International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 5 Issue: 6 145 – 149

Digital Image Processing through Hierarchical Clustering Methods, Tree Classifier of Data Mining

Reena Hooda*

*Assistant Professor, Department of Computer Science & Applications,
Indira Gandhi University Meerpur (Rewari), Haryana, India.
Email: reenah2013@gmail.com

Abstract: digital image processing is one of the research areas for applicability of the data mining techniques. Clustering, k-nearest neighbor, decision tree classifier, neural networks are important techniques of mining that are useful for data analysis in most of the fields. Mining methods can works with the text, numbers, plus the multimedia like video and images also. The present paper target toward the analysis of the digital images stored in a system and this procedure includes many tasks say importing of the images for input, processing the images in that numeric values are computed for the given attributes of the input images, then performing mining operations to deduce the results. Due to advancement of mining tools this task is less time consuming and cost effective too moreover flexibility is provided to the user to select the desired method to get the different view of the outcomes.

Keywords: *Digital Image, Mining, Hierarchical Clustering, Tree.*

*****

1. Introduction

Image processing is computer based technology, a core research area within engineering and computer science includes analysis, interpretation and manipulation of a digital image,performing operations in which input is the image and output is image enhancement, feature extraction, having some parameters and values for that. [1] [3][9].Image processing can be viewed as the graphics or vision.Graphics covers the animated movies having series of pictures whereas vision includes the high level processing of 3-D objects or videos [1]. At the low level it is the image processing and at higher level is the computer vision that includes object detection, recognition, shape analysis, tracking and use of AI & machine learning. [5] Purpose of image processing may be the visualization enhancements like brightness, contrast, sharpness, image wrapping, blending & morphing [6]age retrieval& recognition and pattern recognition etc. [3] All is user dependent and may be biased as no fixed criteria of the change is there. The present paper concentrate on thegraphic image only stored in electronic forms i.e. digital image processing to give reliable processing results, excluding analog image processing in which signals may be distorted, noised during processing. Digital image having at least 2 dimensions x and y can be modeled as multidimensional system. [2] Digital image is nothing but a 2-dimensional array of numbers having range from 1 to 255 [4],as camera occupies a 2-D image out of the real scene [9] called the image acquisition and storage. Computer graphics deals operations on the sored images rather than capturing those, or it may also include creation of edges from the objects then developing pictures and enhance & manipulate. Sometimes applicability of artificial intelligence in image processing is must for instance, in medical diagnosis, or criminal records generation and face recognition. [4] Operations of image processing includes clustering, calculus, linear algebra, probability & statistics[6], graph and tree, regression,distance measures, scatter plotting etc. performed mostly in current application. Techniques used in digital image processing are image editing, image restoration, clustering, KNN, Testing and scoring, neural networks, independent component analysis[2], painting values, comparisons, logistic regression and decision tree etc. Digital image processing or manipulation may include pre-processing, feature extraction, selection of training data, classification method including supervised or unsupervised, information extraction, data compression, find the patterns that can't be seen and plotted manually. [3] [5]In graphics this brightness sometimes found low or high depends upon the surroundings and Luminance i.e. intensity or the quantity of the light reflected by the object for example TV screen is grey but converts to black after switch it on [9]. Contrast is the luminance difference between the surroundings and the object, higher the luminance, larger the the difference is required to get the same contrast. There are only 128 full saturated colors can be distinguished in visible wavelength of 350 mm to780 mm. different types of color modes are RGB having component red, green and blue, HSV having 3 components that are Hue, Saturation and value, third is XYZ mode. Hue indicates dominant wavelength of a color, redness of red, saturation defines the purity of the color or mixing of white color, pink red, blue skyblue and value is the intensity of the light or the luminance value. [9] Astrology, astronomy &satellite pictures, natural sciences, biometrics, correcting illumination,contrasting, corrections or rescaling etc. of personal photos[12], edge detection or the color detection,

region detection like MRI, X-rays, Gamma rays that can be seen by camera not by human eyes [8] are some of major applications of image processing. Others are Segmentation, image compression, biological vision, character recognition say zip code or license or number plate recognition, finger print, product inspection, scouting [9] and repairing [6].Hybrid techniques also incorporate image processing; other applications arereproduction, registration. High resolution display [10], mapping earth's surface [11], machine vision, encoding & transmission of images, remote sensing and small damages etc. [8]Remote sensing includes the detection of the damaged area caused by earthquake through extraction of edges of aerial pictures or satellite images. [8] In robotics it may include the hurdle detection or walk through the edges, video processing includes the quality enhancement through the inclusion of the more number of frames through frame rate conversion [8], biological vision may helpful in the brain study as the 1/3rd of the cerebral cortex of brain is committed to graphicdata that is perceived, processed and interpreted by human brain. [9]Mapping the earth's surface, watching its change in ecosystems, finding new resources is a great application in the area of geography. It merge the computer science, geo, engineering, management, economics, statistics to process the acquired images with machineries, artificial intelligence or the human vision to create a data warehouse. This knowledge base can be used for future projections about the conditions of particular area, environment change, or natural disasters, movements of all living beings and reasons behind those all are based on image processing. [11]

2. Historical Survey

The techniques of digital image processing developed it is stated that during 1960s, the different techniques of image processing came into actuality that introducesby Jet Propulsion Laboratoryof Massachusetts Institute of Technology and Bell laboratories from University of Maryland and others. [2] Initially, cost of systems and software were very high and time consuming however, in 1970s these problems became smaller as results of technological reforms. In 2002,Gradient Domain Image processing announced by RaananFattelthat based on the calculating differences between the pixels not by manipulating values [2]. Processing of digital images applies complex algorithms to perform classification, feature extraction, pattern recognition, projection and transformations etc. [2]The very first image transmitted from London to New York through a submarine cable took 3 years to reach. [8]

3. Methodology used for Digital Image processing

For the current application, hierarchical clustering, tree viewer methods of mining are used. The input images taken from the stamps files of the Tuxpaint, and for image processing Orange Data Miner is used. Hierarchical methods of clustering are Single, Average, Weighted, Complete and Wards methods out of which single, complete and wards methods are applied to view the different hierarchies of clustering. For classify the data before clustering distance vector is used to find the distances between the data. The various distance methods are available for example Euclidean, Manhattan, Cosine, Spearman, pearson,andjaccardetc out of which Cosine distance which seems fine for the images and Euclidean distance between the data values is applied. To get familiar with the Orange, different online tutorials including YouTube tutorials and documentation is studied. For data representation of the images, tree view is also developed plus the validation of the model building test & Score widget is applied to watch the confusion matrix and output of the testing with the output of the model.

4. Model building, Processing and Testing

Common task in image processing may involve the image acquisition, storage, image enhancement, restoration,manipulations to create a knowledge base including the image compression, image segmentation, image descriptions and representations and classifications etc. [9] For the current application, the very first step of the image processing is the importing of the digital images that are stored in the TuxPaint folder in program files in C drive. Then another widget called image Embedding is must to view the additional details for the images that takes the images as the input and gives the data as the output that can be viewed by the data table widget.Data table option is dragged on the canvas to view the stored embedded information of the images that takes input from the image embedding. Before performing the clustering, one has to select the distance measure by right clicking the mouse and selecting the distances widget that takes input from the image embedding. Then Hierarchical clustering is selected that connect two other widgets, one is the distances as input edge and other is the image viewer as the output. The whole process of model building for image processing using different widgets in Orange is shown in Figure 1. Dispersion of the values can also be analyzed through scatter plot along with user defined axis data, jittering percentage, opacity and regression line.

 ISSN: 2321-8169
145 – 149

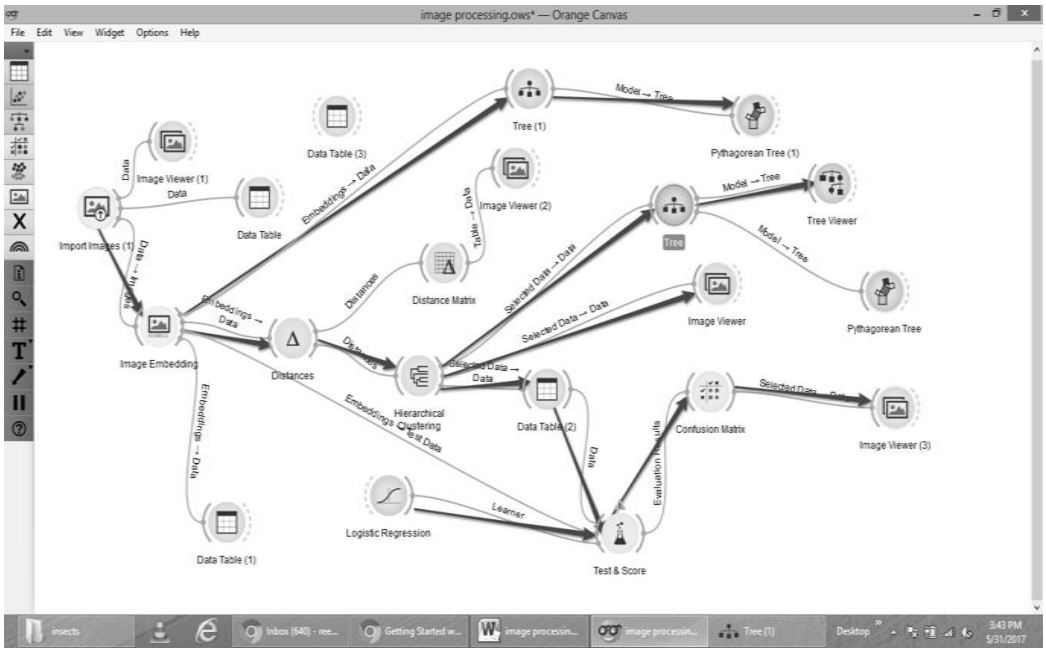

Figure 1: Creating model for image processing in Orange for different analytical operations.

Different distance measures can be opted by double clicking the distances widget and likewise methods of clustering can also be changed by double clicking hierarchical clustering widget also one can select the different option in it. Option are whether pruning on the clustering levels is applied or not, if not then the number of nodes that a user has to defined for hierarchy. Another option is performing clustering by manually selected cluster shown in figure 2 (left) or the height ratio or the top N. When the number of clusters selected as top N, user defined value entered is 14 that will be connected to image viewer as output, it is to be noticed that now after selecting 14, and the clusters of images are titled from 1 to 14 in output view as shown in Figure 2 (right).

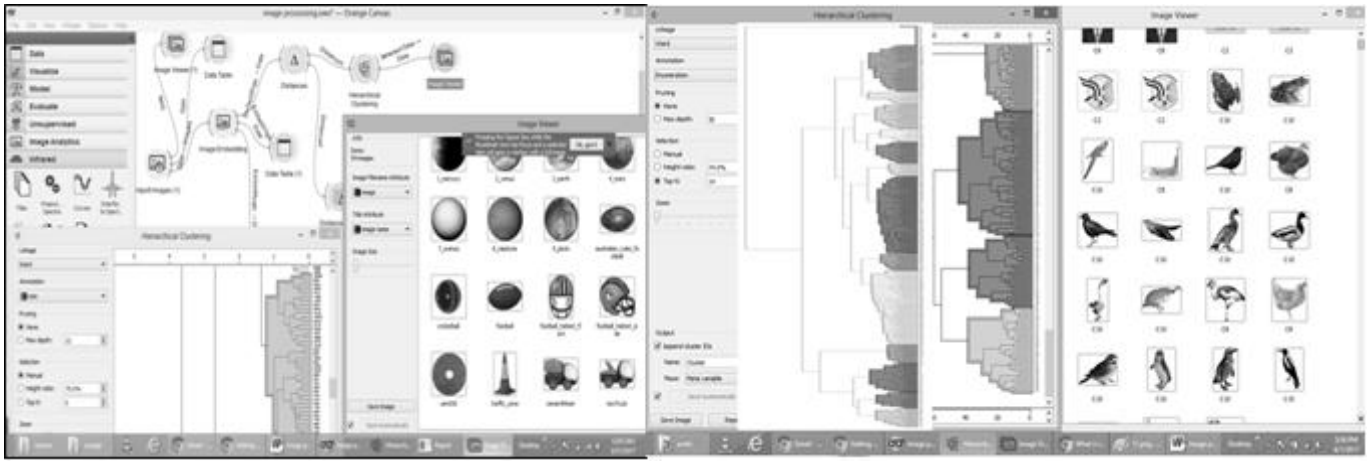

Figure 2: Shows the Hierarchical clustering and corresponding output images for the manually selected (Left) and Top N (right) clusters.

The output can also be adjusted to show the Meta variables. Class name or attributes name. Different annotations can also be given for example, size, width, height, Enumeration, category and image name etc., out of which one can be selected. The images can be further classified as a tree to get the different classes on different nodes. To make the validation test & score widget is employed takes input from the selected data form hierarchical clustering and by using the logistic regression methodology that shows AUC, CA, F1, Precision and recall values 0.949, 0.596,0.552, 0.539 and 0.596., these values are all near to .1000 for the selected cluster as in figure 3 also shows the confusion matrix and the output images. As shown in figure 3, the output of the clustering and test score is the same thus classifier can also be validated.

 ISSN: 2321-8169
145 – 149

Figure 3: Shows the validation of the clustering as output image viewer shows the same results.

Further it is not necessary the output will be same for the image viewer of the model and image viewer of the testing. A good example is shown is figure 4 in which only 8 images are selected out of the 26 images that are more near by selecting the select correct option in the confusion matrix, if option select misclassified is selected all other 14 images are displayed by the image viewer thus this makes the more accurate and reliable results of the classification as shown in figure 4. Test score can work with more than one learner and the results of the each learner are selected independently in the confusion matrix. Advantage of the two or more learners are that there results can be compared to find the most common output values in all the learners. Here output given by the logistic regression is more accurate than the others.

Figure 4: Show the output of test & score with different learners while selecting ‘select correct’ option in confusion matrix.

5. Conclusions

It’s solely the user’s own perspective to analyze the data for a given application. In the present paper efforts are devoted to perform the clustering on the images and view images by selecting the different clusters and options given in widgets. The approach is further validated by imputing the learners to the testing module to watch the similarities or dissimilarities between the hierarchical clustering modules and the testing modules. Further the validation become more significant while comparing the outcomes of different learners by selecting those exclusively in the confusion matrix. The pictorial presentation of Orange made the application development interactive and effective. The further scope of the current application will be finding edges of images and computing different operations by applying the relevant tools of data mining.

References

[1]. Wikipedia [2017].“Image processing”.From Wikipedia, the free encyclopedia.Retrieved from: https://en.wikipedia.org/wiki/Image_processing.

[2]. Wikipedia [2017].“Digital Image processing”.From Wikipedia, the free encyclopedia. Retrieved from: https://en.wikipedia.org/wiki/Digital_image_processing

 ISSN: 2321-8169
145 – 149

[3]. https://www.engineersgarage.com/articles/image-processing-tutorial-applications

[4]. tutorialspoint [2017]. "Digital Image Processing Introduction". Retrieved from: https://www.tutorialspoint.com/dip/image_processing_introduction.htm

[5]. http://www.webopedia.com/TERM/I/image_processing.html

[6]. CS 4640 [2012]: "Image Processing Basics". Retrieved from: http://www.coe.utah.edu/~cs4640/slides/Lecture0.pdf

[7]. https://sisu.ut.ee/imageprocessing/book/1

[8]. https://www.tutorialspoint.com/dip/applications_and_usage.htm

[9]. http://fourier.eng.hmc.edu/e161/lectures/e161ch1.pdf

[10]. http://www.engpaper.com/application-of-digital-image-processing.htm

[11]. Mohammad Sohrabiniaa ,SaeidSadeghianb , DadfarManav [2008]. "APPLICATION OF IMAGE PROCESSING AND IMAGE ANALYSIS METHODS FOR LARGE SCALE MAP REVISION". The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. XXXVII. Part B4. Beijing 2008. Retrieved from :

http://www.isprs.org/proceedings/XXXVII/congress/4_pdf/237.pdf