

Multi Features Content-Based Image Retrieval Using Clustering and Decision Tree Algorithm

Kusrini Kusrini^{*1}, M. Dedi Iskandar², Ferry Wahyu Wibowo³
STMIK AMIKOM Yogyakarta,

Jl. Ringroad Utara Condong Catur Depok Sleman Yogyakarta Indonesia,
Telp:+62274884201/fax:+62274884208

*Corresponding author, e-mail: kusrini@amikom.ac.id¹, dedi.s.kom@gmail.com², ferry.w@amikom.ac.id³

Abstract

The classification can be performed by using the decision tree approach. Previous researches on the classification using the decision tree have mostly been intended to classify text data. This paper was intended to introduce a classification application to the content-based image retrieval (CBIR) with multi-attributes by using a decision tree. The attributes used were the visual features of the image, i.e.: color moments (order 1, 2 and 3), image entropy, energy and homogeneity. K-means cluster algorithm was used to categorize each attribute. The result of categorized data was then built into a decision tree by using C4.5. To show the concept in application, this research built an application with main features, i.e.: cases data input, cases list, training process and testing process to do classification. The resulting tests of 150 rontgen data showed the training data classification's truth value of 75.33% and testing data classification of 55.7%.

Keywords: Image, Classification, Decision Tree, Clustering

Copyright © 2016 Universitas Ahmad Dahlan. All rights reserved.

1. Introduction

Image database has been commonly used by many application domains nowadays such as multimedia search engines, digital libraries, medical databases, geographical databases, e-commerce, online tutoring systems and criminal investigations. Image database could be visualized using image browsing as a way to retrieval systems [1]. Image retrieval can be performed using attributes attached to the image such as creation date, storage location, size or other predefined attributes. However, the search performance produced in this way is highly depended on the expertise of a user in describing the image. In addition, the search could not be performed based on the semantics of the image itself [2].

Technology is evolving towards image searching by using method called content-based image retrieval (CBIR) and also known as query by image content (QBIC) [3]. Instead of taking image's information from external resources or metadata, CBIR approach used the intrinsic features of the image such as color, shape, texture, or a combination of these feature elements [4]. Color as an image feature has been successfully applied in image retrieval application since it has strong correlation with the objects inside the image [5]. Furthermore, the color feature's robustness has been proven in processing scaled and orientation-changed images [6-10]. The image itself is classified into three images i.e. intensity image, indexed image, and binary image [11]. To be more specific, CBIR's retrieval technique is based on information extracted from pixels [12, 13].

CBIR has been widely applied in many research projects. One of application of CBIR is for recognizing porn image [14-15]. CBIR is also used in health area researches. It is used as image administration system that supports the physician task such as diagnosis, telemedicine, teaching and learning new medical knowledge [16-18].

This paper introduces the construction of CBIR by using a case based reasoning concept. In previous research, case-based reasoning has been explained as a concept to build rule for clinical service, specifically in diagnosis problem domain [19]. One algorithm that can be used to build rules as a decision tree is C4.5. It could be also applied to classify image [20-22].

The novelty of this research is the combining used of the clustering method and decision tree developing in image classification. Instead of predetermining the discrete data

manually for classification process, the data clustered by using the k-means clustering method. This process is used to classify each of image features before classification process.

2. Research Method

2.1. Architecture System

The architecture system of CBIR in this paper was formulated as shown in Figure 1. All of image documents were preprocessed in order to make all images have the same format and size. After the preprocessing step have finished, some visual features were then extracted from the images and stored into image databases. In the next step, using the k-means clustering method [23], each of the features was then categorized and stored into categorized image database. These data were then used to build the decision tree using the C4.5 algorithm.

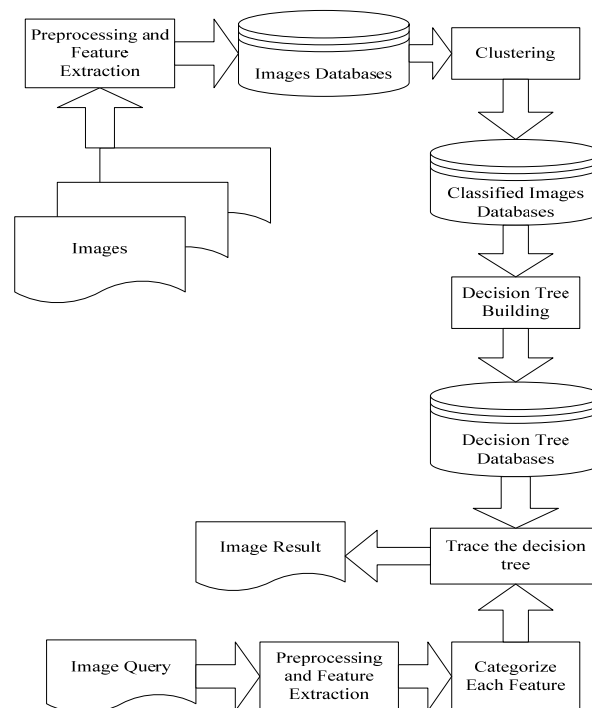


Figure 1. CBIR Architecture

2.2. Image Feature Extraction

Images were preconditioned into some same states, before the process of feature extraction was started, namely size of 140x140 pixel, bmp format, and turned into 8 bit grey image color mode. The features used in this experiment are quite similar to previous study, that are color moment and texture. The color moment was divided into 3 low-order moments while the texture features used were contrast, correlation, energy, homogeneity, and entropy [24], but in this paper used entropy, energy, contrast, and homogeneity [25]. Entropy, energy, contrast, and homogeneity features represent image texture. Texture can be defined as a region's characteristic that is wide enough to form pattern repetition. In another definition, texture is defined as a specifically ordered pattern consisting of pixel structures in the image.

One parameter needed to calculate the value of entropy, energy, contrast, and homogeneity is obtained from co-occurrence matrix, that is a matrix that describes the frequency with which a pair of two pixels with a certain intensity within a certain distance and direction occurs in the image. Co-occurrence intensity matrix $p(i_1, i_2)$ is defined by the following two simple steps:

1. Calculate the distance of every two dots in the image, expressed in a vector of vertical and horizontal directions ($d = (dx, dy)$). The values of dx and dy are also in pixel.

2. Count the number of pixel pairs that have intensity values of i_1 and i_2 and distance of d pixels.

Put the calculation of each pair intensity values into the matrix according to the coordinates, which the abscissa for intensity value i_1 and the ordinate for intensity value i_2 .

Entropy is an image feature used for measuring the disorder of intensity distribution. The entropy calculation is shown in Equation 1.

$$Entropy = -\sum_{i_1} \sum_{i_2} p(i_1, i_2) \log p(i_1, i_2) \quad (1)$$

Energy is a feature for measuring the number of concentration of intensity pairs in co-occurrence matrix. Equation 2 is used to calculate this feature's value. The value will be increased if qualified pixel pairs that matched the co-occurrence intensity matrix requirement are more concentrated at some coordinates of the matrix and will be decreased if they are more dispersed.

$$Energy = \sum_{i_1} \sum_{i_2} p^2(i_1, i_2) \quad (2)$$

The contrast feature is used to measure the strength of different intensity of the image, while homogeneity is used to measure homogeneity of image intensity variation. Image's homogeneity value will be increased as its variation intensity is decreased. The formula to measure the image contrast is shown in Equation 3 while that for measuring the homogeneity is shown in Equation 4.

$$Contrast = \sum_{i_1} \sum_{i_2} (i_1 - i_2)^2 p(i_1, i_2) \quad (3)$$

$$Homogeneity = \sum_{i_1} \sum_{i_2} \frac{p(i_1, i_2)}{1 + |i_1 - i_2|} \quad (4)$$

The notation of p in Equation 2, 3 and 4 denotes the probability and has value in the range of 0 to 1, that is the element value in co-occurrence matrix. Meanwhile i_1 and i_2 are denoted as nearby intensity pair in the x and y direction.

Given a sample of 1-bit image data with a size of 3×3 pixels. Finding the mean of RGB values of each pixel, the result is represented in a matrix as shown in Figure 2.

0	1	1
1	0	1
1	1	1

Figure 2. Image's RGB mean matrix

Feature calculations of entropy, energy, contrast and homogeneity is based on the co-occurrence intensity matrix. Consequently, prior to searching those features values, co-occurrence intensity matrix is needed to be built. In this system, the value of distance (d)=1 and direction in angle 45° ($dx=1$ and $dy=1$) are predetermined. From the above setting, then the value of co-occurrence intensity matrix value from image matrix Figure 2 is obtained as shown in Figure 3.

The second value in row $dx = 1$ and column $dy = 1$ is obtained from pair count with distance 1 and angle 45° between intensity value of image 1 and intensity value of image 1 as illustrated on Figure 4.

		dy	
		0	1
dx	0	1	1
	1	0	2

Figure 3. Co-occurrence Intensity Matrix

0	1	1
1	0	1
1	1	1

Figure 4. Illustration of pair count calculation

The value of $p(0,1)$ is obtained from element value on row $dx=0$ and column $dy=1$ divided by total element value in co-occurrence matrix. Thus, the value of $p(0,1)$ is $\frac{1}{4}$ or 0.25.

The entropy visual feature value is calculated by Equation 1, as follows:

$$\begin{aligned}
 \text{entropy} &= -(p(0,0)\log(p(0,0)) + p(0,1)\log(p(0,1)) + p(1,0)\log(p(1,0)) + \\
 &\quad p(1,1)\log(p(1,1))) \\
 &= -(0.25 \log(0.25) + 0.25 \log(0.25) + 0 \log(0) + 0.5 \log(0.5)) \\
 &= -(-0.15 - 0.15 - 0 - 0.15) \\
 &= 0.45
 \end{aligned}$$

The value of energy visual feature is calculated by Equation 2, as follows:

$$\begin{aligned}
 \text{energy} &= (p^2(0,0) + p^2(0,1) + p^2(1,0) + p^2(1,1)) \\
 &= (0.25^2 + 0.25^2 + 0^2 + 0.5^2) \\
 &= 0.375
 \end{aligned}$$

The value of visual feature contrast is obtained with calculation using Equation 3. That is:

$$\begin{aligned}
 \text{contrast} &= (0-0)^2 \cdot 0.25 + (0-1)^2 \cdot 0.25 + (1-0)^2 \cdot 0 + (1-1)^2 \cdot 0.5 \\
 &= 0 + 0.25 + 0 + 0 \\
 &= 0.25
 \end{aligned}$$

Visual feature homogeneity is obtained with calculation using Equation 4. That is:

$$\begin{aligned}
 \text{Homogeneity} &= \frac{0.25}{1} + \frac{0.25}{2} + \frac{0}{2} + \frac{0.5}{1} \\
 &= 0.25 + 0.125 + 0.125 + 0.5 \\
 &= 0.875
 \end{aligned}$$

Color moment can be defined as simple representation of color feature in colored images. The three low-order moments of color for capturing the information of image's color distribution are mean, standard deviation and skewness [25]. For a color c in the image, the mean of c is symbolized as μ_c , the standard deviation as σ_c and the skewness as θ_c . The values of μ_c , σ_c , and θ_c are calculated using Equation 5, Equation 6, and Equation 7, respectively.

$$\mu_c = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N p_{ij}^c \quad (5)$$

$$\sigma_c = \left[\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (p_{ij}^c - \mu_c)^2 \right]^{\frac{1}{2}} \quad (6)$$

$$\theta_c = \left[\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (p_{ij}^c - \mu_c)^3 \right]^{\frac{1}{3}} \quad (7)$$

Where M and N are the horizontal and vertical sizes of the image and c is the value of color c in the image's row i and column j .

Using Equation 5, the value of visual features color moment order 1 is calculated from the matrix as follows:

$$\begin{aligned}\mu_c &= \frac{1}{3 \times 3} (0 + 1 + 1 + 1 + 0 + 1 + 1 + 1 + 1) \\ &= 0.78\end{aligned}$$

The value of color moment order 2 is obtained using the following steps:

1. Transform each value in the RGB mean matrix in Figure 2 by operating each value in the matrix with the Equation 8.

$$nb = (p_{ij}^c - \mu_c)^2 \quad (8)$$

p_{ij}^c is the matrix cell's value in row i and column j and nb is the new value of the cell. The result is then represented in Figure 5.

0.61	0.05	0.05
0.05	0.61	0.05
0.05	0.05	0.05

Figure 5. Color Moment Order 2 Value Matrix

-0.47	0.01	0.01
0.01	-0.47	0.01
0.01	0.01	0.01

Figure 6. Color Moment Order 3 Matrix

2. The color moment order 2 value σ_c is calculated from the matrix using Equation 2.

$$\begin{aligned}\sigma_c &= \left[\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N nb \right]^{\frac{1}{2}} \\ \sigma_c &= \left[\frac{1}{3 \times 3} 1.556 \right]^{\frac{1}{2}} \\ &= 0.416\end{aligned}$$

The color moment order 3 value is obtained with the following steps:

1. Transform each value in the RGB mean matrix in Figure 2 by operating each value in the matrix with the Equation 9.

$$nb = (p_{ij}^c - \mu_c)^3 \quad (9)$$

p_{ij}^c is the matrix cell's value in row i and column j and nb is the matrix new cell's value.

Using these calculations, the matrix in Figure 2 will be transformed into the color moment order 3 matrix as shown in Figure 6.

2. The color moment order 3 value θ_c is calculated from the matrix using Equation 3.

$$\theta_c = \left[\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N nb \right]^{\frac{1}{3}}$$

$$\theta_c = \left[\frac{1}{3 \times 3} - 0.864 \right]^{\frac{1}{3}}$$

$$= -0.458$$

2.3. K-Mean Clustering

Clustering machine is intended to map the continuous or discrete class data with many variations into discrete class with determined variations. The input of this process is:

1. Initial data, that can be either continuous or discrete data.
2. Delta, that is the value to be used to determine the allowed gap between centroid and mean.

The output of this process is a mapping table containing the discrete class including its centroid value.

The algorithm used in the data clustering machine is derived and extended from k-means clustering algorithm. This data clustering using the K-Means method is commonly done with basic algorithm as follow [26]:

1. Determine count of class (cluster)
2. Determine initial centroid of each class
3. Put each data into class which has the nearest centroid
4. Calculate the data mean from each class
5. For all class, if the difference of the mean value and centroid goes beyond tolerable error, replace the centroid value with the class mean then go to step 3.

In fundamental k-means clustering algorithm, initial centroid value from specific discrete class is defined randomly, while in this research the value is produced from an equation shown in Equation 10.

$$c_i = \min + \frac{(i-1) * (\max - \min)}{n} + \frac{(\max - \min)}{2 * n} \quad (10)$$

Where,

- c_i : centroid class i
 \min : the lowest value of continue class data
 \max : the biggest value of discrete class data
 n : total number of discrete class

The building process of discrete class is shown as follows: 1) Specify the source data; 2) Specify desired total number of discrete class (n); 3) Get the lowest value from source data (\min); 4) Get the highest value from source data (\max); 5) Specify delta (d) to get the acceptable error by Equation 11; 6) For each discrete class, find the initial centroid (c) by Equation 10; 7) For each value in source data, put it into its appropriate discrete class, which is one that has nearest centroid to the value; 8) Calculate the average value of all members for each class (mean); 9) For each discrete class, calculate the difference between its mean and its centroid (s) by Equation 12; 10) For each discrete class, if $s > e$, then replace its centroid value with its mean, put out all values from their corresponding class then go back to step 7.

$$e = d * (\max - \min) \quad (11)$$

$$s = \sum_{i=1}^n | \text{mean}_i - c_i | \quad (12)$$

2.4. C4.5 Algorithm

Decision tree is one of methods in classification process. In decision tree, classification process involving classification rules generated from a set of training data. The decision tree will split the training data based on some criteria given. The criteria are divided into target criteria and determinator criteria. The classification process produces the target criteria values based on determinator criteria values [27].

The algorithm of decision tree building will search for the best deciding criterion for categorizing data homogenously. The criterion will be considered as root node. The next step of

the algorithm is creating branches based on the values inside of the criterion. The training data then categorized based on values at the branch. If the data considered homogen after the previous step, then the process for this branch is immediately stopped. Otherwise, the process will be repeated by choosing another criterion as next root node until all the data in the branch are homogen or there is no remain deciding criterion can be used.

The resulting tree will then be used in classification of new data. The process of classification is conducted by matching the criterion value of the new data with the node and branch in decision tree until finding the leaf node which is also the target criterion. The value at that leaf node is being the conclusion of classification result.

Using C4.5 algorithm, choosing which attribute to be used as a root is based on highest gain of each attribute, the gain is counted using Equation 13 and 14 [28].

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (13)$$

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i \quad (14)$$

Where

S : Case set

A : Attribute

n : Partition Count of attribute A

|S_i| : Count of cases in ⁱth partition

|S| : Count of cases in S

p_i : Proportion of cases in S that belong to the ⁱth partition

3. Results and Analysis

3.1. Result

The result of this research is a computer program to classify images. The program's main features are 1) Data Case Input; 2) Case List; 3) Training Process; and 4) Testing Process.

3.1. Data Input and Case List

The Data Case Input feature is used to store images that will be used in the training process. In this feature, the program will do preprocessing, take the image visual feature's values and then store them in the database. The case data input feature interface is shown in Figure 7. This figure shows that the loaded image processed to some visual features as explained in section 2.2 and the classification text box is filled manually as the expert judgement.

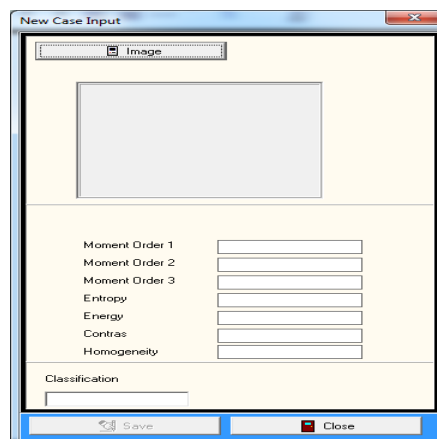


Figure 7. New Case Input Feature

The Case List feature is used to review the cases stored in the database. This feature is also provided with ability to delete a particular data that are assumed not to be used in the next training process. The Case List page is shown in Figure 8.

IDKASUS	MOMENT_1	MOMENT_2	MOMENT_3	ENTROPY	ENERGI	KONTRAS	HOMOGENITAS	KLASIFIKASI
206	66.99213	62.3374	61.23231	7.10014	0.21426	1217.76716	0.59424	HEAD
207	64.5946	60.54507	62.29351	6.89074	0.22427	1203.48622	0.60433	HEAD
208	60.76713	62.07709	64.79096	6.1319	0.30467	1542.54696	0.65187	HEAD
209	49.7188	78.19992	88.21794	5.57595	0.35516	1241.83518	0.68312	HEAD
210	127.4589	38.51756	21.28019	11.78039	0.00038	104.05039	0.26556	CHEST
212	131.32677	75.72968	7907903763.68	11.34127	0.01259	237.61566	0.36151	CHEST
213	131.75063	45.38014	89077041.0508	12.23899	0.00038	116.5133	0.26389	CHEST
214	127.19217	50.09205	-2545607890.32	11.96628	0.00105	101.71468	0.2817	CHEST
215	109.6842	48.26142	39.45755	11.82768	0.00049	144.65222	0.29647	CHEST
217	146.53413	50.91604	-982395117.220	12.71639	0.00019	197.32599	0.19138	CHEST
220	112.41392	44.94624	15.47945	11.83102	0.00027	202.05361	0.26372	CHEST
221	111.22993	64.13765	20173166.9282	11.43984	0.00542	595.3463	0.35113	CHEST
222	133.68657	54.6671	-2712572866.37	11.88567	0.0007	90.95803	0.30974	CHEST
224	128.87647	51.62903	24.05426	12.0164	0.00035	81.77252	0.28513	CHEST
225	127.59607	47.95226	-1066289605.28	11.36801	0.00069	50.30663	0.3633	CHEST
226	136.28543	49.39541	-236090709.433	12.08429	0.00037	443.11757	0.27842	CHEST
227	109.4575	46.35664	-2464239411.60	11.36485	0.00243	76.17348	0.32544	CHEST
228	123.83767	46.29532	-61863413.9355	12.00126	0.00041	133.37682	0.26394	CHEST
229	113.0441	48.10433	-1300428145.38	11.76993	0.00143	335.51347	0.30607	CHEST
235	29.99023	52.59285	58.77356	4.28302	0.50671	330.58669	0.76443	HEAD
236	36.73217	53.16427	53.5615	5.22498	0.40967	345.74935	0.70694	HEAD
237	50.3523	68.16462	73.78752	7.83879	0.16582	1742.07595	0.53752	HEAD
238	43.17677	63.80342	71.96961	6.73665	0.25745	1172.62187	0.60626	HEAD
239	61.37257	70.78824	76.18371	8.42434	0.13102	1154.94914	0.48276	HEAD
240	55.1181	47.15702	43.54283	10.85013	0.02829	703.9093	0.29622	HEAD
242	90.27853	95.04155	70.52623	6.97026	0.20882	2001.78041	0.59225	HEAD
245	83.76383	77.77055	68.65383	8.91211	0.08427	1031.2124	0.46205	HEAD

Figure 8. List of Cases Feature

3.2. Training Process using K-Means Cluster and C4.5 Algorithm

The training process is used to produce the decision tree from the cases in the case list. This process contains 2 sub processes, they are categorization process and decision tree building process. The categorization process is performed for each of image visual feature by using K-Means Cluster algorithm as explained in section 2.3. The input of categorization process is all values of images in a features and the result is the cluster and the centroid of each data. The cluster selected will be the one that has the nearest centroid. The decision tree building process is performed using C4.5 algorithm. The training process is then visualized in the form of decision tree as in Figure 9 and in a table of rule as shown in Figure 10.

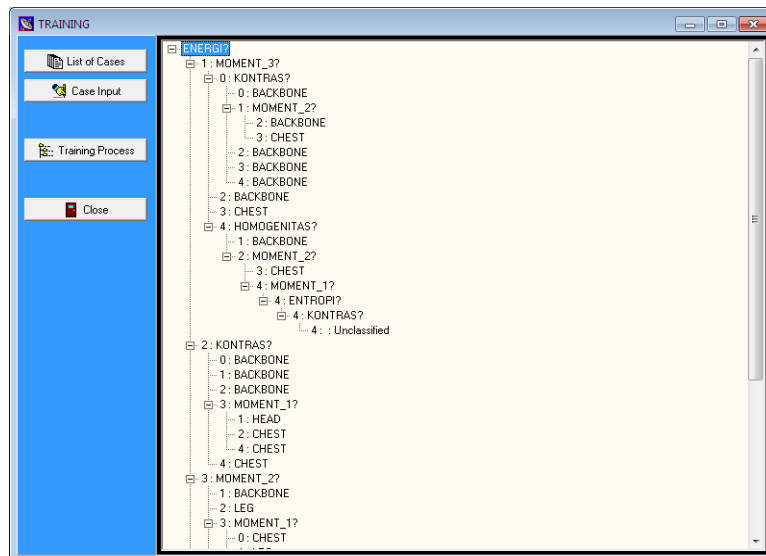


Figure 9. Tree Feature

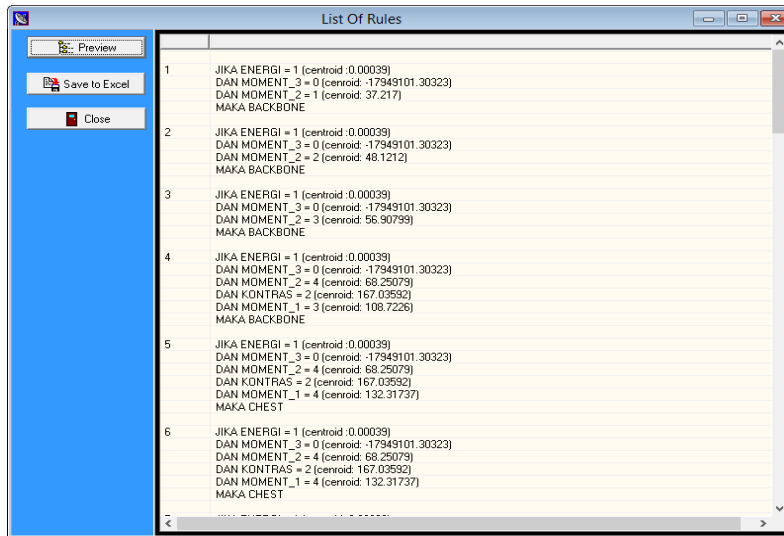


Figure 10. List of Rules Feature

3.3. Testing Process

The testing feature is used to do classification to newly incoming images. The classification is carried out by comparing between the new image's visual feature value and the rule constructed in the prior training process. The testing feature is shown in Figure 11.

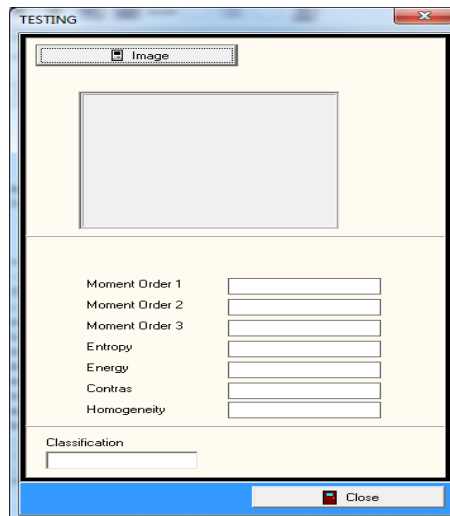


Figure 11. Testing Feature

3.2. Analysis

The testing process was carried out with 150 rontgen image data and resulted some classifications. The data is shown in Table 1.

Table 1. Cases Classification

Classification Count	
Backbone	50
Chest	50
Head	50

The testing was carried out by comparing between the classification from system's output and the expert judgment in Case Input Data as explained in section 3.1. The testing to all training data resulted to true classification for 113 data. It means 75.3% of truth value. The detail result of testing training data is shown in Table 2.

Table 2. Testing Result of All Training Data

Classification	Testing	Result	Count
Backbone	Backbone	True	30
Backbone	Chest	False	12
Backbone	Unclassified	Unclassified	8
Chest	Backbone	False	10
Chest	Chest	True	34
Chest	Head	False	1
Chest	Unclassified	Unclassified	5
Head	Chest	False	1
Head	Head	True	49

Beside the testing process to all training data, it is also conducted 10 experiments using 120 data. The data consists of 40 Backbone, 40 chest, and 40 head image data with different combination. The test result is shown in Table 3.

Table 3. Testing Results of the 120 Training Data

Iteration	True	False	Unclassified
1	88	22	10
2	88	23	9
3	85	25	10
4	85	25	10
5	88	23	9
6	92	19	9
7	97	15	8
8	97	17	6
9	94	19	7
10	90	20	10
Average	90.4	20.8	8.8
%	75.33	17.33	7.33

The testing to testing data was done by 10-fold cross validation. It is used 120 training data and 30 testing data for each of experiments. The test result with this model is shown in Table 4. From the Table 4, it is known that the testing process to the data test producing correct classification of 55.7% wrong classification of 33.7% and unclassified data about 10.7%.

Table 4. Testing Results of the Test Data

Iteration	True	False	Unclassified
1	14	16	0
2	20	6	4
3	20	9	1
4	18	9	3
5	16	7	7
6	16	9	5
7	16	11	3
8	16	9	5
9	19	7	4
10	12	18	0
Average	16.7	10.1	3.2
%	55.7	33.7	10.7

In order to compare the result, the same dataset is tested into classification algorithm without clustering preprocessing. This time, the class range is defined manually. The testing to all training data resulted to true classification for 23 data. It means 15.3% of truth value. The result of 10 experiments using 120 training data with different combination without clustering process is shown in Table 5.

Table 5. Testing Results of the 120 Training Data Without Clustering Process

Iteration	True	False	Unclassified
1	17	20	83
2	20	18	82
3	21	17	82
4	20	17	83
5	18	17	85
6	19	19	82
7	20	19	81
8	17	16	87
9	16	15	89
10	16	18	86
Average	18,4	17,6	84
%	15,3	14,7	70

The result of testing to testing data that was done by 10-fold cross validation that is not using clustering process. The result is shown in Table 6. From the Table 6, it is known that the testing process to the data test without clustering process producing correct classification of 17% wrong classification of 16% and unclassified data about 67%.

Table 6. Testing Results of the Test Data Without Using Clustering Process

Iteration	True	False	Unclassified
1	7	3	20
2	3	4	23
3	3	9	18
4	4	2	24
5	7	8	15
6	4	3	23
7	3	3	24
8	6	6	18
9	7	7	16
10	7	3	20
Average	5,1	4,8	20,1
%	17	16	67

Comparing the above processes of classification with and without clustering algorithm, it shows that the using of clustering can increase the accuracy of classification process. Since it is done manually, the classification without clustering algorithm is less accurate to define the class range based on data distribution.

4. Conclusion

From the discussion section above, it can be concluded that: (1) The use of decision tree can be applied to classify image with multi visual features, and (2) Clustering algorithm can help to improve categorization process in each of this research's selected image visual features. The accuracy of the process using the training data reached 75.33%. The accuracy of the process using the testing data reached 55.7%. It is higher than classification without using clustering process that only reached 15.3% for training data and 17% for testing data.

Based on the conclusion, there are some opportunities and challenges for further researches to improve this research, among others determining the classification of data which are still in unclassified category in order to get a better accuracy value and examining other visual features that are able to increase classification accuracy value.

References

- [1] Schaefer G. *Interactive Browsing of Large Image Databases*. IEEE Conference - Sixth International Conference on Digital Information Processing and Communications (ICDIPC). Beirut. 2016: 168-170.
- [2] Karmakar S, Dey S, Sahoo S. *Towards Semantic Image Search*. IEEE Tenth International Conference on Semantic Computing (ICSC). California. 2016: 483-487.
- [3] Ambika P, Samath JA. *Visual Query Exploration Process and Generation of Voluminous Image Database for CBIR Systems*. IEEE Conference – Third International Conference on Computing Communication & Networking Technologies (ICCCNT). Coimbatore. 2012: 1-7.
- [4] Jyothi B, Latha YM, Reddy VSK. *Relevance Feed Back Content Based Image Retrieval Using Multiple Features*. IEEE International Conference on Computational Intelligence and Computing Research (ICCIIC). Coimbatore. 2010: 1-5.
- [5] Minarno AE, Kurniawardhani A, Bimantoro F. Image Retrieval Based on Multi Structure Co-Occurrence Descriptor. *TELKOMNIKA Telecommunication Computing Electronics and Control*. 2016; 14(3): 1175-1182.
- [6] Guizhi Li. Improving Relevance Feedback in Image Retrieval by Incorporating Unlabelled Images. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2013; 11(7): 3634-3640
- [7] Sreedhar J, Raju SV, Babu AV. Query Processing for Content Based Image Retrieval. *International Journal of Soft Computing and Engineering*. 2011; 1(5): 122-131.
- [8] Patil C, Dalal V. *Content Based Image Retrieval Using Combined Features*. International Conference & Workshop on Emerging Trends in Technology. New York. 2011; 11: 102-105.
- [9] Alnihoud J. Content-Based Image Retrieval System Based on Self Organizing Map, Fuzzy Color Histogram and Subtractive Fuzzy Clustering. *The International Arab Journal of Information Technology*. 2012; 9(5): 452-458.
- [10] Haridas K, Thanamani AS. Well-Organized Content Based Image Retrieval System in RGB Color Histogram, Tamura Texture and Gabor Feature. *International Journal of Advanced Research in Computer and Communication Engineering*. 2014; 3(10): 8242-8248.
- [11] Sai NST, Patil RC. Image Retrieval Using Bit-Plane Pixel Distribution. *International Journal of Computer Science & Information Technology (IJCSIT)*. 2011; 3(3): 159-174.
- [12] Hussain SJ, Kumar RK. A Novel Approach of Content Based Image Retrieval and Classification using Fuzzy Maps. *International Journal of Research Studies in Computer Science and Engineering*. 2015; 2(3): 91-94.
- [13] Youness C, Khalid EA, Mohammed O. New Method of Content Based Image Retrieval based on 2-D ESPRIT Method and the Gabor Filters. *Indonesian Journal of Electrical Engineering and Computer Science*. 2015; 15(2): 313-320.
- [14] Liu Y, Xiao C, Wang Z, Bian C. *An One-Class Classification Approach to Detecting Porn Image*. Conference on Image and Vision Computing. New York. 2012; 27: 284-289.
- [15] Karavarsamis S, Pitas I, Ntarmos N. *Recognizing Pornographic Images*. Multimedia and Security. New York. 2012; 12: 105-108.
- [16] Ramu B, Thirupathi D. *A new methodology for diagnosis of appendicitis using sonographic images in image mining*. International Conference on Computational Science, Engineering and Information Technology. New York. 2012; 2: 717-721.
- [17] Kumar A, Nette F, Klein K, Fulham M, Kim J. A Visual Analytics Approach Using the Exploration of Multidimensional Feature Spaces for Content-Based Medical Image Retrieval. *IEEE Journal of Biomedical and Health Informatics*. 2015; 19(5): 1734-1746.
- [18] Ramos J, Kockelkorn TTJP, Ramos I, Ramos R, Grutters J, Viergever MA, Ginneken BV, Campilho A. Content-Based Image Retrieval by Metric Learning From Radiology Report: Application to Interstitial Lung Diseases. *IEEE Journal of Biomedical and Health Informatics*. 2016; 20(1): 281-292.
- [19] Haolin Gao. Which Representation to Choose for Image Cluster. *TELKOMNIKA Indonesian Journal of Electrical Engineering*. 2014; 12(8): 6181-6189
- [20] Kusriani. Klasifikasi Citra dengan Pohon Keputusan. *Jurnal Ilmiah Teknologi Informasi (JUTI)*. 2008; 7 (2): 47-56.
- [21] Kusriani, Harjoko, A. *Decision Tree as Knowledge Management Tool for Image Classification*. Knowledge Management International Conference. Malaysia. 2008: 190-195.
- [22] Kusriani. Differential Diagnosis Knowledge Building by using CUC-C4.5 Framework. *Journal of Computer Science (JCS)*. 2010; 6(2): 180-185.
- [23] Kusriani K. *Grouping of Retail Items by Using K-Means Clustering*. Information Systems International Conference. 2015; 3: 495-502.

-
- [24] Chaugule A, Mali SN. Evaluation of Texture and Shape Features for Classification of Four Paddy Varieties. *Journal of Engineering*. 2014; 2: 1-8.
- [25] Dubey RS, Choubey R, Bhattacharjee J. Multi Feature Content Based Image Retrieval. *International Journal on Computer Science and Engineering (IJCSSE)*. 2010; 2(6): 2145-2149.
- [26] Masruro A, Wibowo FW. Intelligent Decision Support System For Tourism Planning Using Integration Model of K-Means Clustering and TOPSIS. *International Journal of Advanced Computational Engineering and Networking*. 2016; 4(1): 52-57.
- [27] Agrawal GL, Gupta H. Optimization of C4.5 Decision Tree Algorithm for Data Mining Application. *International Journal of Emerging Technology and Advanced Engineering*. 2013; 3(3): 341-345.
- [28] Dai W, Ji W. A MapReduce Implementation of C4.5 Decision Tree Algorithm. *International Journal of Database Theory and Application*. 2014; 7(1): 49-60.