Classification of Images Using Decision Tree

Dr. Emad K. Jabbar
ComputerScience Department,Universityof Technology/Baghdad
E-mail:emadalfatly@yahoo.com
Mayada jabbar kelain
Ministry of Higher Education and Scientific Research/Baghdad

Revised on: 3/3/2013 & Accepted on: 9/5/2013

ABSTRACT

In this paper, the proposed system is based on texture features classification for multi object images by using decision tree (ID3) algorithm. The proposed system uses image segment tile base to reduce the block effect and uses (low low) Wavelet Haar to reduce image size without loss of any important information. The image texture features like (Entropy, Homogeneity, Energy, Inverse Different Moment (IDM), Contrast and Mean) are extracted from image to build database features. All the texture features extracted from the training images are coded into database features code. ID3 algorithm uses database features code for classification of images into different classes. Splitting rules for growing ID3 algorithm are Entropy, Information Gain used to build database rules, which depend on if_then format. The proposed algorithm is experimented on to test image database with 375 images for 5 classes and uses accuracy measure. In the experimental tests 88% of the images are correctly classified and the design of the proposed system in general is enough to allow other classes and extension of the set of classification classes.

Keywords: Texure Feature Extraction, Data Mining ,Decision tree,ID3 Algorithm.

تصنيف الصور باستخدام شجرة القرار

الخلاصة

في هذا البحث ،النظام المقترح مبني على اساس تصنيف الخصائص النسيجية للصور التي تحوي على كائنات متعددة باستخدام شجرة القرار بخوارزمية (ID3). في النظام المقترح استخدمنا (Segment tile base) لتخلص من التأثير الكتلي واستخدمنا (wavelet haar LL) لتقليل حجم الصورة بدون فقدان اي معلومات مهمة .ان الخصائص النسيجية للصورة مثل (العشوائية ، التجانس ،القوة، لحظة اختلاف الانعكاس، التناقض، المعدل) استخلصت من الصورة لبناء قاعدة بيانات صفات الصورة .إن جميع الخصائص النسيجية التي تم استخلاصها من الصورة إثناء عملية التدريب تم تحويلها الى رموز في قاعدة بيانات لاستخدامها في بناء شجرة القرار لتصنيف الصور وذلك بالاعتماد على مجموعة قوانين التي تم بناءها باستخدام ID3 .إن الخوارزمية المقترحة تم تجربتها على مجموعة صور اختباريه تصل الى 375 صورة لخمسة أصناف وباستخدام مقاييس الدقة، كانت نتائج الاختبار 88% من صور الاختبار صنفت بشكل صحيح.

الكلمات المرشدة :استخلاص الخصائص النسيجية ، تنقيب البيانات ، شجرة القرار،خوارزمية ID3 .

INTRODUCTION

Images are produced by a variety of physical devices including still and video cameras , X-ray, electronic microscope , radar, and are used for a variety of purposes, including entertainment, medical, business , industrial, military, civil ,traffic, security ,and scientific. Image processing allows one to enhance image features of interest while attenuating details irrelevant to a given applications, and then extracting useful information about the scene from the enhanced image. This information is to used extract knowledge to take decision [1]. Image mining deals with the extraction of implicit knowledge, image data relationship, or other patterns not explicitly stored in the image databases. The images from an image database are first preprocessed to improve their quality. These images then undergo various transformations and feature extraction to generate the important features from the images. With the generated features, mining can be carried out using data mining techniques to discover significant patterns. The resulting patterns are evaluated and interpreted to obtain the final knowledge, which can be applied to applications [2]. The decision tree induction is a well-known methodology used widely in various domains, such as artificial intelligence, machine learning, data mining, and pattern recognition .It is a predictive model which usually operates as a classifier. The construction of a decision tree, also called decision tree learning process, uses a training dataset consisting of records with their corresponding features and a label attribute to learn the classifier. Once the tree is built, it can be used to predict the label attribute of unidentified records that are from the same domain [3].

TEXTURE FEATURE EXTRACTION

Texture, the pattern of information or arrangement of the structure found in an image, is an important feature of many image types. In a general sense, texture refers to surface characteristics and appearance of an object given by the size, shape, density, arrangement, proportion of its elementary parts. Due to the significatnace of texture information, texture feature extraction is a key function in various image processing applications, remote sensing and content-based image retrieval. Texture features can be extracted in several methods, using statistical, structural, model-based and transform in- formation, in which the most common way is using the Gray Level Co-occurrence Matrix (GLCM). GLCM contains the second-order statistical information of spatial relationship of pixels of an image[4]. In general, GLCM could be computed as follows. First, an original texture image D is re-quantized into an image G with reduced number of gray levels, N_g. A typical value of N_g is (16). Then, GLCM is computed from G by scanning the intensity of each pixel and its neighbor, defined by displacement d and angle $ø$.Finally, scalar secondary features are extracted from this co-occurrence matrix such as[5] :

- **Entropy** [6]: measures the disorder of an image and it achieves its largest value when all elements in P matrix are equal. When the image is not texturally uniform many GLCM elements have very small values, which implies that entropy is very large. Therefore, entropy is inversely proportional to GLCM energy.

$$\textbf{Entropy} = -\sum_{i,j}^{n} \boldsymbol{P}\ (\boldsymbol{i},\mathbf{j})\log \boldsymbol{P}\ (\boldsymbol{i},\mathbf{j}) \qquad \ldots(1)$$

- **Energy** [7]: is also called Uniformity or angular second moment. It measures the textural uniformity that is pixel pair repetitions. It detects disorders in textures. Energy reaches a maximum value equal to one.

$$\textbf{Energy} = \sum_i \sum_j P_d{}^2(i,j) \quad \ldots(2)$$

- **Contrast** [7]**:** is a measure of the degree of spread of the grey levels or the average grey level difference between neighboring pixels. The contrast values will be higher for regions exhibiting large local variations. The GLCM associated with these regions will display more elements distant from the main diagonal, than regions with low contrast. Local statistics contrast and GLCM contrast are strongly correlated.

$$\textbf{Contrast} = \sum_i \sum_j P_d(i,j)^2(i,j) \quad \ldots(3)$$

- **Homogeneity** [6] :It measures image homogeneity as it assumes larger values for smaller gray tone differences in pair elements. It is more sensitive to the presence of near diagonal elements in the GLCM. The GLCM contrast and homogeneity are strong, but inversely correlated in terms of equivalent distribution in the pixel pairs population. It means homogeneity decreases if contrast increases while energy is kept constant.

$$\textbf{Homogeneity} = \sum_i \sum_j \frac{\mathcal{P}_d(i,j)}{1+(i-j)2} \quad \ldots(4)$$

- **IDM** [6]**:** It measures image homogeneity. This parameter achieves its largest value when most of the occurrences in GLCM are concentrated near the main diagonal.

$$\textbf{IDM} = \sum_{i,j} \frac{1}{1+(i-j)2} \mathbf{p(i,j)} \quad \ldots(5)$$

- **Mean** [7]**:** The GLCM Mean is expressed in terms of the gray level co-occurrence matrix. Consequently, the pixel value is weighted not by its frequency of occurrence by itself (as in common mean expression), but by the frequency of its occurrence in combination with a certain neighboring pixel value.

$$\textbf{Mean} = \sum_{i,j=0}^{n-1} i \ (\mathrm{p}(\mathrm{i},\mathrm{j}) \quad \ldots(6)$$

DATA MINING

Data mining is a term that describes different techniques used in a domain of machine learning, statistical analysis, modeling techniques and database technologies that can be used in different industries. With a combination of these techniques, it is possible to find different kinds of structures and relations in the data, as well as to derive rules and models that enable prediction and decision making in new situations.

It is possible to perform classification, estimation, prediction, affinity grouping, clustering and description and visualization [8]. Classification of data objects based on a predefined knowledge of the objects is a data mining and knowledge management technique used in grouping similar data objects together. It can be

defined as supervised learning algorithms as it assigns class labels to data objects based on the relationship between the data items with a pre-defined class label and there are many classification algorithms available in literature but decision trees are the most commonly used because of their ease of implementation and are easier to understand compared with other classification algorithms [9].

DECISION TREE

A decision tree is a tree in which each branch node represents a choice between a number of alternatives, and each leaf node represents a decision. Decision trees are commonly used for gaining information for the purpose of decision-making. A decision tree starts with a root node on which it is for users to take actions. From this node, users split each node recursively according to decision tree learning algorithm. The final result is a decision tree in which each branch represents a possible scenario of a decision and its outcome [10].

ID3 Decision Tree Induction Algorithm

ID3 is a simple decision tree learning algorithm developed by Ross Quinlan (1983). ID3 produces a decision tree which can classify the outcome value based on the values of the given attributes. ID3 algorithm which is a supervised learning, with the ability of generating rules through a decision tree .To construct the decision tree, calculate the entropy of each features of the training images by using ID3 algorithm and measure the information gained for each features and take maximum of them to be the root[11].

- **Entropy:** The idea of entropy of random variables was proposed by Claude Shannon .There are several ways to introduce the notion of entropy. Quantities of the form [12]. Will be recognized as that of entropy as defined in certain formulations of statistical mechanics where pi is the probability of a system being in cell i of its phase space. We shall call R= -∑pi log pi the entropy of the set of probabilities $p_1 \ldots p_n$. If x is a chance variable we will write R(x) for its entropy ;thus x is not an argument of a function but a label for a number, to differentiate it from R(y) say, the entropy of the chance variable y [12].

- **Information Gain**: Given entropy as a measure of the impurity in a collection of training examples, we can now define a measure of the effectiveness of an attribute in classifying the training data. The measure we will use, called information gain, is simply the expected reduction in entropy caused by partitioning the examples according to this attribute. More precisely, the information gain, Gain(S, A) of an attribute A**,** relative to a collection of examples ***S,*** is defined as[10]:

$$\textbf{Gain(S,A)= Entropy(S) } -\sum\left(\left(\frac{|Sv|}{|S|}\right) \textbf{ * Entropy(Sv) } \ldots(7)\right.$$

where ∑ is over each value V of all the possible values of the attribute A,
SV= subset of S for which attribute A has value V ,
|SV|=number of element in SV,
|S|= number of element in S.

The proposed ID3 algorithm is as follows,

ID3 (Learning Sets S, Attributes Sets A, Attributesvalues V)
Return Decision Tree.

- Begin
- Load learning sets first, create decision tree root node 'rootNode', add learning set S into root node as its subset.
- For rootNode, we compute Entropy(rootNode.subset) first
- If Entropy (rootNode.subset)==0, then rootNode.subset consists of records all with the same value for the categorical attribute, return a leaf node with decision attribute:attribute value;
- If Entropy(rootNode.subset)!=0, then compute information gain for each attribute left(have not been used in splitting), find attribute A with Maximum(Gain(S,A)). Create child nodes of this rootNode and add to rootNode in the decision tree.
- For each child of the rootNode, apply ID3(S,A,V) recursively until reach node that has entropy=0 or reach leaf node.
- End ID3.

PROPOSED ALGORITHM

The main idea of the proposed algorithm depends on the fact that any image has multi unique features. These features are different from image to another depending on their objects color and texture. In this paper, an algorithm is proposed to classify image by image mining and extracting features from the image depending on texture such as (Entropy, Energy, Contrast, Mean, Inverse Difference Moment and Homogeneity). All the texture features extracted from the training images are classified into database features and then used to choose the closest one to the requested image.To specify these features we use gray level single – channel images to extract their features and contract a feature vectors by using Co-occurrence matrix for each textured image, then feature vectors are classified into groups by using hierarchical classification techniques to use them with decision tree. The structure of the proposed approach consists of two phases (Training phase, testing phase) as shown in Figure (1) Each phase has specific functions.

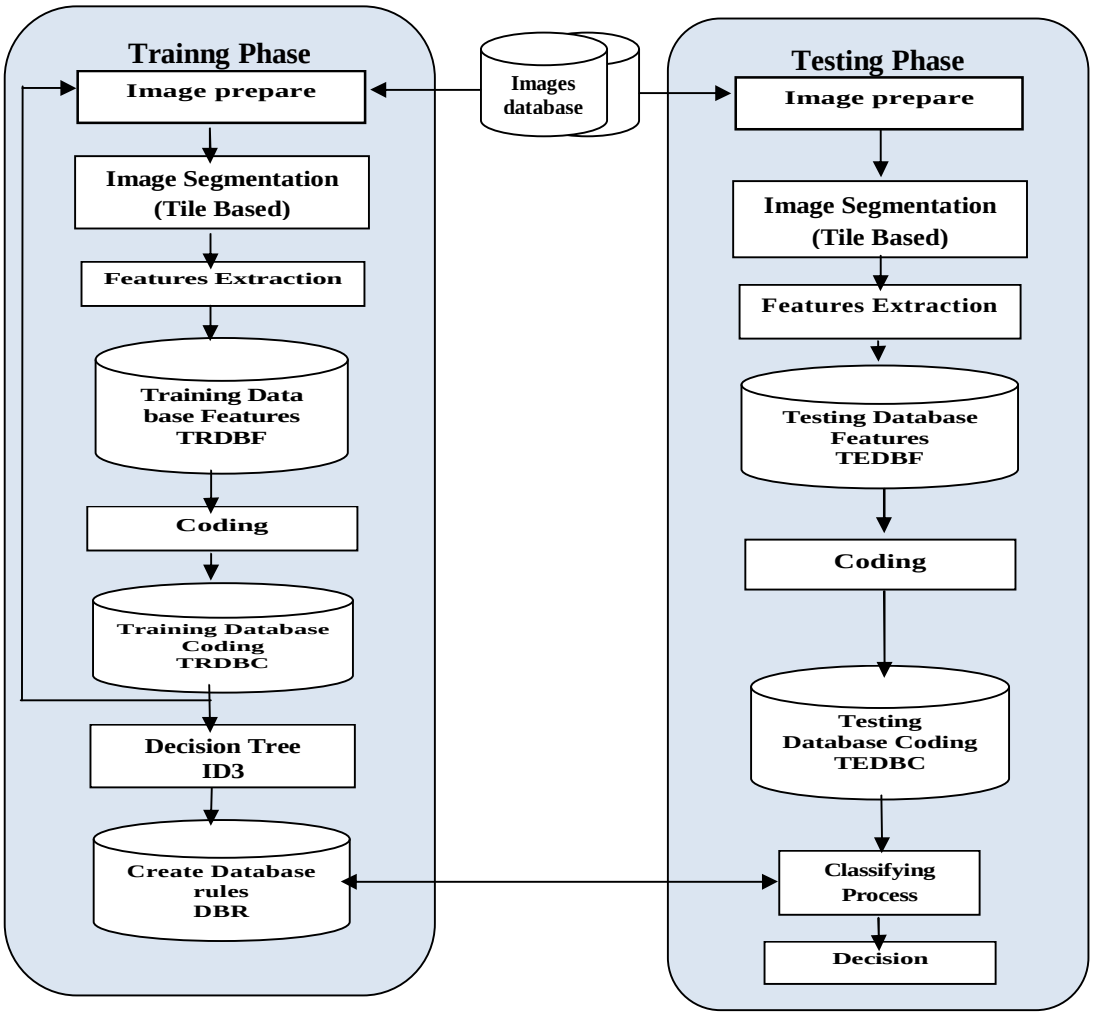

Figure (1) Structure of Proposed CIUDT.

Training Phase

This phase consists of many main steps such as:

Step 1: - Image preparing

In this step the images are loaded and image size reduced into NxN as shown in Figure (2) that will reduce the computing time and storage space.

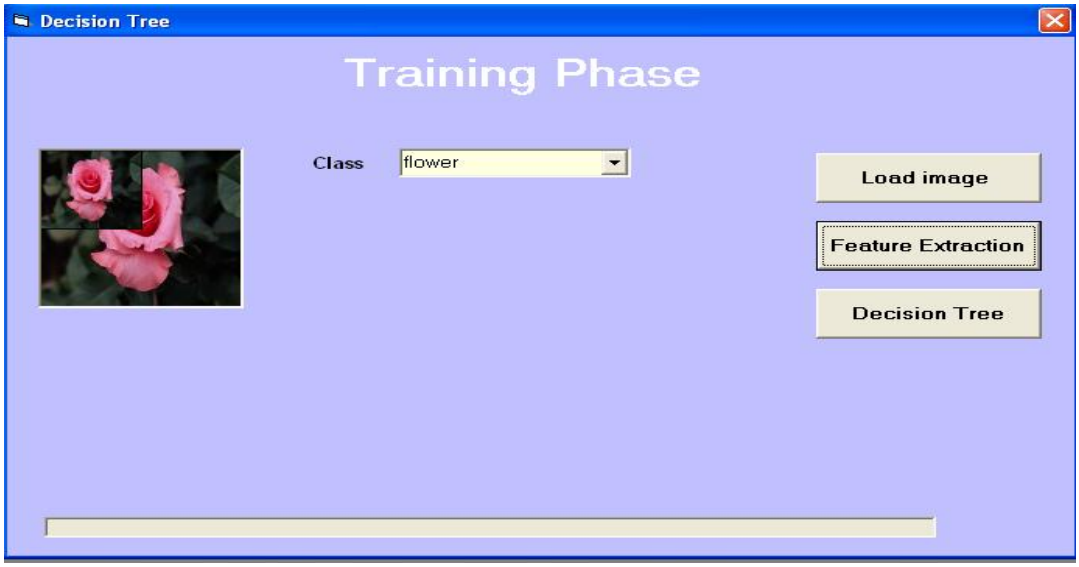

Figure (2) The user interface of training image.

Step 2: - Image Segmentation (Tile based)

In this step we use tile base to get important part of image by dividing the image to NXN parts and neglecting any part that has entropy less than threshold.

Step 3: - Feature Extraction

In this step the low level features textures are extracted from each image . RGB color block is converted to graylevel and quantized into16 levels to be used with co-occurrence matrix to extract texture features. For each gray block (N blocks) compute co-occurrence matrices of theta (0^o, 45^o, 90^o and 135^o) and distance 1. For each GLCM a set of 6 features (Entropy, Energy, Mean, Contrast, IDM, and homogeneity) is computed and saved in a database called (TRDBF) as shown in Figure (3).

a	b	c	d	e	f	class
13.643583	2.688965	0.542816	2.57815	6.582031	44.527344	horse
13.472475	2.733271	0.806763	2.619996	14.761719	45.105469	horse
13.419602	2.780022	0.601013	2.672169	6.550781	30.925781	horse
14.059173	2.401719	0.532318	2.24227	17.75	56.007812	horse
14.186347	2.456751	0.507233	2.312515	14.398438	52.929688	horse
12.305212	2.684998	0.737732	2.575798	6.867188	29.554688	horse
12.134076	2.936561	0.90799	2.858568	5.882812	40.515625	horse
13.420408	2.645471	0.675232	2.519737	9.171875	37.984375	horse
14.638075	2.754469	0.499207	2.661234	7.167969	42.519531	horse
5.664042	3.228125	2.913086	3.198897	2.609375	46.75	horse
9.473803	3.18484	1.962799	3.131982	5.171875	44.773438	horse
13.661682	2.628692	0.562286	2.506556	8.375	46.765625	horse
15.29452	2.632389	0.417755	2.516476	9.425781	46.917969	horse
9.316912	2.997459	1.460236	2.956016	4.214844	22.972656	horse
10.786045	3.036124	1.161194	2.984183	7.417969	29.425781	horse
9.354463	3.116341	1.296692	3.074839	3.121094	26.363281	horse
10.320734	3.05824	1.456451	2.998075	4.855469	29.332031	horse
7.440756	3.140309	2.080109	3.102489	3.917969	44.425781	horse
5.019239	3.25226	3.367859	3.22597	3.699219	41.449219	horse
14.964729	2.51303	0.466125	2.375582	12.945312	45.804688	horse

Figure (3) TRDBF.

Where(a=Entropy, b=Homogeneity, c=Energy, d=Contrast, e=IDM(Inverse Different Moment, f=Mean)

Step 4: -Coding

In this step the feature vectors values are converted into codes due to max and min attribute values and these codes are stored in TRDBC that can be understood by decision tree as shown in Figure (4) to make decision and used to increase facilities of the work in the decision tree.

a	b	c	d	e	f	classs
a3	b1	c2	d1	e2	f3	horse
a3	b1	c2	d1	e4	f3	horse
a3	b1	c2	d1	e2	f2	horse
a4	b1	c2	d1	e4	f3	horse
a4	b1	c2	d1	e4	f3	horse
a3	b1	c2	d1	e2	f1	horse
a3	b2	c2	d2	e2	f3	horse
a3	b1	c2	d1	e2	f2	horse
a6	b1	c2	d1	e2	f3	horse
a2	b2	c4	d2	e1	f3	horse
a3	b2	c4	d2	e2	f3	horse
a3	b1	c2	d1	e2	f3	horse
a6	b1	c1	d1	e2	f3	horse
a3	b2	c3	d2	e2	f1	horse
a3	b2	c2	d2	e2	f1	horse
a3	b2	c3	d2	e1	f1	horse
a3	b2	c3	d2	e2	f1	horse
a2	b2	c4	d2	e2	f3	horse
a2	b2	c4	d2	e2	f3	horse
a6	b1	c1	d1	e3	f3	horse

Figure (4) TRDBC.

Step 5: -Decision tree

This step which can build decision tree of the input image as shown in Figure (5) based on the values of the given features attributes by using ID3 algorithm.

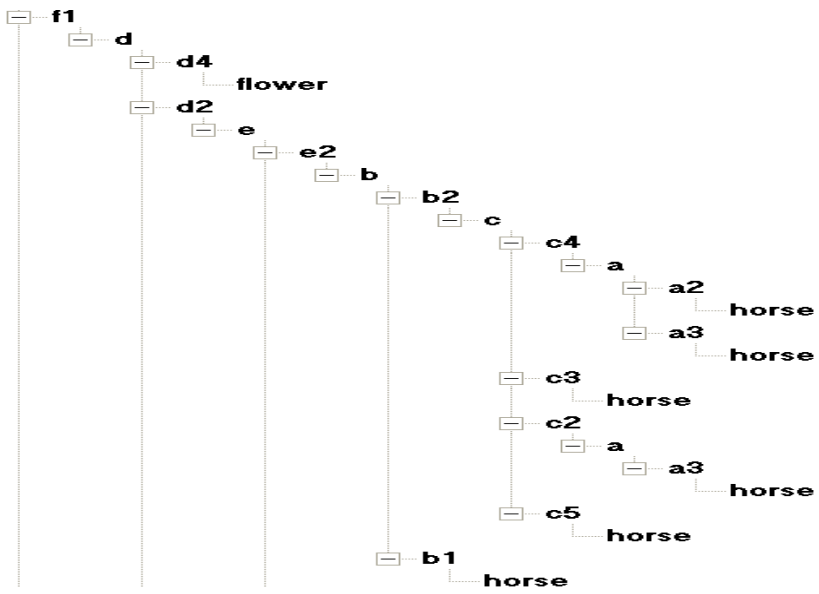

Figure (5) Sample of Decision Tree.

Step 6: - Create Database Rules

As a result many rules are generated and stored in database rules (DBR) due to if-then format which aid in Test phase to classify images and make decision.

Testing phase

All steps in test phase will be as in training phase except one procedure which is called classifying process step as follows:

1. Read input data from TEDBC
2. Locate rules that satisfy image test features in DBR
3. If found then add 1 to count class (kind k)
4. Repeat steps from 1 until not end of file
5. Find max counter of (class kind1, class kind2,…….,class kind k)
6. Make decision.

EXPERIMENTS AND RESULT

The training images in this paper belong to a subset, which is a part of WANG.The subset used consists of 5 classes, that is (dinosaurs , flowers, horses, food and cars) each class contains 75 images, so that the total numbers of used images are 375. Accuracy measure is the most widely used measurement method to evaluate the classification images.For our experiment on 375 images the classification image accuracy by use of decision tree with texture features algorithm is given as many different results, some of which with good accuracy when they have no complex background and no affected block as well as till segmentation aid is used as to remove any features of image that are unwanted data. To compute the accuracy of the decision the following equation is used and gives the accuracy ratio to classify the images. Table (1) explain is the accuracy of kind of images.

$$\textbf{Accuracy} = \frac{\textbf{Number of Correctly Classified Images}}{\textbf{Total Number of Images}} \textbf{ X 100\%} \quad \textbf{…}(8)$$

Table (1) Accuracy of all kinds of images.

Type of Classification	Correct classifiers	Wrong classifiers	No Decision	Accuracy
dinosaur	100	0	0	100%
horse	89	6	5	89%
flower	90	4	6	90%
food	80	9	11	80%
car	81	7	12	81%
Average				88%

From Table (1)the classification result may be one of the following states: Correct classifiers, Wrong classifiers, and No decision.

Correctly classifiers represents the classified image.

Wrong classifiers represents unclassified images.

No decision state happens when the number count of kind class i is equal to number count of kind class j.

Comparison of CIUDT with Others Systems

This subsection evaluates the classification accuracy of the proposed system and compare it with some of the existing system as shown in Fig .6. The result of the paper is compared with the performance of MUFIN [13], Metode Proposta [14], and CBIC [15], LNBNN [16], DTC [17]. It is noted this proposed has higher accuracy due to Table (2).

Table (2) Comparison of accuracy of the proposed system (CIUDT) with previously existing systems.

Proposed System Name	Algorithm	Accuracy
MUFIN	Decision tree (ID3)	80.5
DTC	Decision tree (J4.8)	86.66%
Metode Proposta	Rede Neural+ Naïve Bayes	85,02%
CBIC	Neural network	79,5%
LNBNN	Naive Bayes Nearest Neighbor	71,09%
(CIUDT)	Decision tree (ID3)	88%

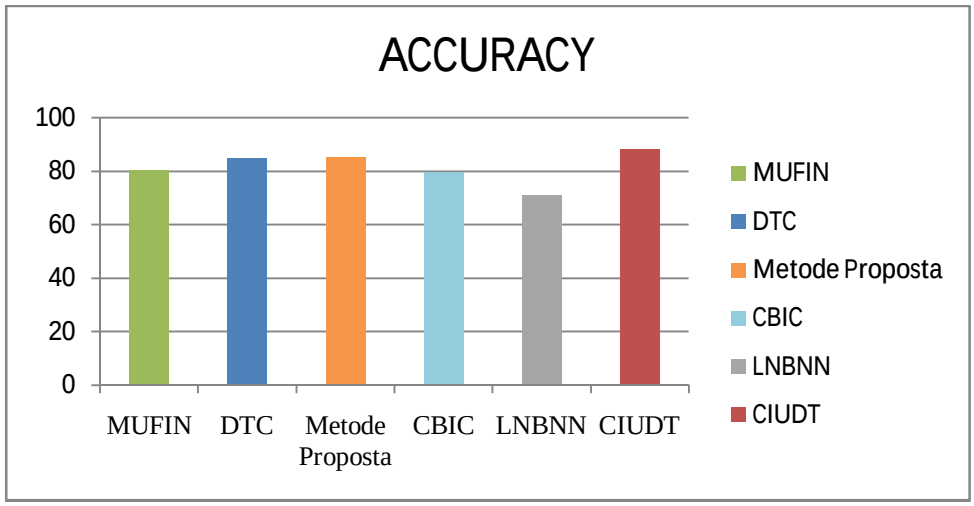

Figure (6) Comparison of Accuracy of Proposed System (CIUDT) with other systems.

From Figure (6) it is noted that the proposed system has higher accuracy than others because we use tile segmentation for the images and that aids in removing any unwanted data and texture feature is used with decision tree.

Contributions

In this paper the main contributions are as follows:

1- Use of textures features with decision tree to classify multi object image while the other researches use texture to classify single object image.
2- Use of tile segmentation by combining the data obtained from each segment to get increase in number of quantifiable features and neglect any part that has little entropy . This aid in preventing any block effect.
3- Use of Low Low sub bands from Haar discrete wavelet transform since Haar can decompose signal into different components in the frequency domain, Low High, High Low, High High, Low Low. Low Low of them represents average component and three other are detail components.

CONCLUSIONS

1- The process of classifying single object image is more easily than multi object image.
2- To prevent any block effect you can divide the image to many sub images as happens when we use segment tile.
3- Decoding method of feature values is a suitable way to reduce data ranges and reduce time of execution of tree building.
4- Quantization method of pixels colors value aids in reducing data size without losing any important data.
5- Classification of simple background image is easier than that of complex background image and there are no general classification systems.

REFERENCES

[1]. Silver. B., "An Introduction to Digital Image processing ", 2000, Available at: www.cognex.cm
[2]. Deshpande. D.S., "Association Rule Mining Based On Image Content", International Journal of Information Technology and Knowledge Management , Volume 4, No. 1, pp. 143-146, 2011.
[3]. Leonidaki E.A., Georgiadis. D. P. , N. D. , "Decision Trees for Determination of Optimal Location and Rate of Series Compensation to Increase Power System Loading Margin," IEEE Transactions on Power Systems , Vol. 21, pp.1303-1310,2006.
[4]. Pham.T.A.," Optimization of Texture Feature Extraction Algorithm" MSc Thesis, Computer Engineering, Mekelweg CD Delft ,The Netherlands , Available at http://ce.et.tudelft.nl/ 2010.
[5]. kusumaningrum. R. R., A. M., " Color and Texture Feature for Remote Sensing – Image Retrieval System: A Comparative Study ", IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 5, No 2, 2011.
[6]. Van. E.L.," Human-Centered Content-Based Image Retrieval" Nijmegen, 2005.
[7]. Gadkari, D. "Image Quality Analysis Using GLCM" , MSc. Thesis, College of Arts and Sciences at the University of Central Florida ,2004.
[8]. Bach. M. P., D. Ć., "Data Mining Usage in Health Care Management:Literature Survey and Decision Tree Application"[1]Ekonomski fakultet, Sveučilište u Zagrebu; [2]Valicon d.o.o., 2007.
[9]. Matthew. N., " Comparative Analysis of Serial Decision Tree Classification Algorithms", International Journal of Computer Science and Security, (IJCSS) Volume (3) : Issue (3), 2005.

[10]. Peng.W., J. C., H. Z., " An Implementation of ID3 - Decision Tree Learning Algorithm", University of New South Wales, School of Computer Science & Engineering, Sydney, NSW 2032, Australia ,2002.

[11]. Harish. D.V.N., Y. S., K.N.V., P. A., "Image Annotations Using Machine Learning and Features of ID3 Algorithm," International Journal of Computer Applications, Vol. 25, No. 5, pp. 0975–8887, 2011.

[12]. Shannon. C. , "A Mathematical Theory of Communication ",Reprinted with corrections from The Bell System Technical Journal,Vol.27,pp.379-423,623-656,July,October ,1999.

[13]. Surynek. P., I. L., "Automated Classification of Bitmap Images Using Decision Trees", Faculty of Mathematics and Physics, Department of Theoretical Computer Science and Mathematical Logic, Malostranské námestí 25, Praha, 118 00, Czech Republic,2011.

[14]. Kalva. P. R., F. E. , A. L. , "WEB Image Classification using Combination of Classifiers" ,IEEE Latin America Transactions, Vol. 6, No. 7, December,2008.

[15]. Park S. B., J.W., S. K. ,"Content-based Image Classification Using a Neural Network", Pattern Recognition Letters 25, 287–300 ,2004.

[16]. Mcann. M., D. G., "Local Naive Bayes Nearest Neighbor for Image Classification", University of British Columbia , Technical Report, 2011.

[17]. Pooja A.P., J. J., K. S., "Classification of RS Data Using Decision Tree Approach "International Journal of Computer Applications (0975 – 8887) Volume 23– No.3, 2011.