before clustering.

\[ \text{Table 1. The percentage retrieval accuracy with variation of feature weighted scheme prior to clustering.} \]

| Img | Equal | S1  | S2  | S3  | S4  |
|-----|-------|-----|-----|-----|-----|
| F 1 | 75.23%| 78.99%| 87.52%| 83.69%| 68.03%|
| F 2 | 74.13%| 77.65%| 85.52%| 83.69%| 67.33%|
| F 3 | 75.20%| 78.99%| 87.52%| 83.69%| 68.03%|
| F 4 | 75.15%| 78.99%| 85.52%| 82.00%| 67.33%|
| F 5 | 74.13%| 76.40%| 87.52%| 83.69%| 68.03%|
Based on Table 3.1, it is shown that the retrieval accuracy with the S2 scheme yields the highest accuracy of over 85% and the S4 weight variation yields the lowest accuracy of less than 69%. The S2 scheme is a feature weighting where 50% feature shapes, 30% color features and 20% texture features. From Table 1 it can be seen that performance of texture features is more dominant in face image retrieval. The retrieval accuracy graph with the same feature weight and feature weight varies before clustering is shown in Figure 6.

**Figure 6.** The graph of retrieval accuracy with weighted of performance feature before clustering.

In Figure 5.6, it shows that S2 weight of performance feature variation yields the highest retrieval accuracy compared to other weight variations.

4.2.2. After clustering. Clustering stages and variation of weighting schemes affect the speed of the retrieval process and the level of retrieval accuracy. Percentage of retrieval accuracy with difference of feature weight scheme and number of clusters can be seen in Table 2.

| K   | Equal | S1     | S2     | S3     | S4     |
|-----|-------|--------|--------|--------|--------|
| 3   | 72.13%| 75.59% | 92.67% | 82.38% | 63.33% |
| 5   | 75.23%| 77.45% | 95.60% | 86.36% | 65.40% |
| 7   | 76.40%| 77.75% | 95.60% | 86.38% | 65.67% |
| 9   | 76.40%| 77.45% | 95.67% | 86.87% | 65.66% |
| 10  | 76.40%| 77.65% | 95.67% | 86.67% | 65.69% |
| 15  | 74.1% | 75.15% | 93.55% | 84%    | 64.2%  |
| 20  | 73.1% | 74.25% | 93%    | 83.3   | 64.67% |

Based on Table 2 shows that the variation of S2 weighting scheme with a 50% weight form feature, 30% color features and 20% texture features as well as cluster number variations between 5 to 10 clusters yields a retrieval average accuracy of more than 95% higher than the cluster others. While S4 weighted variation with 30% feature weight form, 50% color feature weight and 20% texture feature weight with variation of cluster number 3 yields the lowest accuracy of less than 64%.

The retrieval accuracy graph on the Face image database uses the same and varied feature weights that are S1, S2, S3 and S4 with variations of cluster numbers ranging from 3 to 20 can be seen in Figure 6.
Figure 7. Graph of retrieval accuracy with weighted of performance feature after clustering.

Figure 7 shows that retrieval accuracy tends to rise to a variation in the number of clusters ranging from 5 to 10 and tends to decrease in the number of clusters greater than 10. This test is performed on 600 face image data with 480 training images and 120 test images. The test was taken randomly with 24 data displayed.

4.2.3. Performance feature test. To know the performance of image features in supporting the accuracy of the system conducted testing. Test method used in this study is K-fold Cross Validation. In this test uses 3 feature values that have 7 dimensions for shape features, 9 dimensions for color features and 16 dimensions for texture features. The total data used in this test is 600 image data that is 120 face1 image (F1), 120 face2 image (F2), 120 face3 images (F3), 120 face4 image (F4) and 120 image face5 (F5).

Image performance test feature using K-fold with value 2, 3, 6 and 9 with SMO classification. Variations of K-fold values to determine the magnitude of K-fold values that produce the percentage accuracy, precision and recall reach the highest level. Table comparison of accuracy, precision and recall value can be seen in Table 3 and confusion matrix value on K-fold with k = 3 value shown by Table 4.

| K-Fold | Accuracy (%) | Precision (%) | Recall (%) |
|--------|--------------|---------------|------------|
| 2      | 94.99        | F1 100        | F1 100     |
|        |              | F2 97.2       | F2 85.1    |
|        |              | F3 100        | F3 100     |
|        |              | F4 99.2       | F4 99.6    |
|        |              | F5 100        | F5 95.0    |
|        |              | F1 100        | F1 100     |
|        |              | F2 98.1       | F2 85.1    |
| 3      | 95.62        | F3 100        | F3 100     |
|        |              | F4 96.8       | F4 100     |
|        |              | F5 88.8       | F5 97.5    |
|        |              | F1 100        | F1 100     |
|        |              | F2 98.1       | F2 85.1    |
| 6      | 95.62        | F3 100        | F3 100     |
|        |              | F4 96.8       | F4 100     |
|        |              | F5 88.8       | F5 97.5    |
|        |              | F1 100        | F1 100     |
|        |              | F2 98.1       | F2 85.1    |
| 9      | 95.62        | F3 100        | F3 100     |
|        |              | F4 96.8       | F4 100     |
|        |              | F5 88.8       | F5 97.5    |
Table 4. Confusion matrix of face image feature with K-Fold = 3.

| Actual | F1 | F2 | F3 | F4 | F5 | Total |
|--------|----|----|----|----|----|-------|
| F1     | 119| 0  | 0  | 0  | 0  | 119   |
| F2     | 15 | 103| 0  | 3  | 0  | 121   |
| F3     | 0  | 0  | 119| 0  | 1  | 120   |
| F4     | 0  | 0  | 0  | 120| 0  | 120   |
| F5     | 0  | 2  | 0  | 1  | 117| 120   |
| Total  | 134| 105| 119| 124| 118| 600   |

Table 4 shows that starting at K-fold with values of 3, 6 and 9 has a high accuracy value of more than 95% compared to K-fold of 2. At K-fold with value 3 obtains a mean value of precision equal to 95.7% and the average recall value of 95.5%. Furthermore, accuracy, precision and recall on K-fold 6 and 9 also has the same precision, precision and recall value.

4.2.4. Computation time. In testing the face tracking time is done by the variation of the number of clusters of the face image database. The test result is the average time of the face tracking which can be seen in Table 5.

Table 5. The average face time trial test results with variation of cluster number.

| Number of cluster | The tracking of time (milliseconds) | Equal | W1 | W2 | W3 | W4 |
|-------------------|-----------------------------------|-------|----|----|----|----|
| 3                 | 93                                | 93    | 93 | 93 | 93 | 93 |
| 4                 | 62                                | 62    | 62 | 62 | 62 | 62 |
| 5                 | 56                                | 56    | 56 | 56 | 56 | 56 |
| 6                 | 55                                | 55    | 55 | 55 | 55 | 55 |
| 7                 | 54                                | 54    | 54 | 54 | 54 | 54 |
| 8                 | 53                                | 53    | 53 | 53 | 53 | 53 |
| 9                 | 51                                | 51    | 51 | 51 | 51 | 51 |
| 10                | 49                                | 49    | 49 | 49 | 49 | 49 |
| 15                | 47                                | 47    | 47 | 47 | 47 | 47 |
| 20                | 45                                | 45    | 45 | 45 | 45 | 45 |

Based on table 5 shows that on the same cluster, it tends to have the same average.

5. Conclusion
Each feature performs differently in support of CBIR-based retrieval accuracy. In this study proves that the use of different feature-weighted schemes on each normalized feature value and cluster determination affect the increase of face retrieval image accuracy.

The research results show that the shape feature in the face image is the dominant factor to determine the level of similarity in the retrieval process, while the color features and texture features are the complementary features. The average computational time required for a retrieval is 5 milisecond.

References
[1] Sani 2007 Aplikasi Image Retrieval berdasarkan tekstur dengan menggunakan transformasi Haar Wavelet Seminar Nasional Sistem dan Informasi, Bali
[2] Vijay K 2008 A Content Based Approach to Image Database Retrieval Journal of Computer Applications 1(4)
[3] Herry S 2009 Sistem Pencarian Citra Digital Menggunakan Content Based Image Retrieval Proceeding Seminasif
[4] Hiremath and Jagadesh 2012 Content Based Image Retrieval Based on Color, Texture and Shape
Feature Using Image and Its Complement *International Journal of Computer Science and Security* 1(4)

[5] Sumana K and Remesh A 2015 Efficient Technique Using Color, Texture and Shape Feature In Sketches *IJAETAE* 5

[6] Lingadalli K R and Ramesh N 2015 Content Based Image Retrieval Using Color, Shape and Texture *International Advanced Research Journal in Science, Engineering and Technology*

[7] Jumi, Harjoko A and Ashari A 2015 Content Based Image Retrieval for Asset Management based on Weighted Feature and K-Means Clustering *JATIT* 77(1)

[8] Gonzales R C and Woods R E 2008 *Digital Image Processing, Third Edition* (New Jersey: Pearson Prentice Hall)

[9] Harralick R M, Shanmugam K and Dinstein I 1973 Texture Features for Images Classification *IEEE Transaction on Systems, Man and Cybernetics*

[10] Acharya T and Ray A K 2005 *Image Processing Principle and Applications* (USA: John Willey & Sons)

[11] Castleman K R 1996 *Digital Image Processing* (New Jersey: Prentice Hall Inc.)