Recognition of Similar Objects Using a Hybrid Classifier

S. M. Mirsadri, H. Bolandi, F.Saberi
Dept. of Electrical Engineering, the Iran University of Science and Technology
smohsen_mirsadri@ee.iust.ac.ir

Abstract

A novel method for classification of objects based on a hybrid of decision theoretic and structural methods is presented in this paper. The circumstances of scaling and presence of noise are included as the part of the study. Images are degraded by some known types of noise like Gaussian, Salt & Pepper and Speckle and iterative filtering algorithm based on classification results and using alpha-trimmed mean filter will be used. Fuzzy clustering algorithm is used for thresholding and background removal in cluttered images and spurious parts are reduced using morphological operations. Input database is made up of images having similar shapes lied on their most usual appearance. Feature vectors are composed of Moment invariant and Interior angles of polygons and would be extracted after normalizing the object boundary with respect to size and orientation. Interior angles are extracted from a shape, described using a new polygonal approximation technique. Similarity measurement is done by combining two classifiers, Euclidean distances in Decision theoretic and String matching in Structural methods. In order to investigate the reliability of presented method in presence of noise, the classification results obtained from a hybrid method are compared with those of the Decision theoretic or Structural methods.

1. Introduction

Recognizing objects on the basis of their visual appearance is one of the major goals in computer vision. This task has shown to be challenging mainly because of the large variability in objects appearance. Objects may appear in cluttered backgrounds, be in variant distances from the camera (different size and orientation) and become degraded during acquisition process.

Object recognition is an essential part of any high-level image analysis system and finds objects in the real world from an image of the world, using object models which are known a priori. It is now widely used in many applications such as computer vision, military, biology, psychology, medicine, etc. The task of recognition can be broken into three distinct phases: Image acquisition and preprocessing, feature extraction and classification.

First stage operates on the original image and extracts regions containing possible targets, also performs thresholding, noise reduction and background removal within region of interest.

If an image is being sent electronically from one place to another, via satellite or wireless transmission, or through network cable, errors are expected to occur in the image signal. Such degradations may include noise which is defined to be any degradation in the image signal, caused by external disturbance. Gaussian noise is an idealized form of white noise, which is caused by random fluctuations in the signal (It can be observed by watching a television which is slightly mistuned to a particular channel). Salt and pepper (Impulse) noise can be caused by sharp, sudden disturbances in the image signal and its appearance is randomly scattered white or black (or both) pixels over the image. Noise reduction can be done using spatial or frequency domain filters [1]. Thresholding helps to specify objects by their intensity values, and then background removal can be done using objects sizes and their connectivity.

Feature extraction stage selects the features from each ROI for input to the classification stage. A feature is some attribute of the object that is considered important in describing and recognizing the object in relation to other objects. The feature detector applies operators to images and identifies locations of features that help in forming a hypothesis of the object.

Classification stage decides on the type of input and will consist in the comparison of image data with the model database. The model database contains all the models known to the system. The information in the model database depends on the approach used for the recognition. The recognition is based on the computation of the similarity between image and model. The four best known approaches for pattern recognition are: 1) template matching, 2) decision theoretic, 3) syntactic or structural matching, and 4) neural networks [2]. These techniques are not necessarily independent and sometimes a pattern

0-7695-3039-7/07 $25.00 © 2007 IEEE
DOI 10.1109/IMSCCS.2007.84

recognition technique exists with different interpretations. Attempts have been made to design hybrid systems involving multiple techniques [5].

One of the simplest and earliest approaches to pattern recognition is based on template matching. Matching is a generic operation in pattern recognition which is used to determine the similarity between two entities (points, curves, or shapes) of the same type. A pattern to be recognized is matched (often by correlation) against the stored template [3].

For template matching there are two types of templates [2]: global, and local. A global template represents the complete object under investigation whereas a local template utilizes several local features of the object. Brunelli and Poggio [7] reviewed synthetic discriminant functions based on least square estimation. Remagnino et. al. [8] studied a technique which allows the identification of similar image patches from two consecutive views. The template matching, while effective in some application domains, has a number of disadvantages [3]. For instance, it would fail if the patterns are distorted due to the imaging process, view-point change, or large intra-class variations among the patterns. In general, template matching is only appropriate in simple pattern matching problems. In the decision theoretic or statistical approach, each pattern is represented in terms of d features or measurements and is viewed as a point in a d-dimensional space. The goal is to choose those features that allow pattern vectors belonging to different categories to occupy compact and disjoint regions in a d-dimensional feature space [3]. The statistical approach uses global shape features, such as moments, autoregressive coefficients and Fourier descriptors to recognize shapes [7].

Structural recognition deals with symbolic information properties and relationships. It is based primarily in representing objects such as strings, trees, or graphs and the recognition rules are then based on those representations [4].

Neural network models attempt to use some organizational principles (such as learning, generalization, adaptivity, fault tolerance and distributed representation, and computation) in a network of weighted directed graphs in which the nodes are artificial neurons and directed edges (with weights) are connections between neuron outputs and neuron inputs [3].

The organization of the paper is as follows: section 2 discusses feature extraction and some image preprocessing operations including noise reduction, thresholding, background removal, boundary extraction and normalization. In Section 3, classifier basis and utilized structures are described. Section 4 gives conclusions.

2. Preprocessing and feature extraction

A feature is some attribute of the object that is considered important in describing and recognizing the object in relation to other objects. The feature extraction aspect of image analysis seeks to identify inherent characteristics of objects found within an image. Feature extraction is applied on image arrays to produce a list of descriptors or feature vectors. Prior to extract features, some preprocessing operations like removing background of the image, noise reduction and pattern normalization which will contribute in defining a compact representation of the patterns are done.

2.1. Preprocessing

A brief description of the image preprocessing steps and changes in the images that take place after these steps are as follows:

2.1.1. RGB to grayscale conversion. RGB (red-green-blue) color images are converted to gray-scale images.

2.1.2. Size Correction. The size of images is reduced to correct dimension 150 × 200 pixels, in order to perform preprocessing operations on source and target images having the same size. Scaling is done using spatial transformation, which can be expressed as [1]:

$$(x, y) = T\{(\omega, z)\}$$

$$\begin{cases} x = S_x \omega \\ y = S_y z \end{cases} \qquad S_x, S_y = Scaling \quad Factors \tag{2- 1}$$

S_x : Ratio of 200 over width of an image
S_y : Ratio of 150 over height of an image

2.1.3. Image restoration. If an image is being sent electronically from one place to another, via satellite or wireless transmission, or through network cable, errors are expected to occur in the image signal. These errors will appear on the image output in different ways depending on the type of disturbance in the signal. Image restoration concerns the removal or reduction of degradations which have occurred during the acquisition of the image. Such degradations may include noise which is defined to be any degradation in the image signal, caused by external disturbance. In this paper images are degraded by Gaussian, Salt & Pepper and Speckle noises and iterative filtering algorithm based on classification results and using alpha-trimmed mean filter will be used.

The alpha-trimmed mean filter combines order-statistics and a mean filter and replaces each pixel with the mean value of its neighbors in S_{xy} *trimmed* of a given fraction, α of its lowest and highest values in S_{xy} and is given by:

$$f(x,y)=\frac{1}{mn-2\alpha(mn-1)}\sum_{(s,t)\in S_{xy}^{tr}} g(s,t). \quad (2\text{-}2)$$

It becomes necessary to apply a filter multiple times in some cases and a number of iterations is obtained from classification results.

2.1.4. FCM thresholding. Thresholding process is applied and the gray level of the pixels between 0–255 is changed to either 0 or 255. Level is a normalized intensity value that lies in the range [0 1]. Threshold value is determined by a Fuzzy c-means (FCM) method which is an iterative data clustering technique wherein each data point belongs to a cluster to some degree that is specified by a membership grade. This algorithm starts with an initial guess for the cluster centers, which are intended to mark the mean location of each cluster. The initial guess for these cluster centers is most likely incorrect. Additionally, fcm assigns every data point a membership grade for each cluster. By iteratively updating the cluster centers and the membership grades for each data point, fcm iteratively moves the cluster centers to the right location within a data set. This iteration is based on minimizing an objective function that represents the distance from any given data point to a cluster center weighted by that data point's membership grade.

2.1.5. Regional minima. Regional minima are connected components of pixels with a constant intensity value, and whose external boundary pixels all have a higher value. Regional minima [1] of a binary image are computed and the value of an image becomes 1 corresponding to the pixels that belong to regional minima and 0 otherwise.

2.1.6. Flood-fill operation. Flood-fill operation on background pixels of an image is performed and the image regions and holes are filled [1]. It is done through changing connected background pixels (0's) to foreground pixels (1's) and stopping when it reaches object boundaries.

2.1.7. Spur Reduction. Spur Reduction [1] is used for removing spurious parts of the image. An approach is based on combining two morphological operations, dilation and erosion. Dilation is an operation that “grows” or “thickens” objects in a binary image and erosion “shrinks” or “thins” objects in a binary image.

2.1.8. Boundary extraction. Boundary extraction is used to find points on the image surface such that any neighborhood of those contains some points in the surface and some not in the surface. Boundary of the object with longest boundary is selected.

2.1.9. Size normalization. Size normalization is done using spatial transformation [1]. Scaling factors (S_x, S_y) are defined as the ratio of Major and Minor axis length of a source image boundary over those of target image respectively. Major (Minor) axis is the line segment connecting two farthest (nearest) points on the boundary and is one of the measurable properties of a region or boundary called Region descriptors.

2.1.9. Orientation normalization. Extracted boundary of the image is normalized with respect to orientation and that is done by computing the angle between major axis of image boundary and x-axis using regional descriptors, then aligning the image boundary with respect to it. Alignment is done using spatial transformation, which can be expressed as [1]:

$$(x,y)=T\{(\omega,z)\}\ , \quad \begin{cases} x=\omega\cos\theta - z\sin\theta \\ y=\omega\sin\theta - z\cos\theta \end{cases} \quad (2\text{- }3)$$

2.2. Feature extraction

Three types of features are used in this paper: Moment Invariants [11], [12] , freeman chain codes and interior angles. Extracted feature vector should be invariant to position, orientation, and scale aspects of the object.

2.2.1. Moment invariants. A set of invariant moments, $\{\phi\}$, may then be defined as follows [1]:

$$\begin{aligned}
&\phi_1=\eta_{20}+\eta_{02}\\
&\phi_2=(\eta_{20}-\eta_{02})^2+4\eta_{11}{}^2\\
&\phi_3=(\eta_{30}-3\eta_{21})^2+(3\eta_{21}\ -\eta_{03})^2\\
&\phi_4=(\eta_{30}+\eta_{12})^2+(\eta_{21}\ +\eta_{03})^2\\
&\phi_5=(\eta_{30}-3\eta_{12})(\eta_{30}+\eta_{12})[\ (\eta_{30}+\eta_{12})^2\\
&-3(\eta_{21}\ +\eta_{03})^2\ ]+(3\eta_{21}-\eta_{03})(\eta_{21}+\eta_{03})\\
&\times[\ 3(\eta_{30}+\eta_{12})^2-(\eta_{21}+\eta_{03})^2\ ]\\
&\phi_6=(\eta_{20}-\eta_{02})[\ (\eta_{30}+\eta_{12})^2-(\eta_{21}+\eta_{03})^2\ ]\\
&+4\eta_{11}(\eta_{30}+\eta_{12})(\eta_{21}+\eta_{03})\\
&\phi_7=(3\eta_{21}-\eta_{03})(\eta_{30}+\eta_{12})[\ (\eta_{30}+\eta_{12})^2\\
&-3(\eta_{21}\ +\eta_{03})^2\ ]+(3\eta_{21}-\eta_{03})(\eta_{21}+\eta_{03})\\
&\times[\ 3(\eta_{30}+\eta_{12})^2-(\eta_{21}+\eta_{03})^2\ ]
\end{aligned} \quad (2\text{-}4)$$

This set of invariant moments makes a useful feature vector for the recognition of objects which must be detected regardless of position, size or orientation.

This feature vector is represented as $f_{ij},\ j=1,2,...,7$ where i specifies a class and j indicates moment invariants.

2.2.2. Interior Angles. Interior angles are used to represent a boundary by a sequence measured angles of a boundary corners. Typically, this representation is based on 4- or 8- connectivity of the segments [1]. Obtaining the Interior angles of a boundary would result in a long sequence with small variations that are not necessarily representative of the general shape of the image. Thus the boundary is subsampled with a grid separation equal to approximately 10% the width of the image. This feature vector is represented for class i as S_{ij} , $j=1,...,N$ and because of the normalization operations which were done through the preprocessing stage, is invariant to variation in size and orientation.

These feature vectors are computed for source and target images and will be used in the classification stage. “Figures 2.1.” through “2.3.”, shows the process of extracting features for source and target images.

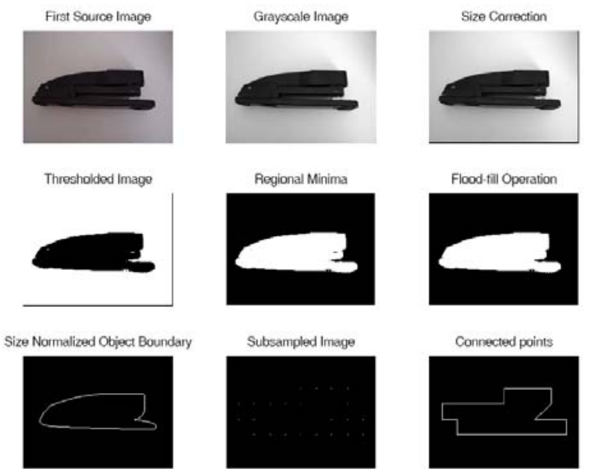

Figure 2.1. Preprocessing and feature extraction of the first model image. Thresholding level of the image (0.305) was obtained after 19 iterations of FCM algorithm.

Computed features are as follows:

$S_{1j} =$

[333333333333333333333333333333333333333333233333333333333333333322
2222222222222222224333333333333333333333333333333333333333333323333
333333333333333323333333333333333333333333333333333333333333333333
333333333333333333333333333333333333333333333333333333333333333333
333333333333333333333333333333323333333333333333333343333333333333
333333323333333333333333333323333333333333333333333333333333333333
333333333333333333333333333333333333333333333343333333333333333333
3233333333333333333333]

Above code can be interpreted as: 1, 2, 3, 4 represent 0°, 90°, 180°, 270° respectively.

$f_{1j} =$

[1.9667 3.2741 3.5378 2.7192 5.8428 4.3043 4.7260]

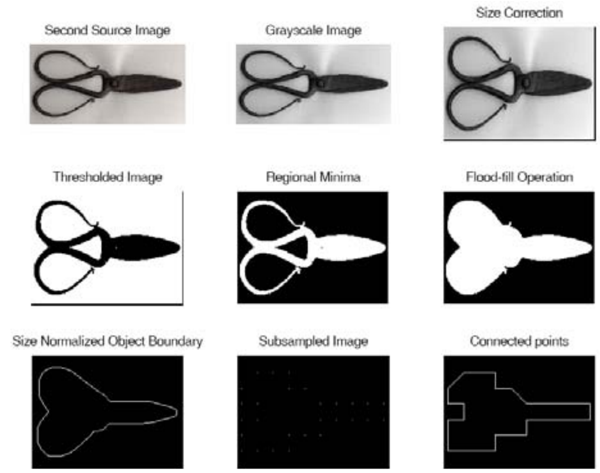

Figure 2.2. Preprocessing and feature extraction of the second source image. Thresholding level = 0.388, No. of Iterations = 20

Computed features are as follows:

$S_{2j} =$

[33333333333333333333323333333333333333333343333333333333333333322
222222222222322223223333333333333333333333333333333333333333333333
333333333333333333333333333333333333323333333333333333333323333333
333333333333333333333333333333333333333333333333333333333333333333
333333333343333333333333333333332333333333333333333333333333333333
333333343333333333333333333323333333333333333333333333333333333333
333333333333333333333333323333333333333333333333333333333333333333
323333333333333333333343333333333333333333343333333333333333333323
3333333333333333333222222222222222222222233333333333333333333]

$f_{2j} =$

[2.3014 3.3330 5.7366 0.4836 2.0895 1.1597 0.9984]

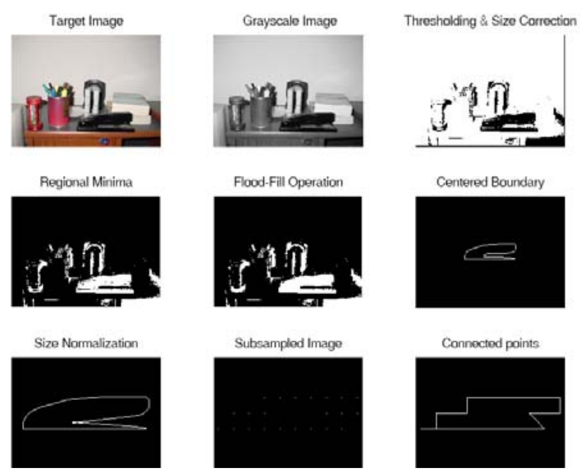

Figure 2.3. Preprocessing and feature extraction of the cluttered target image. Thresholding level = 0.407, No. of Iterations = 22

Computed features are as follows:

k_{1j} =

[333333333333333333333333333333333333333333333323333333333333333333323
33333333333333333333333333333333333333334222222232323232322221333
33333333333333333333333333333333333333333333333333333333333333333333
33333333333333333333333333333333333333333333333333333333333333333333
33333333333333333333333333331333333333333333333343333333333333
33333332333333333333333333333333333333333333334333333333333333
33332333333333333333333333333333333333333333333333333333333333333333
333333333333333333333]

k_{2j} =

[1.5208 2.5869 2.4265 2.1112 4.3450 3.3912 4.3669]

3. Classification

Classification stage decides on the type of input image and will consist in the comparison of image data with the model. The degree of difficulty of classification problem depends on the variability in feature values for object in the same category, relative to the difference between feature values for objects in different categories. Classification is based on the computation of the similarity between image and model.

3.1. Decision theoretic method

In the decision theoretic or statistical approach, similarity measurement is done by computing distance between the feature vectors of unknown object and that of a model. Model object (ideal feature values) for each class which is known is represented as f_{ij} and also measured features of the unknown object are represented as k_{2j}. To decide the class of the object, its distance with each class is measured and the object would be assigned to the nearest class. Measuring distance is done by computing the Euclidean distance between f_{ij} and k_{2j} :

$$d_i = \left[\sum_{j=1}^{7} (k_{2j} - f_{ij})^2 \right]^{-1/2} \tag{3-1}$$

Then the object is assigned to the class i of n such that:

$$d_R = \min_{i=1}^{n} \left[\frac{d_i}{7} \right]$$

$$\left(\frac{d_i}{7}: \text{ Mean of 7 difference values}\right) \tag{3-2}$$

3.2. Structural method

Structural recognition deals with symbolic information properties and relationships. It is based primarily in representing objects such as strings, trees, or graphs and the recognition rules are then based on those representations. By using the string matching, measurement of similarity can be viewed the same as measuring distances and positions like in comparing two region boundaries [4]. For doing this, extracted interior angles of boundaries related to source and target images, which are normalized with respect to size and orientation, are converted to strings and the similarity is measured performing the match between corresponding symbols. Let λ denote the number of matches between these two strings, where a match is said to occur in the zth position if

$$f_{ij_z} = k_{1j_z}$$

The measure of similarity is the ratio [1]:

$$S = \frac{\lambda}{\max(|f_{ij}|, |k_{1j}|) - \lambda} \tag{3- 3}$$

where $|\arg|$ is the number of symbols of the strings in the argument. Then the object is assigned to the class i such that:

$$d_S = \max_{i=1}^{n} [S] \tag{3- 4}$$

n : Number of image classes

"Figures 3.1." through "3.3.", shows the process of extracting features for noisy target images degraded by Gaussian, Salt & Pepper and Speckle noises.

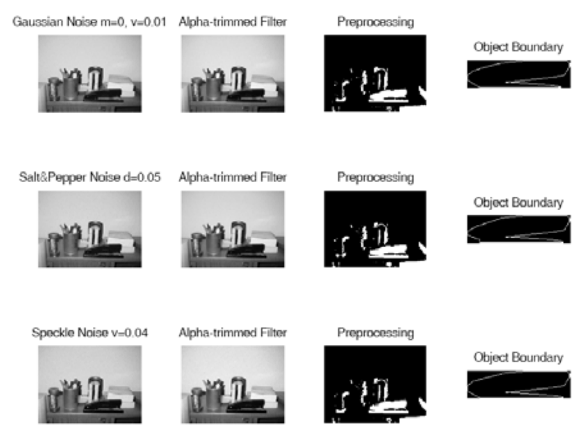

Figure 3.1. Preprocessing and feature extraction of the cluttered noisy target image. (Top) Image degraded by gaussian noise of mean = 0 and var = 0.01. (Middle) Salt&Pepper noise with density = 0.05. (Bottom) Speckle noise of var= 0.04.

For classification of objects degraded by low noises, application of one alpha-trimmed filter is sufficient and recognition can be done using either Decision theoretic or Structural method as shown in "Table 1."

But when images are degraded by noises with high amplitudes, the classification results, as shown in “Table 2.” doesn’t match together. In this case the filter is applied multiple times, until the results of both methods assign the object to the same class.

TABLE 1
Classification results of noisy images

Noise	-	Gaussian	Salt & Pepper	Speckle
Mean	-	0	-	0
Variance	-	0.01	-	0.04
Density	-	-	0.05	-
Distance with class 1	0.3345	0.4754	0.5025	0.4564
Similarity with class 1	9.7333	2.8333	2.8333	2.8333
Distance with class 2	0.8594	0.7955	0.7795	0.8042
Similarity with class 2	2.8431	1.5128	1.5128	1.5128
Assigned Class	1	1	1	1

“Figure 3.2.”, shows the extracted boundary of images which are degraded by high amplitude noises and have been filtered one time.

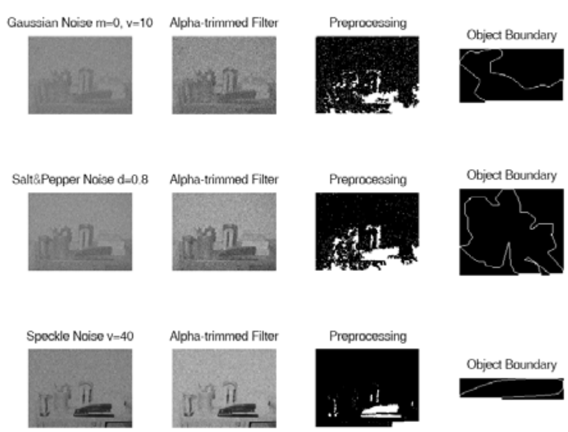

Figure 3.2. Preprocessing and feature extraction of the cluttered noisy target image. (Top) Image degraded by gaussian noise of mean = 0 and var = 10. (Middle) Salt&Pepper noise with density = 0.8. (Bottom) Speckle noise of var= 40.

“Figure 3.3.”, shows the same process, but applying a filter 3, 4 and 1 times to images degraded by Gaussian, Salt & Pepper and Speckle noises respectively.

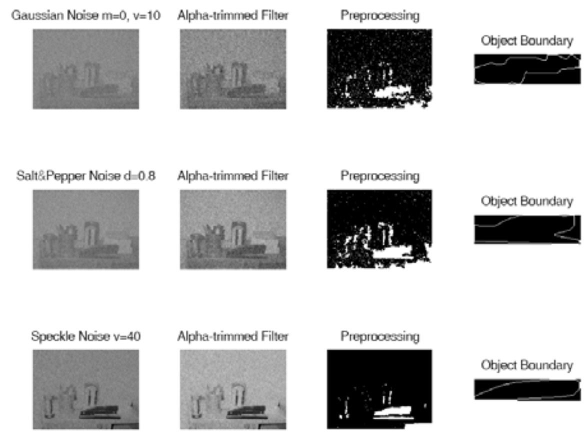

Figure 3.3. Preprocessing and feature extraction of the cluttered noisy target image. (Top) Image degraded by gaussian noise of mean = 0 and var = 10 and filtered three times. (Middle) Salt&Pepper noise with density = 0.8 and filtered four times. (Bottom) Speckle noise of var= 40 and filtered once.

TABLE 2
Classification results of images degraded by high noises

Noise	Gaussian		Salt & Pepper		Speckle	
Mean	0	0	-	-	0	0
Variance	10	10	-	-	40	40
Density	-	-	0.8	0.8	-	-
Filter iterations	1	3	1	4	1	1
Distance with class 1	0.5751	0.7275	0.8067	0.1401	0.5413	0.4898
Similarity with class 1	4.4124	2.6316	1.9538	8.6600	2.5778	2.5778
Distance with class 2	0.6345	1.1841	0.7630	1.0948	0.8902	1.1456
Similarity with class 2	4.8690	1.4810	1.8421	2.7692	1.5021	1.1502 1
Assigned Class	-	1	-	1	1	1

4. Conclusions

Object recognition in presence of the noise, clutter background and variations in size has been reported in this paper. Recognizing objects on the basis of their visual appearance is one of the major goals in computer vision. A key feature of any object recognition algorithm aiming to perform well in realistic scenarios is robustness. The hybrid classifier represented above, is robust to change in size, cluttered background and degradation in many cases.

Acknowledgement

Supported by Electrical Power Industries Laboratories Company (EEPIL).

References:

[1] Rafael C. Gonzalez, Richard E. Woods and Steven L. Eddins, Digital Image Processing Using Matlab, Pearson Education, Inc., 2004.

[2] S. F. Bahgat, A hybrid moment method for identification of 3D-objects from 2D-monocromatic images, IJICIS, VOL. 4, NO. 1, January 2004.

[3] Anil K. Jain, Robert P.W. Duin and Jianchang Mao, Statistical pattern recognition: A review, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, No. 1, 2000.

[4] Lourdes A. Funtanilla, GIS pattern recognition and rejection analysis using MATLAB, Texas A&M University, 2004.

[5] Liang-Kai Huang and Mao-Jiun J.WANG, Efficient shape matching through model-based shape recognition, Elsevier Science Ltd., Pattern Recognition, pp. 207-215, Vol. 29, No. 2, 1996.

[6] K. S. Fu, A step toward unification of syntactic and statistical pattern recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 200-205, Vol. 5, No. 2, March 1983.

[7] R. Brunelli and T. Poggio, Template matching: matched spatial filters and beyond, Massachusetts Institute of Technology, A. I. Memo No. 1549, 1995.

[8] L. Devroye, L. Gyorfi and G. Lugosi, A Probabilistic Theory of Pattern Recognition, Springer Verlag, Berlin,1996