Image Understanding Using Decision Tree Based Machine Learning

Chesta Agarwal

chesta.agarwal@gmail.com

Abhilasha Sharma abhilasha sharma87@yahoo.com

Delhi College of Engineering, Delhi University, Delhi, India

Delhi Technological University, Delhi, India

ABSTRACT

Image Understanding, a discipline that concerns the interpretation of an image and analysis of the image to give a decision about the image and the actions represented in it. Decision tree is a tree based classification, widely used in data mining, which classifies the input data set into predefined classes. Decision tree approach is used here to train the image understanding system to perform supervised machine learning. The various low level characteristic features (color, shape, texture) of the image form the various attributes of the decision tree among others. This paper presents the application of the decision tree approach for image understanding. It also discusses an algorithm to calculate the relative distance between the retrieved results, as a sub process required in the proposed approach. The paper describes the production rules required to generate the decision tree. An example study is used to describe the image understanding process in a descriptive manner.

KEYWORDS: Image Understanding, Expectation Values, Production Rules, Machine Learning

1. INTRODUCTION

Image understanding is the advanced image processing sub domain in which artificial intelligence techniques are used to interpret images by locating, characterizing, and recognizing objects and other features in the scene (Dictionary of Cloud Computing, [1]). This process has earlier been mainly done by the human cognitive thinking, but with increasing demand of critical image processing applications the need of an automated image understanding system has emerged, thereby marking the image understanding process as a promising discipline. Apart from the decision making process, the image understanding process includes several sub processes like object identification, image retrieval and recognition of patterns based on their features. The human cognitive thinking can be done by various supervised and un-supervised machine learning techniques. scientists have suggested methods like sparse coding [22], knowledge based inference [26], fuzzy logic [9], neural networks [8] [18] and bidirectional associative memory mapping among the various techniques of machine learning used in the image understanding process. The decision tree approach is discussed in section 2. The various components of a general image understanding process are detailed in

section 3. It explains the inter-relation between the object recognition, pattern recognition and the artificial intelligence based machine learning. The Section 4 describes the proposed approach of the image understanding process. It describes the application of the decision tree as a supervised machine learning approach with various components of an image understanding system. The different steps of the image analysis process are explained with the help of a critical defence application. An algorithm to calculate the relative distance between the recognised objects using the various feature extraction techniques is discussed. The different attributes of the decision tree and the production rules required to generate the decision tree are also described in this section. It also states the various expectation values (based on research) and their correlated production symbols and decision outcomes. Results and a brief conclusion of the proposed approach for image understanding focussing on the potential of the image understanding system are presented in Section 5.

2. DECISION TREE BASED APPROACH

A decision tree is defined as a connected, acyclic, undirected graph, with a root node, zero or more internal nodes (all nodes except the root and the leaves), and one or more leaf nodes (terminal nodes with no children), which will be termed as an ordered tree if the children of each node are ordered (normally from left to right) [25]. A tree is termed as univariate, if it splits the node using a single attribute or a multivariate, if it uses several attributes. A binary tree is an ordered tree such that each child of a node is distinguished either as a left child or a right child and no node has more than one left child or more than one right child. For a binary decision tree, the root node and all internal nodes have two child nodes. All non-terminal nodes contain splits [25].

A Decision Tree is built from a training data set, which consists of objects, each of which is completely described by a set of attributes and a class label. Attributes are a collection of properties containing all the information about one object. Unlike class, each attribute may have either ordered (integer or a real value) or unordered values (Boolean value). These attributes form the internal nodes of a decision tree, while the values of these attributes represent the branches of the tree. Leaf node represents a class of a classifying attribute. It contains records belonging to the same class. Decision tree is

traversed from top to bottom by performing test on each internal node that comes in the way until the leaf node is encountered. If attribute is continuous rather than discrete then a threshold is formed and tests are performed on that threshold value. Several methods ([14] [13] [12]) have been proposed to construct decision trees. These algorithms generally use the recursive-partitioning algorithm, and its input requires a set of training examples, a splitting rule, and a stopping rule. The decision trees are generally categorised as *Classification tree* (predict a class for a new instance) and *Regression tree* (predict a real number).

From the above discussions, it is evident that a decision tree can be used to classify a pixel by starting at the root of the tree and moving through it until a leaf is encountered. At each non leaf decision node, the outcome for the test at the node is determined and attention shifts to the root of the subtree corresponding to this outcome. This process proceeds until a leaf is encountered. The class that is associated with the leaf is the output of the tree. A class is one of the categories, to which pixels are to be assigned at each leaf node. The number of classes is finite and their values must be established beforehand. The class values must be discrete. A tree misclassifies a pixel if the class label output by the tree does not match the class label.

Amongst other decision methods, decision trees have various advantages. Decision trees have the ability to handle discrete attributes as well as continuous attributes. Unlike other classification techniques which require data normalization, removal of missing attribute values etc.; the decision tree just requires simple data preparation. Decision trees are very easy to understand and visualize. Hence they can be easily tested and validated. In other words statistical test cases can easily be applied on the decision tree model to check the reliability of the model. Decision trees work well for large amount of data. Large amount of data can be analysed in lesser time. They are known for their good performance against larger data sets.

However, the decision trees also have some limitations. Practical decision-tree learning is based on greedy algorithm where locally optimal decisions are made at each level of the tree. Such algorithms cannot guarantee to return the globally optimal solutions. That is why the problem of learning with Decision tree is NP Completes. A tree generated may overfit the training examples due to noise or too small a set of training data. This is called over fitting in classification.

3. IMAGE UNDERSTANDING

When we look at an image, we can know some information without any other hints, for example, dark or light, etc. We can easily infer the different objects present in the image, the number of instances of objects of each type and their relationship with each other. That is because there exists a knowledge base in our experience and we can use the knowledge and extract the global information existing in an image. Moreover, we will take the image to be processed based on the knowledge, e.g., a dark image need to be

enhanced in lightness, etc. However, such dependency on human thought process is not always reliable, especially in various critical applications of the scientific domain. Various image processing applications of medical, defence and other critical applications require image analysis understanding operations. Image Understanding is a sub domain of the broader science of image processing. It involves analysing the image and generating an inference on the various aspects of the image, as per the application involved, a process known as scene analysis. The level of understanding in an image increases with the context the image and its data are linked together. The relationship between the context of the image data and the level of image understanding is represented in the Figure 1.

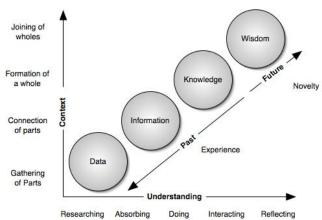


Figure 1: Relationship between image data and level of image understanding [2]

The process of image understanding is used to predict the unknown i.e. the future; this would require the knowledge of the past. The understanding of the image is said to be at the researching stage when the parts of the image objects are gathered together to obtain data. This data when linked to each other through relationships leads to information which is absorbed by the system. The absorbed information is used to form the knowledge base through certain processes acting on it to refine the relationships of the data and generate the individual modules of the system. The knowledge generated can be used to reflect into the future and thereby join the entire system as a whole. This reflection into the future comes through the machine learning in the automated image understanding system. Thus, the level of description of the image data is closely related to the level of understanding of the system. Image understanding however requires both past experience and an insight to predict the future.

3.1: Steps involved in Image Understanding

The process of image understanding or scene analysis would require simulating the process of human intuitive thinking. Apart from the steps of image processing common to most systems, this process involves many processes which can be explained by human intuitive thinking, but always implemented by the numerical methods. For example, according to some transformations or statistics of an image,

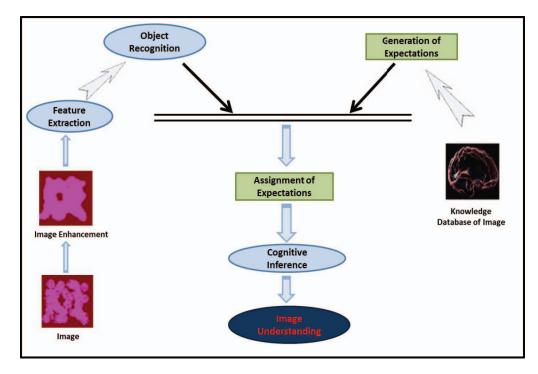


Figure 2: Image Understanding steps

we want to find out the overall characteristics in an image, e.g., in histogram statistics, it can be seen that the image is biased to dark or light and by frequency transformation, the noise can be roughly predicted and filtered. Corresponding to each result, the images are mapped by a global feature and a linguistic meaning, just like human think. And, after the image improvement, sub-global information of an image is considered to extract the object [7]. The object in an image can be easily recognized by human intuition. That is, there is a knowledge base in the brain to know the structure of an object. The task of understanding analysed data is complex because it is based purely on human processes, which comprise many different, important thought processes, including:

Perception: differentiating, recognising, perceptual categorisation, orientation,

Attention: unintentional attention, free attention, *Memory:* declarative, sensory, short-term, long-term (episodic, semantic, contextual), implicit memory, *Thinking, calculating, abstracting*,

Forming concepts,

Taking decisions, planning, predicting, problem solving, Understanding

Human thinking processes leading to interpreting and describing complex data form the foundation for designing computer data analysis systems. The thought processes taking place in the human brain are used as the basis for designing a computer system which attempts to imitate natural processes and thus aims at the automatic, computer analysis of data [15]. Each of the human thought process

involved in the image analysis is simulated though the various image processing techniques to accomplish the task of setting an automatic image analysis system.

The various steps of an image processing system can be represented as in Figure 2.

The various steps in an image understanding system can be enumerated as follows:

Image Enhancement: Image Enhancement process prepares the input image for the further image processing steps. It denoises the image using Fourier Transform and other approaches. Image enhancement also resizes the image to make it suitable for the subsequent feature extraction techniques to be applied.

Feature Extraction: Image understanding requires that the individual components of the image are represented by the features that best represent their properties. For example, a rocky terrain would be more suitably represented by the texture features ([3] [6] [23] [20]) rather than the color ([17] [19]) or shape features ([5] [10] [11] [16]). Also, a clearly distinguishable object would have sharper and distinct shape features as compared to the color and texture features.

Object Recognition: Object recognition stage of the image understanding process is used to extract the important objects in the image (scene). A relationship is then set up between the objects identified. The Object recognition is used to identify the important characteristic traits of the

image and hence help in better descriptive analysis of the image (scene).

Generation of Expectations: Based on the previous knowledge database and the research done from the existing knowledge, expectation values are generated. These expectations are used to ease the decision making process. For example, an image with bright clear sky and a bright sun would definitely represent an image of broad day light. Hence the expectation value of an object representing the sun would be higher towards representing a day and lower otherwise. Expectation values represent the probability of a result in case a particular object is recognised in the image.

Assignment of Expectations: The expectation values generated according to the knowledge base are assigned to the objects recognised by the feature extraction phase on the input image. The various objects are assigned an expectation value.

Cognitive Inference: Based on the expectation values of the individual objects in the image, the overall expectation value is assigned to the image. This expectation value is thereby used by the image understanding system to generate the final decision about the scene. The various techniques used to generate the result are based on machine learning (either supervised or unsupervised). They are based on the fact that the system is trained on a training set to generate an inference about the characteristic of the image.

4. DECISION TREE BASED IMAGE UNDERSTANDING: PROPOSED APPROACH

The Image Understanding process involves generating a decision about the objects in the scene represented by the image. A general image understanding system can be illustrated as in Figure 3.

The various functional blocks of the system include an Image Handler, a Pattern Recognizer, a Feature Extractor, a Similarity Comparator, a Database Handler, an Image Analyser, a Decision Generator and a GUI handler amongst its major components.

The *Image Handler* module analyzes the image format of the input image. It reads the input image and converts the image into a data matrix which stores the color components and intensity values according to the different color spaces for each pixel. If RGB color space is considered for example, then the data matrix would be a 3 dimensional matrix of the dimension *Height X Width X 3*; storing the R, G and B component values for each pixel. The Image handler module is also responsible for converting the data matrix back to the image file while saving the results generated by the system. The *GUI handler* module manages the entire GUI of the system required to ease the end user in specifying the query and to view the results. The *Feature*

Extractor module is responsible for extracting the various low level and high level

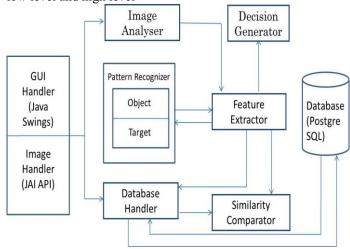


Figure 3: Block Diagram Of An Image Understanding System

features of the image using the various techniques for extracting the color, texture and shape parameters of the image. It would calculate the features of the image as a whole and of various sub images within the main image. The features are then saved in the database in the form of tuples of the different tables. The feature vector comprises of the pixel coordinates of the top left corner of the sub image along with the feature parameters corresponding to the method used for feature extraction. A typical feature vector looks like,

$$F_i = \{ x_i, y_i, f_1, f_2, f_n \}$$
 (1)

Where, F_i is the feature vector corresponding to the ith sub image. x_i and y_i are the pixel coordinates of the top left corner of the rectangular sub image. And f_1 , f_2 ,.... f_n are the parameters generated by the various techniques. The feature vectors corresponding to all the sub images or the image as a whole are stored in the database for future retrieval by the Similarity Comparator module. The storing of these feature parameters and their retrieval is done from the database by the Database Handler module. The calculation of the image parameters is a one-time task thereby reducing the running time and the response time of the system required to generate the results. The system does the task of understanding the input query and calculating the similarity between the input query image and the images or the sub images in the database. This task is done by the Similarity Comparator module of the system. The similarity may be measured using the Euclidean distance measuring technique. The Image Analyser is used to analyse the relative distance between the objects identified in the image. It also analyses the ground texture (metalled road, sandy plain, etc.) represented in the image. Using the information about the objects, the texture and other information gathered from the image (by the Pattern Recogniser module), the decision tree approach is used with expectations values assigned to the image according to the various characteristic features of the image. The generation of the decision tree and the analysis of the image, based on the low level feature parameters of the image, is done by the *Decision Generator* module.

The integration of the decision based approach with the image understanding process is detailed and exemplified by the subsequent sections.

An Application of Image Understanding System

Image Understanding system can be used in various critical applications like defence and medical imaging. An example pertaining to the defence application; is to locate the air bases constructed by various militant groups settled around the high security areas. It distinguishes a government established airport from the one built be militant groups for air attacks. It is also used to distinguish an airport hangar from a runway. The similar approach can also be used in case of shipyards by providing a different image database and training the system to generate the decision accordingly. The image understanding scheme proposed here uses object recognition i.e. image retrieval to identify the aircrafts. The retrieved results are used to generate the inference with the help of the decision trees.

The various steps involved in generating the final inference about the image are:

1. *Image Enhancement:* The de-noising approach of image centering is performed prior to the Fourier Transformation. *Image Centering:* This process minimises the error propagated in the satellite input image feed. It involves a scalar point-to-point multiplication of the image pixel matrix with the scaling factor.

Scale Factor: $(-1)^{(x+y)}$

- 2. Feature Extraction: Image understanding requires that the individual objects in the image are represented by the low level features that best represent their properties. The aircrafts in the input satellite feed are represented by the moment invariant shape feature ([10] [16]). The aircraft is a clearly distinguishable object hence, it would have sharper and distinct shape features as compared to the color ([17] [19]) and texture features ([6] [20] [21]). The rocky terrain is represented by the texture features (GLCM [6] [19] and Gabor Filter [3] [4] [27]) instead of the color or shape features. The color features (color moments, color histogram and fuzzy color moments [19]) are used to further distinguish the sandy plains from the other plain textures in the image. Thus depending on the application, the different feature extraction techniques can be used to represent various visual characteristics of the image.
- 3. Object Recognition: Object recognition stage of the image understanding process is used to identify the different objects on the basis of the features extracted during the

feature extraction phase. It categorises the different objects into classes and sets a relationship between the objects of a similar class. The different objects in the scene are identified by computing the Euclidean distance between the various image templates.

- 4. Generation of Expectations: The decision tree being a supervised machine learning process uses the expectation values as generated by the research work from the previous knowledge database. The presence of the aircrafts in the image would have a higher expectation value. Expectation values represent the probability of a result in case a particular object is recognised in the image.
- 5. Assignment of Expectations: The expectation values generated according to the knowledge base are assigned to the objects recognised by the feature extraction phase on the input image. The various attributes of the decision tree are represented by the features and the objects identified in the feature extraction phase and the object recognition phase respectively. Apart from these attributes, the decision tree has an attribute of the relative distance between the objects identified. This relative distance between the shape objects is evaluated as per the algorithm 1. The assignment of expectations is done as per Table 1. The decision tree (Figure 4) generated through the production rules is presented further to.

Assumption: All the object templates are considered as nodes and are referred to by the whole numbers in the sequence they are retrieved by the object recognition phase.

Variable List: The current node being evaluated = currentNode; The node closest to the node under evaluation = nearestNode; Minimum distance between the two nodes under evaluation = minDistance; Total number of nodes to be evaluated = totalNodes; Array for storing the relative distance between the nodes = distance[totalNodes-1]; Temporary variable storing the temporary distance evaluated = tempDistance

Algorithm 1: To calculate relative distance between the shape based objects

Input: The array node[][] containing the list of nodes of object templates and their x-y coordinates.

1. Initialise the nearestNode with the first input node (marked by 0) and the distance[] array with a very large value, say, INFINITY.

```
i.e. nearestNode = 0;
distance[] = INFINITY;
```

2. For all the nodes except the last node perform the following sequence of steps:

- Swap node[nearestNode][0] with the node[currentNode +1][0];
 Swap node[nearestNode][1] with the node[currentNode +1][1];
- Swap distance[nearestNode-1] with distance[currentNode];
- 5. Calculate average distance from the distance[] array.

The decision tree shown in figure 4 can be represented by the production rules which assign the expectation values to the decision path. The various symbols used in the production system are:

m = color & texture features; o = shape objects; t = number of templates; d = relative distance between templates.

```
of templates, d = 1 relative distance 1. R1: E \leftarrow (m) (\sim0)

2. R2: E \leftarrow (\sim m) (\sim0)

3. R3: E \leftarrow (m) (0) (t) (d)

4. R4: E \leftarrow (m) (0) (\simt) (d)

5. R5: E \leftarrow (m) (0) (t) (\simd)

6. R6: E \leftarrow (m) (0) (\simt) (\simd)

7. R3: E \leftarrow (\sim m) (0) (t) (d)

8. R4: E \leftarrow (\sim m) (0) (\simt) (d)

9. R5: E \leftarrow (\sim m) (0) (t) (\simd)

10. R6: E \leftarrow (\sim m) (0) (\simt) (\simd)
```

Here if the color and texture features suggest a metalled surface then the symbol m will be true and it will have the value assigned as stated in the expectation table (table 1). Otherwise the symbol representing un-metalled surface will be represented by the symbol $\sim m$. The presence of shape objects, representing the aircrafts is represented by the symbol o, while $\sim o$ denotes the absence of any aircrafts in the input GIS image. The symbol t represents that the number of templates representing the shape objects in the image are more than the threshold value t_1 , while $\sim t$ represents the number of templates to be less than the threshold t₁. The relative distance between the retrieved shape templates is calculated as per algorithm1 and is represented by the symbol d, if the relative distance is less than the threshold distance d₁. If the distance is greater than the threshold d_1 then it is represented by the symbol \sim d. All the symbols are assigned expectation values according to table 1. These values are then used to assign values to the symbol E, which represents the final expectation value at a decision path.

Symbol	Expectation Value
m	0.95
~m	0.05
o	0.9
~0	0.1
t	0.7
~t	0.3
d	0.6
~d	0.4

Table 1: Expectation values for the production symbols

6. Cognitive Inference: Based on the expectation values of the individual objects in the image, the overall expectation value is assigned to the image with the help of decision tree. This expectation value is thereby used by the image understanding system to generate the final decision about the scene as per Table 2. The various techniques used to generate the result are based on machine learning (either supervised or unsupervised).

Expectation Value (E)	Decision Outcome
0.095	Metalled Terrain
0.005	Rocky Terrain
0.3591	Busy Airport Hangar
0.1539	Airport Hangar
0.2394	Busy Airport Runway
0.1026	Airport Runway
0.0189	Busy Airbase Hangar
0.0081	Airbase Hangar
0.0126	Busy Airbase Runway
0.0054	Airbase Runway

Table 2: Decision outcomes for the various expectation values

Observation & Results

The decision tree based machine learning approach generates the output according to the training data set used. Table 3 illustrates some results generated according to the proposed approach. The results identify the various components of the image and generate a decision accordingly. The estimation values generated by the research impact the decision tree outcome drastically. The higher expectation values depict the presence of the various objects.

CONCLUSION

The decision tree based machine learning approach is effectively applied to the image understanding system. However, the correctness of the results generated according to the proposed approach largely depends on the following: estimation values (generated by the research), object recognition process, techniques used for extracting the different color, texture and shape features. The estimation values used in the decision tree based image understanding system are particular for the specified example. The approach and the algorithm can however be applied to any example system. The decision tree based approach used in the image understanding system targets the ease of the decision tree generation as compared to the other approaches used for the cognitive inference in an image understanding system.

REFERENCES:

- [1]. A Dictionary of Computing. 2004. Web link: Encyclopedia.com:http://www.encyclopedia.com/doc/1O11-imageunderstanding.html
- [2]. Continuum of understanding. Web link: http://www.nwlink.com/~donclark/performance/understanding.html
- [3]. A. Jain, N. Ratha, S. Lakshmanan, "Object detection using Gabor filters", In Pattern Recognition, vol. 30, pp. 295-309, 1997.
- [4]. A. K. Jain, and F. Farroknia, "Unsupervised texture segmentation using Gabor filters," *Pattern Recognition*, Vol.24, No.12, pp. 1167-1186, 1991.
- [5]. A. Sajjanhar, and Lu, G., A Grid Based Shape Indexing and Retrieval Method, *Australian Comput. J.*, 1997, vol. 29, issue 4, pp. 131–140.
- [6]. B.S. Manjunath and W. Y. Ma, "Texture Features for browsing and Retrieval of Image Data", IEEE Transaction on Pattern Analysis and Machine Learning, Vol. 18, No. 8, Aug, 1996
- [7]. Bor-Tow Chen , Yung-S heng Chen, Wen-Hsing Hsu, Image Processing & Understanding based on the Fuzzy Inference Approach, 0-7803-1896-X/94, IEEE 1994
- [8]. E., Barnard and D., Casasent, "Image Processing For Image Understanding With Neural Nets"
- [9]. F., Jurij, KoSir, Andrej, TasiE, "Formal system based on fuzzy logic applied to digital image scene analysis"
- [10]. Hu, M.-K., Visual Pattern Recognition by Moment Invariants, *IEEE Trans. Information Theory*, 1962, vol. 8, issue 2, pp. 179–187.

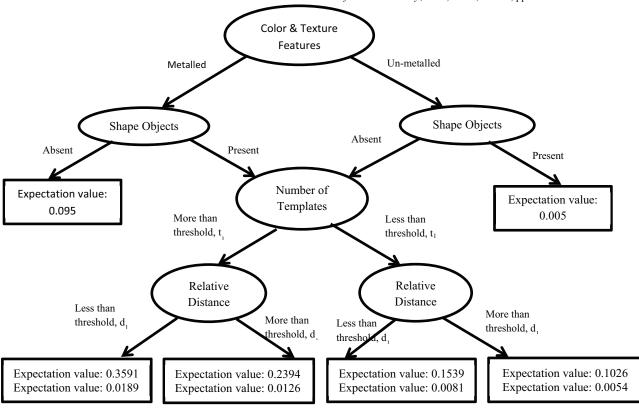


Figure 4: Decision Tree Generated

- [11]. J. D. Foley, A. van Dam, S. K. Feiner, and J. F. Hughes, *Computer graphics: principles and practice*, 2nd ed., Reading, Mass, Addison-Wesley, 1990.
- [12]. J. R., Quinlan, 1975. Machine Learning, Vol. 1, No. 1, University of Sydney
- [13]. J.R., Quinlan, 1993. C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, San Mateo, California, 302 p
- [14]. L., Breiman, J.H., Friedman, R.A., Olshen, C.J., Stone, 1984. Classification and Regression Trees, Wadsworth International Group, Belmont, California, 358 p.
- [15]. L., Ogiela, Marek R. Ogiela, Cognitive Approach to Bio-Inspired Medical Image Understanding, IEEE 2010, 978-1-4244-6439-5/10
- [16]. L. Yang, and F. Algregtsen, "Fast computation of invariant geometric moments: A new method giving correct results," *Proc. IEEE Int. Conf. on Image Processing*, 1994.
- [17]. M. Stricker, and M. Orengo, "Similarity of color images," SPIE Storage and Retrieval for Image and Video Databases III, vol. 2185, pp.381-392, Feb. 1995
- [18]. M. Zhenjiang, Yuan Baozong, "A NN Image Understanding System for Maps and Animals Recognition", at IEEE Tencon, '93, Beijing
- [19]. N.S.Vaselier, "Content-based Image Retrieval Methods". Programming and Computer Software, 2009, Vol. 35, No. 3, pp. 158-180

- [20]. P. Brodatz, "Textures: A photographic album for artists & designers," Dover, NY, 1966
- [21]. P.S.Hiremath, S. Shivashankar, and Jagadeesh Pujari, "Wavelet Based Features for Color Texture Classification with Application to CBIR". International Journal of Computer Science and Network Security, Vol. 6 No. 9A, Sept, 2006
- [22]. Q., Hou, HePing Pan, Juan Li, Ti Wu, "Natural Image Understanding via Sparse Coding", DOI: 978-1-4244-5824-0, 2010 IEEE
- [23]. T. Barbu, "Content-based Image Retrieval using Gabor Filtering" in 20th International Workshop on Database and Expert Systems Application, 2000
- [24]. T. Barbu, "Novel automatic video cut detection technique using Gabor filtering", Computer and Electrical Engineering, DOI:10.1016/j.compeleceng.2009.02.003.
- [25]. T. H., Cormen, C. E., Leiserson, and R. L., Rivest, 1989. Introduction to Algorithms, MIT Press, Cambridge, MA
- [26]. T Tjahjadl and R Henson, "A Knowledge Based System For Image Understanding"
- [27]. V., Manian and Marcel J. Castro, "Color Image Classification using Gabor Filters", Computing Research Conference Proceedings, May, 2000

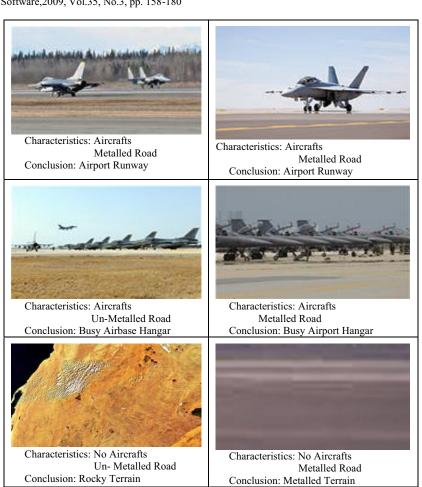


Table 3: Results generated