ISSN:0975-9646

Color and Shape Feature Extraction and Matching in Pill Identification Systems

Dr.E.Annasaro
Director, Deptartment of MCA,
S.N.R.Sons College of Arts and Science, Coimbatore .

A.Hema
PhD. Research scholar,
Manonmaniam Sundaranar University, Tirunelveli.

Abstract-Image mining is the concept used to detect unusual patterns and extract implicit and useful Informations.Mining Image data is one of the essential features in the present scenario. Image data is the major one which plays vital role in every aspect of the systems like business for marketing, hospital for surgery, engineering for construction, Web for publication and so on.Image mining is an extension of Data mining.This paper describes about the implementation and tests to determine whether a given pill is legal or illegal based on the feature retrieval like shape, color, text etc. In this paper color and shape feature extraction were performed with color histogram for color feature and geometrical gradient vector algorithm for the shape. The color and shape feature of the pill is compared only which is matched with text feature. The pills matched with all the features like text, color, shape is said to be legal pill and mismatch with any one of the feature is declared as illegal pill.

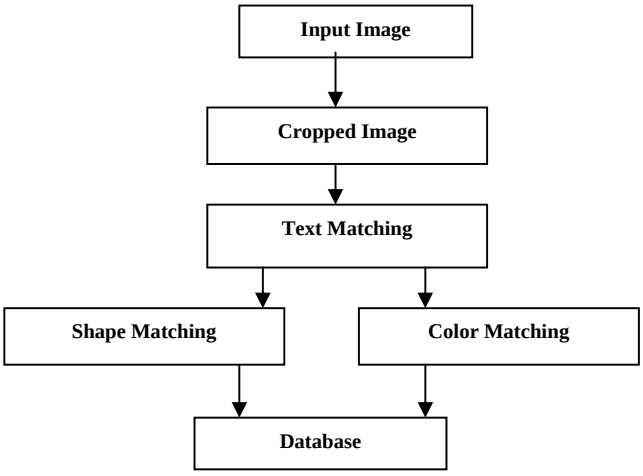
Keywords: Feature extraction,Matching,Feature retrieval, Image mining.

1.INTRODUCTION

The main aim of this paper is to identify the illegal pills and to prevent the life of the people. Diseases are most common nowadays and fortunately the drug has been found for almost all diseases but unfortunately some illicit Drugs, widely circulated in international market are one of the major factors influencing criminal activities which lead to additional enforcement and tracking expense for law enforcement units[4]. It is a very pathetic situation where people without knowing how to identify legal and illegal drugs leads to their life end .In other side some of the companies still involved in preparing the banned drugs from kids to adults. So, it is mandatory to have legal pill identification system in terms of size,color,shape and Imprint. The proposed system is used to avoid errors occurring in prescription and dispensing of a drug. Identification of legal and illegal pills help physicians and patients to decrease instability and to increase patient confidence in health care system.The legal pills details can be collected from the pharmaceutical companies enrolled in the database. Such database have information like chemical description and physical description like shape,color,text or Imprint.Imprint is an printed mark on pills, tablet or capsule. It may be a set of digits, text, symbol or any combination of them. The proposed system is used to identify legal and illegal pills from the features of the pill itself without any user- specified text keywords.Presently, some of the keyword based identifier tools available and provide the details like for what the medicine is used and what are the side effects occur by consuming that pill,but it doesn't gives information whether it is an legal or illegal pill.To overcome the above mentioned problem the performance of the proposed system was evaluated using several legal and illegal pill images and the obtained results is presented in this paper.

2. PROPOSED METHODOLOGY

The three common steps are feature extraction,matching, and feature retrieval is used for identifying pill images.Feature extraction is to extract image features in a various form[6]. Feature matching involves matching these features to yield a result which is similar.The methodology diagram is given below. As a first step the input image is cropped and text feature is compared with database image. In the second step the pills which is matched with the text alone undergo an comparison with the features like shape and Color. Only the pills matched with all the features will be marked as legal pills and stored in the database.



Methodology Block Diagram

The original pill image cannot be processed as such because it may be of irregular size or it may contain noise or it may be blurred or it may contain an improper background. In order to avoid all these factors preprocessing is necessary. Thus preprocessing is used to enhance the fine details of a Pill. It is carried out by two steps as Interpolation and Cropping image .Interpolation is used to enlarge the size of the image without modifying details of images which gives fine feature of pill images. The method used for Interpolation process is Bicubic Interpolation. The main aim of introducing this method is to reduce blurring, error and jagged edges during up-sampling. The edge region of image is determined using Wavelet interpolation method. Image Cropping is the process used for pick out text from the image. If we apply Geometrical gradient vector method on image before cropping ,which takes the edges for sharp part of images and not obtaining the text separately for matching. So, the cropping method is used and text matching is performed for query images.

3.FEATURE EXTRACTION

The proposed system consists of three steps for identifying drug pill images such as Preprocessing, Feature Extraction, Feature matching and Retrieval of drug pill image. Preprocessing is the process used to enhance the details of pill image for further processing.Feature Extraction gives the various characteristics of pill by transforming image into numerical form.

3.1Text feature extraction

The feature Text or Imprint be extracted using the Geometrical Gradient feature transformation algorithm In text matching the query pill image is taken as input image which is matched to database image with the help of geometrical gradient feature transform. In cropping process, text feature on image is extracted for matching each letter of text in query image with database image. The cropped image is imposed to canny edge detection. Use of canny is to take wide range of edge in image so that an outer area of text is obtained. An edge in an image may point to a variety of directions, so the canny algorithm uses four filters to detect horizontal, vertical and diagonal edges in the blurred image. The edge detection operator (Roberts, Prewitt, Sobel) returns a value for the first derivative in the horizontal direction and the vertical direction . Row and column wise scanning is performed to detect edge pixel from that each letter in the text are extracted separately. These letters are compared with database image separately. Take the center pixel for each letter and calculate Euclidean distance between center pixel and remaining pixel value. Euclidean distance is a method used to find "ordinary" distance between two points that one would measure with a ruler, and is given by the Pythagorean formula. In the Euclidean plane, if P= (p1, p2) and Q= (q1,q2) then distance is given by,

$$D = \sqrt{(p1 - q1)^2 + (p2 - q2)^2}$$

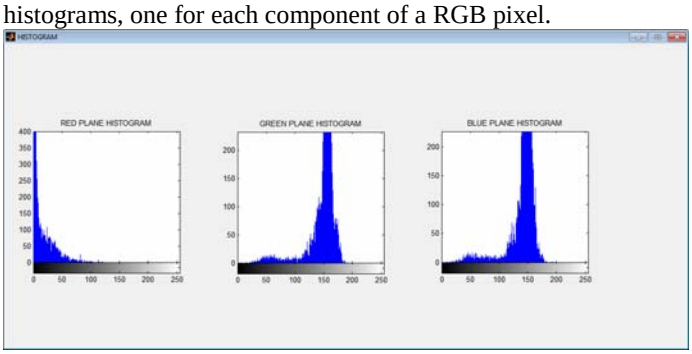
p1, p2 - Pixel coordinates in center position of image
q1, q2 – Pixel coordinates of other pixel in image

Using above formula distance are calculated which are arranged in vector form. The same process is repeated for database image. The difference between these two image distances are determined. This method is also adaptable for text in different scale and rotation. The query image which contains minimum difference with database image is taken as matched one.That minimum difference is considered as threshold value.

3.2 Color feature extraction

The image which is matched with text feature alone is compared with color feature using color histogram method. Histogram method is used for color matching of image and provides summarization of distribution of data. It is an representation of distribution of color in an image. Some range of pixel values is taken and repetition of values in ranging of those pixels is calculated. A color histogram describes the brightness distribution for any of the colors individually. This can be more helpful when trying to assess whether or not individual colors have been clipped.The histogram for Red, Green, Blue color values of pill images are calculated separately.This is done by taking grayscale value in X-axis and number of repetition of that value at image in Y-axis. These values are varying for every image[5]. It gives histogram for image that will be matched with database image alone be considered.

A more common approach to comparing the color content of a query image to that of database images is that of comparing color histograms. The methodology relies on the fact that images are generally represented as a series of pixel values, each corresponding to a visible color[10]. Color histograms are computed for each image so as to identify relative proportions of pixels within certain values.A most basic form of color retrieval involves specifying color values that can be searched for in images from a database. Computers represent all visible colors with a combination of some set of base color components, generally Red, Green and Blue (RGB)[5]. The image retrieval utilized during their experimentation computes similarity based on the similarity of three different histograms, one for each component of a RGB pixel.



3.3 Shape feature extraction

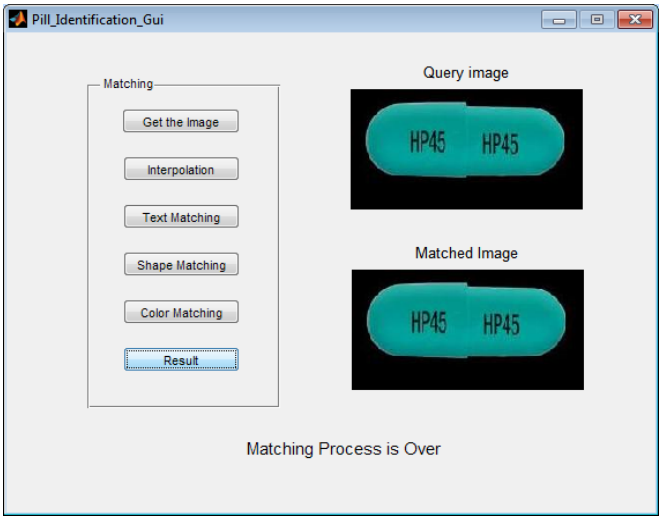
Here, the images which are matched with the features like text,color be compared with shape feature. Geometrical gradient vector method is used for the extraction of shape. In this feature extraction first background is to be removed from pill image and perform shape matching with database image.

This is done by taking 3x3 windows at each corner of the image. The median values are calculated for the same which is taken as threshold level. Thresholding is non-linear operation that converts gray-scale value into binary image where two levels are assigned to pixels that are below or above threshold value[11]. It is called as tolerant value. The tolerant values are reduced to 0. It makes image as clearly visible from background. Thus the image is separated from background. Gradient operator is applied to that image which takes only the high frequency part of image and we get boundary of pill image.Geometrical gradient transform is taken into the boundary. It trace the edge pixel on boundary having minimum values. These values give center position of pill image. Euclidean distances are calculated same as mentioned in text matching.

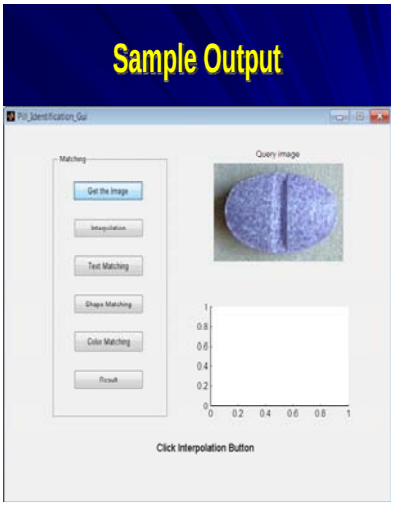
3.4 Feature Matching

Matching process is performed by comparing the distance of both database and query image. During shape matching, the image must be invariant for scale and rotation. So the following processes are done for accurate shape matching.The image of pixel having minimum distance is assigned to one variable which is repeated for every pixel in an image. It gives the rotation of variables in query image. This method gives the query image's pixel position as same as database image pixel position. The reason for doing this process is, if query image is same as database image but the direction of query image is different which leads to mismatch of query image. So such problems can be avoided by this method of shape matching.L1 normalization method is used to match all the three features of the given query image with the features of database image.

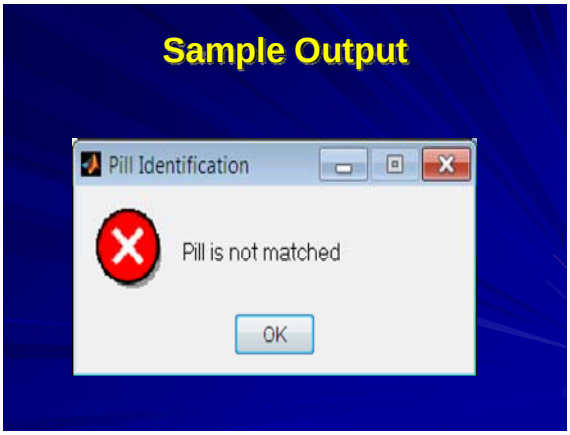
4.EXPERIMENTAL RESULTS



Fig(1): Output Result for the Matched Pill

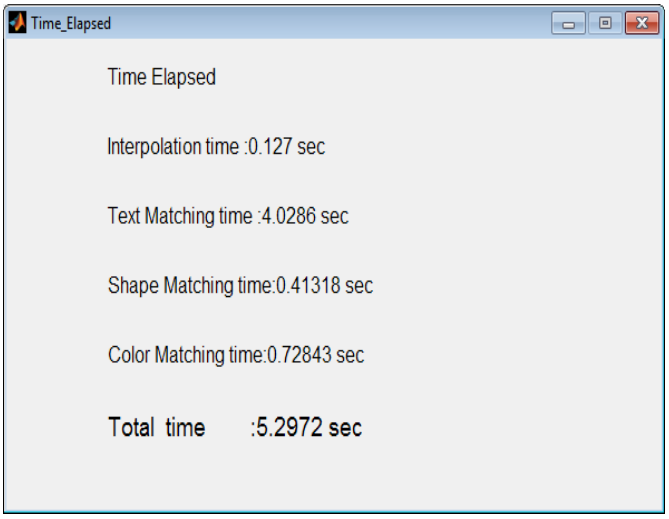


Fig(2):Sample Pill Image



Fig(3): output result for the mismatched pill

The Interpolated pill is applied for text matching and also compared to obtain the matching with color and shape as shown above.If the given query pill is matched with all the three text,color,shape features then the result shows as Legal Pill. If the pill is mismatched with any one of the three features it shows the result as Pill in not matched due to mismatch occurred either in text,color,shape which is declared as illegal.The average accuracy of the proposed system after matching with text,color,shape feature is 86.9%.Hence it needs a total elapsed time of 5.29 Sec for the entire process and elapsed time for each individual feature is also shown .Thus the results from various experiment shows that algorithm used in the proposed system is efficient when compared to other algorithm.



Fig(4) : Time duration of each feature extraction and entire process

To assess the result of images proposed in this work the evaluation parameters such as Sensitivity,Specificity and Accuracy been calculated.These parameters are determined for different criterian parameters such as True positive(TP),True negative(TN),False positive(FP),False negative(FN) as follows.

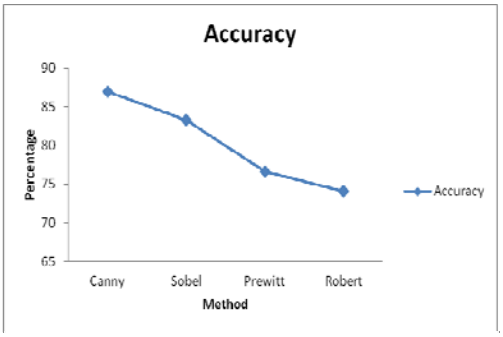
Accuracy is calculated in the proposed system using F-measure technique as follows and the formula is.

Accuracy=((TP+TN)/(TP+TN+FP+FN))*100

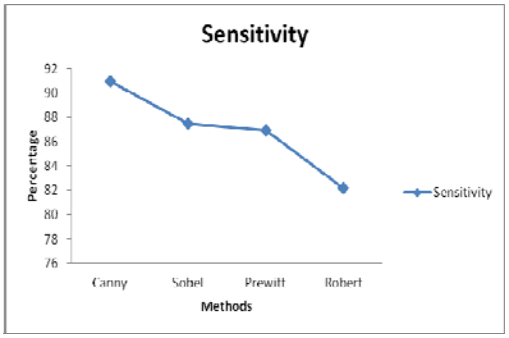
TP=True Positive, TN=True Negative.
FP=False Positive, FN=False Negative.
Accuracy Percentage = ((Matched Pill)/ (Matched Pill +Miss Matched Pill))*100

Sensitivity = True Positive / True Positive + False Negative

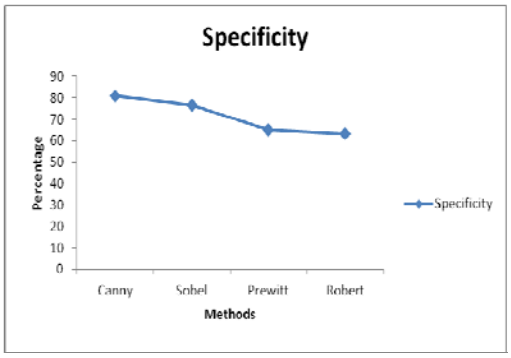
Specificity = True Negative/ True Negative + False Positive



Fig(5) Comparative study on Accuracy diagram



Fig(6) Comparative study on Sensitivity diagram



Fig(7) Comparative study on Specificity diagram

Table 1: Comparitive analysis of algorithm

Method	True Positive	True Negative	False Positive	False Negative	Accuracy	Sensitivity	Specificity
Canny	672	391	93	67	86.91742	90.93369418	80.78512397
Sobel	655	364	110	94	83.31971	87.44993324	76.79324895
Prewitt	564	373	201	85	76.61488	86.90292758	64.9825784
Robert	572	334	193	124	74.08013	82.18390805	63.37760911

5. CONCLUSION

The techinques used in the proposed System to identify whether a given pill is an legal or illegal by matching wih three different features as Text, color,Shape is proved to be efficient when comparing with existing system.However furthur improvements can be made by adding too many images in the database.This research paper also reduces the artifacts,blurring, and jagged edges.The identification of legal and illegal pills is a challenge because even the same tablet vary with variation in illumination,Imprint size and view point.Inclusion of Rotation normalization features improve pill identification system more and also more work is furthur needed to examine the failure cases which are planned in future.

REFERENCES

1. Enhanced Image Mining Techniques for Drug Pill Image, Mrs. A.Hema, Dr.(Mrs)E.Annasaro, International Journal of Computer Trends and Technology- volume4Issue2- 2013.
2. *Automatic Text Extraction in Video Based on the Combined Corner Metric and Laplacian Filtering Technique Kaushik K.S1, Suresha D2, ISSN: 2278 – 1323 International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 2, Issue 6, June 2013.*
3. A survey in need of image mining techniques,A.Hema1, E.Annasaro, International Journal of Advanced Research in Computer and Communication EngineeringVol. 2, Issue 2, February 2013.
4. *.Lee, Y.B., Park, U., Jain, A.K. and Lee, S.W. (2012)Pill-ID: Matching and retrieval of drug pill images,Pattern Recognition Lett,, Volume 33, Issue 7, Pp. 904–910.*
5. A Study of Color Histogram Based Image Retrieval, Rishav Chakravarti, Xiannong Meng, 2009 Sixth International Conference on Information Technology: New Generations.
6. *CBIR USING COLOR HISTOGRAM PROCESSING,P.S.Suhasini , DR. K.sri ama krishna,DR. I. V. murali krishna, Journal of Theoretical and Applied Information Technology © 2005 - 2009 JATIT. All rights reserved.Vol6. No1.(pp 116 - 122).*
7. Geradts, Z. and Bijhold, J. (2002) Content based information retrieval in forensic image databases, J.Forensic Sci., Vol. 47, No.2, Pp.285–292.
8. *Hu, M.K. (1962) Visual pattern recognition by moment invariants, IEEE Trans. Inform. Theor., Vol. 8, No.2,Pp.179–187.*
9. .Fuzzy C-Means Clustering Atgorithm for euality Inspection of Fruits based on fmage Sensors Data, Ebrahim.Aghajari, D.C. Gharpure
10. *SEGMENTATION AND HISTOGRAM GENERATION USING THE HSV COLOR SPACE FOR IMAGE RETRIEVALShamik Sural, Gang Qian and Sakti Pramanik..*
11. Extended Abstract: Image Matching for Branding Phishing Kit Images,Chengcuiz zhang, rajan kumar kharel, Jason Britt, Song Gao.
12. http://en.wikipedia.org/wiki/Health_care_industry.
13. *http://www.pharmer.org/images.*
14. http://en.wikipedia.org/wiki/Health_care_industry.