**EECS**

**University of Tennessee**

**Pattern Recognition – ECE 571**

**Final Project – Vehicle Recognition Using Vehicle Silhouettes**

**Submitted by:**

Yazeed Alkhrijah                         Mahendra Duwal Shrestha

yalkhrij@vols.utk.edu         mduwalsh@vols.utk.edu

**Abstract**

This project uses pattern recognition procedures for vehicle recognition using vehicle silhouettes. The objective is to integrate different classification techniques to get best accuracy of classification.

Specific properties of shape features measures of 2D image can be studied to gain knowledge about which vehicle the image belongs to. The dataset was collected by JP Siebert at the Turing Institute, Glasgow, Scotland in 1986-87. The original purpose was to find a method of identifying 3D objects within a 2D image by application of combination of shape feature extractors to the 2D silhouettes of the objects.

Four types of vehicle were chosen for this project and images of painted black vehicles were captured with different angles, and used for feature extraction. The dataset was already provided for our project. The given dataset was then normalized and introduced to dimensionality reduction technique. The effects of dimensionality reduction and not using dimensionality reduction technique were measured using different types of classifiers on those datasets. Both supervised and unsupervised learning were used for classification purpose. Since just a single dataset was provided 3 fold cross validation evaluation technique was implemented for all classifiers using supervised learning. Best three high performance giving classifiers were selected for classifier fusion.

Three types of classifier fusion methods were implemented which were plurality, Naive Bayes and Behavior Knowledge Space. After implementing all three fusions, classification accuracy were measured for each fused classifier.

**Introduction**

The dataset we've chosen for this the project is vehicle silhouettes. Vehicle recognition has been used for decades now and been used in various fields in real world. Since shape and design of each brand of vehicle are different than others, silhouettes formed by different types of vehicle in 2D image have different properties depending upon the type of vehicle. Specific properties of shape features measures of the image can be studied to gain knowledge about which vehicle the image belongs to. So identifying vehicle type which is 3D based on the shape features extracted from 2D image with all possible angles forms a very good pattern classification problem.

Vehicle recognition has been used in different areas. It can be used for solving criminal cases. It can be used as tool for data mining or statistics of vehicle type in specific location. This can also be used in advertising market.

**Data set**

The data was originally collected by JP Siebert at the Turing Institute, Glasgow, Scotland in 1986-87. The original purpose was to find a method of identifying 3D objects within a 2D image by application of combination of shape feature extractors to the 2D silhouettes of the objects.

For the experiment, four "Corgie" model vehicles were chosen: a double decker bus, Chevrolet van, Saab 9000 and an Opel Manta 400, with the expectation that the bus, van and either one of the cars would be readily identifiable but would be more difficult to distinguish between the cars. The vehicles were painted black to minimize highlights and the images were captured by a camera looking downwards at model vehicle from fixed angle of elevation. The images were then threshold to produce binary vehicle silhouettes. The images of a vehicle were captured with different orientations of vehicle facing head on, rear on and rotating.

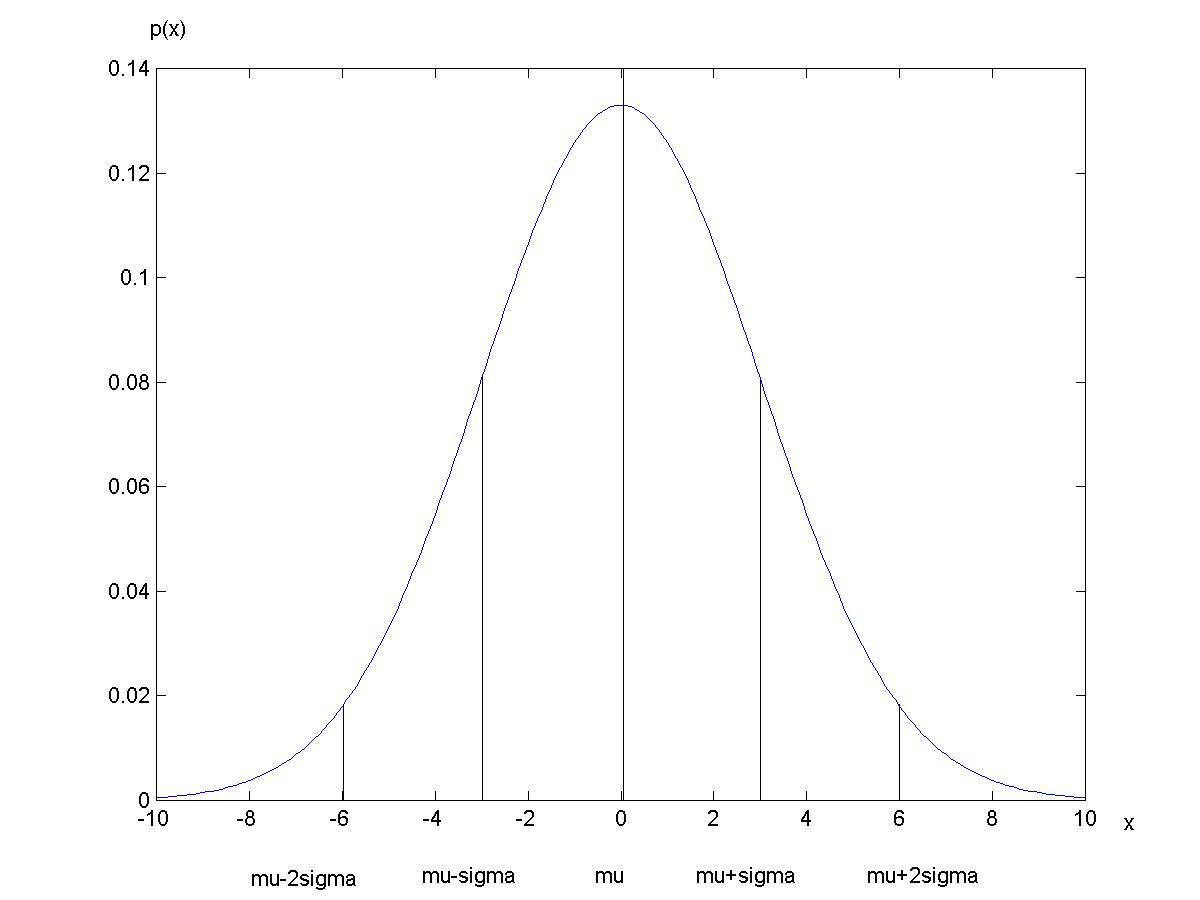
The data set consist of 18 different features for each sample extracted by the HIPS (Hierarchical Image Processing System) extension BINATTS from the silhouettes. This extracts a combination of scale independent features utilizing classical moments based measures and heuristic measures.

|  |  |
| --- | --- |
| Feature | How it is extracted |
| Compactness | (average perim)^2/area |
| Circularity | (average radius)^2/area |
| Distance circularity | area/(av.distance from border)^2 |
| Radius ratio | (max.radius-min.radius)/average radius |
| Pr.axis aspect ratio | (minor axis)/(major axis) |
| Max.length aspect ratio | (length perp. max length)/(max length) |
| Scatter ratio | (inertia about minor axis)/(inertia about major axis) |
| Elongatedness | area/(shrink width)^2 |
| Pr.axis rectangularity | area/(primary axis length\*primary axis width) |
| Max.length rectangularity | area/(max.length\*length) |
| Scaled variance along major axis | (2nd order moment about minor axis)/area |
| Scaled variance along minor axis | (2nd order moment about major axis)/area |
| Scaled radius of gyration | (major var+minorvar)/area |
| Skewness about major axis | (3rd order moment about major axis)/sigma\_min^3 |
| Skewness about minor axis | (3rd order moment about minor axis)/sigma\_maj^3 |
| Kurtosis about major axis | 4th order moment about major axis)/sigma\_min^4 |
| Kurtosis about minor axis | 4th order moment about minor axis)/sigma\_min^4 |
| Hollows ratio | (area of hollows)/(area of bounding polygon) |

Table1: explanation for feature extraction

**Technical Approach**

The dataset chosen had 18 features. The data was normalized before using. Though it’s not quite large number of dimension, the dataset was introduced to dimension reduction techniques FLD and PCA which reduced number of dimensions to 3 and 6 respectively. Since only one dataset was provided, for cross validation purpose the dataset was divided into three folds and different classification techniques were used on three folds datasets following supervised learning while unsupervised learning used the dataset as a whole. The best performing three classifiers were then chosen to integrate and generate fused result lookup table. Three different techniques of fusion were implemented to build fused classifier. The dataset was classified using the fused lookup table and accuracy of classification was measured.

**Gaussian Density Distribution:**

Multivariate Gaussian density is given by:



When d=1,

**Maximum Posterior Probability (MPP):**

Bayesian decision formula for MPP is given by:



Where P(ωj|x) is posterior probability; p(x| ωj) is conditional probability density function or likelihood; p(x) is normalization constant and is given by:



P(ωj) is prior probability.

Discriminant function for multivariate Gaussian density can be represented as:



**Case I:**

If features of sample data are statistically independent and have same variance, then we’ll have



where

The Euclidean distance is also called minimum distance because it is the distance of x from each of the c mean vectors. And when  is same for all classes the discriminant function is measuring minimum distance.



**Case II:**

If covariance matrices of all classes are same and not equal to scalar multiplication of identity matrix, then and discriminant function will be:



**Case III:**

If covariance matrices of all classes are arbitrary, then discrimanant function will be:



**Fisher’s Liner Discriminant (FLD) method**

It is one of commonly used dimension reduction methods which tries to discriminate transformed data best possible way. This method tries to project mean vector of each class as further as possible and minimize scatterings of sample data within each class as best as possible. Since this uses class information of training samples, it is supervised learning. Using this method, number of dimensions d can be reduced to c-1 where c is number of classes and is supposed less than d. If we consider two class case, then data of d dimension is projected onto a line. So we want to find projection vector ‘**w**’ such that the data can be best separated. And projected data points are given by,



And



is within class scatter matrix and given by



And **m1** and **m2** are mean of each class.

For more than two classes case,

Between mean scatter matrix is given by,



And solving v = λ v (generalized eigenvalue problem),

will at most give c-1 distinct eigenvalues and thus we can obtain projection vector.

**Principal Component Analysis (PCA) method**

It is also known as K-L transform. This method tries to represent the data in best possible way. It finds a new feature space (m- dimensional) that is sufficient to describe data in original feature space where m<d.

A vector **x**described in terms of a set of basis vectors **b***i*.



The basis vectors (**b***i*) should be linearly independent and orthonormal, that is,



If we want to consider m (m<d) components of **y** and still represent **x** though with some error, we will calculate the first *m* elements of **y** and replace the others with constants.



Error:

Instead of plain error, mean square of error is used to quantify error and can be expressed as



Eigenvectors of covariance matrices of all classes are optimal basis vectors. Eigenvectors can be sorted corresponding to its eigenvalues in descending order and eigenvector with largest eigenvalue is principal component. Dimensions of data can be reduced to m from d by omitting eigenvectors corresponding to eigenvalues whose total sum (summing from smallest to larger eigenvalue) is not than error tolerance.

**kNN(k- Nearest Neighbors)**

To estimate *p*(*x*) from *n* samples, we can center a cell at *x* and let it grow until it contains *kn* samples, and *kn* can be some function of *n*



Normally, we let



Given *c* training sets from *c* classes, the total number of samples is

Given a point **x** at which we wish to determine the statistics, we find the hypersphere of volume ***V*** which just encloses *k* points from the combined set. If within that volume, *km* of those points belongs to class *m*, then we estimate the density for class *m* by



Using Bayesian decision rule and posterior probability,



So from the above posterior probability expression, we can derive a decision rule which tells us to look in a neighborhood of the unknown feature vector for *k* samples. If within that neighborhood, more samples lay in class *i* than any other class, we assign the unknown as belonging to class *i*. This is using majority voting principle.

For distance calculation, most commonly used is Euclidean distance metric while we can also use Minkowski distance.

In the project, distance from a sample data to all data in training set was calculated and sorted in ascending order. Then picking only k of the nearest distances, majority of class of training data in k samples was identified.

**k-means**

k-means is one of the clustering algorithms which assumes there are k clusters and starts by initializing cluster centers to arbitrary values.

Step 1: Begin with an arbitrary set of cluster centers and assign samples to nearest clusters

Step 2: Calculate sample mean of each cluster using samples in each cluster

Step 3: Reassign each sample to the cluster with the nearest cluster mean

Step 4: If assignment of sample to cluster does not change cluster, then stop; else repeat from step 2

**Winner takes all**

Winner takes all is improvement in k-means algorithm to increase speed of convergence. The winner (cluster center which is nearest to a sample) is updated on the fly immediately after a sample is classified to the cluster by pulling cluster towards the sample.

Step 1: Begin with an arbitrary set of cluster centers*i*

Step 2: Calculate sample mean of each cluster using samples in each cluster

Step 3: For each sample **x**, find the nearest cluster center, which is called the winner and assign the sample to the cluster

Step 4: Modifyusingnew = old + (**x**- old) whereis learning parameter and typically small value in order of 0.001

Step 5: If assignment of sample to cluster does not change cluster, then stop; else repeat from step 2



x

x-

**Kohonen Feature Map**

It is an extension of the winner-take-all algorithm and is also called self-organizing feature maps. In this algorithm, a problem dependent topological distance is assumed to exist between each pair of cluster centers. When the winner cluster center is updated so are other cluster centers as its neighbors in the sense of topological distance.



x

x-









The winning cluster center and its neighbors are trained based on the following formula:

are cluster centers

As k increases, decreases. (However, for this project it is chosen small value constant)

are coordinates of cluster centers

are coordinates of winning cluster center

The last expression tells that closer the neighbors are in topology, more affected they are.

**BackPropagation**

Back propagation is feed forward neural network algorithm where feedback is only used during training of network. It is a method of trainingneural network used in conjunction with an optimization method, which is least mean square in our case.

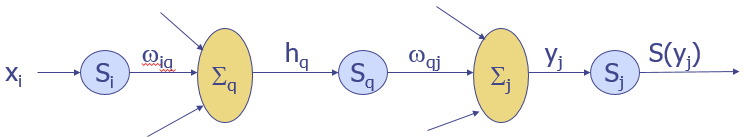


Figure: backpropagation three layer network

Where the objective function is ,

The backpropagation learning rule is based on gradient descent where the weights are initialized with random values, and changed in a direction that will reduce the error to find weight values.



Back propagation uses sigmoid function as threshold function which is continuous and derivable.

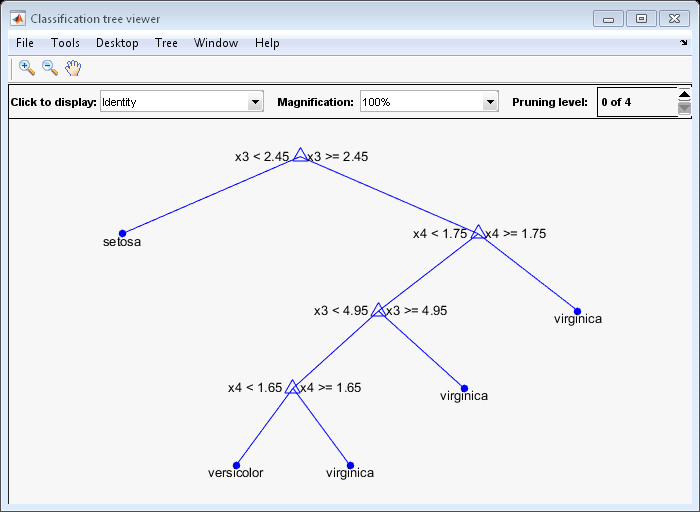


**Decision Tree**

Decision tree is one of the predictive modeling approaches used in statistics, data mining and machine learning. Tree models where the target variable can take a finite set of values are called classification trees and tree models where target variable can take continuous values are called regression trees. The term Classification And Regression Tree (CART) analysis is a term used to refer to both classification and regression procedures.

Decision tree learning constructs a decision tree from class labeled training samples. In a decision tree, each internal (non-leaf) node denotes a test on an attribute, each branch represents the outcome of a test, and each leaf (or terminal) node holds a class label. The topmost node in a tree is the root node.

Since it is more expressive and simple in training, most of the decision trees are binary tree. We seek property query at each node that makes data reaching the immediate leaf nodes contain as less impurity as possible.



**SVM**

Support Vector Machines are motivated by training the linear machines with margins, but rely on preprocessing the data to represent patterns in a high dimension — typically much higher than the original feature space. In this project, we used predefined function to apply classification by using SVM for multiclass data set.



To apply SVM:

For Non-separable Cases we can use:

1. SVM with soft margin
2. Kernel trick

**Classifier fusion**

Classifier fusion is integration of different classifiers for classifying a dataset. A classifier fusion is applied to get better accuracy, which means getting the classification result from multiple classifiers instead of only one and use all the results to generate one. In this project, we applied three classification fusion algorithms, which are majority voting, Naïve Bayes and behavior knowledge space.

**Majority voting**

There are three methods for applying majority voting:

1. Unanimity (100%)

To classify a sample to a certain class we have to have an agreement between different classifiers used for fusion.

1. Simple majority (50%+1)

To classify a sample to a certain class we have to have more than 50%+1 of the classifiers are agree to this class.

1. Plurality (most votes)

We pick the highest selected class as classification results.

**Naïve Bayes**

In Naïve Bayes,it is assumed classifiers are independent of each other.Confusion matrix can be generated based on known label of samples and to which class label it was classified by classifier. Then Bayesian decision rule can be applied as a decision rule to classify each sample based on confusion matrix:



Where

L classifiers, i=1,..,L

c classes, k=1,…,c

si: class label given by the ith classifier, i=1,…,L, s={s1,…,sL}.

**Behavior knowledge space**

In this algorithm,for a particular combination of output labels from each classifier, number of samples from actual class of each sample is calculated. Then the voting goes to the class containing maximum number of samples for that combination of output labels from different classifiers.And a lookup table can be generated for fused result of classifiers.

**CrossValidation**

Cross validation is a model evaluation method that is better than residuals. The problem with residual evaluations is that they do not give an indication of how well the learner will do when it is asked to make new predictions for data it has not already seen. One way to overcome this problem is to not use the entire data set when training a learner. Some of the data is removed before training begins. Then when training is done, the data that was removed can be used to test the performance of the learned model on new data. This is the basic idea for a whole class of model evaluation methods called cross validation.

**n-fold cross validation**

The dataset is divided into n different sets and training and testing method is carried out for n times where one set will be testing set and other n-1 set will be training. This is carried out for n different sets of data as testing set and performance accuracy is averaged.

**Experiments and Results**

The dataset was first normalized and then FLD was used to reduce dimensions from 18 to only 3. PCA was also performed on normalized with tolerance of 0.1 which resulted to reduced dimensions of 6.

All three datasets from normalization, FLD and PCA were each divided into three folds for better evaluation of performance of classifiers.

**Maximum posterior probability(MPP)**

After applying MPP for the three folds consist of normalized data set, we get the following error rate:

1. Case I

|  |  |
| --- | --- |
| Fold | Error rate |
| 1 | 0.475177 |
| 2 | 0.507092 |
| 3 | 0.453901 |
| Average error rate=0.4787 | |

Table x: Error rate from mpp case I

1. Case II

|  |  |
| --- | --- |
| Fold | Error rate |
| 1 | 0.230469 |
| 2 | 0.191489 |
| 3 | 0.212766 |
| Average error rate=0.21157 | |

Table X: Error rate from mpp case II

Comments:

We couldn’t perform mpp case III for normalized data that because the dimensions of the data set are 18 and the code couldn’t find the determinant of 18 by 18 covariance matrix. So, we chose case II as best case, and we performed mpp for reduced dimension data set by using case 2.

MPP for reduced dimension data set:

1. Case II for dimension reduced data set by using PCA

|  |  |
| --- | --- |
| Fold | Error rate |
| 1 | 0.47447 |
| 2 | 0.482269 |
| 3 | 0.471631 |
| Average error rate=0.4761 | |

Table X: Error rate from mpp case II for reduced data set by using PCA

1. Case II for dimension reduced data set by using FLD

|  |  |
| --- | --- |
| Fold | Error rate |
| 1 | 0.3014 |
| 2 | 0.3085 |
| 3 | 0.33333 |
| Average error rate=0.3144 | |

Table X: Error rate from mpp case II for reduced data set by using FLD

Comments:

After performing MPP for normalized data set and reduced dimension data set, we found the normalized data set gave us the best accuracy.

**KNN**

To accelerate the performance we used KNN with partial distance. After applying KNN for deferent K we got the following results:

1. Normalized data set fold1:

|  |  |
| --- | --- |
| K value | Error rate |
| 1 | 0.28 |
| 3 | 0.28 |
| 10 | 0.32 |
| 15 | 0.29 |
| 23 (the square root of the data set) | 0.32 |

Table x: the error rate for fold1 after applying KNN

1. Normalized data set fold2:

|  |  |
| --- | --- |
| K value | Error rate |
| 1 | 0.28 |
| 3 | 0.30 |
| 10 | 0.28 |
| 15 | 0.30 |
| 23 (the square root of the data set) | 0.30 |

Table x: the error rate for fold1 after applying KNN

1. Normalized data set fold3:

|  |  |
| --- | --- |
| K value | Error rate |
| 1 | 0.26 |
| 3 | 0.28 |
| 10 | 0.28 |
| 15 | 028 |
| 23 (the square root of the data set) | 0.31 |

Table x: the error rate for fold1 after applying KNN

Comments:

The result from different K’s are close. We chose 3 as a best value for K, so we decided to apply KNN with K=3 for reduced dimension data set.

For Reduced dimension data set we get the following results:

1. KNN with K=3 for dimension reduced data set by using PCA

|  |  |
| --- | --- |
| Fold | Error rate |
| 1 | 0.361 |
| 2 | 0.361 |
| 3 | 0.368 |
| Average error rate=0.363 | |

Table X: Error rate from KNN with K=3 for reduced data set by using PCA

1. KNN with K=3 for dimension reduced data set by using FLD

|  |  |
| --- | --- |
| Fold | Error rate |
| 1 | 0.28 |
| 2 | 0.30 |
| 3 | 0.28 |
| Average error rate=0.363 | |

Table X: Error rate from KNN with K=3 for reduced data set by using FLD

Comments:

After applying KNN for reduced dimension data set by using PCA, the accuracy for the classifier was reduced and the error rate was increased compared to results for KNN for normalized data set. However, the results from reduced dimension data set by using FLD were almost the same with normalized data set.

**Back Propagation**

For normalized data set:

|  |  |
| --- | --- |
| Fold | Error rate |
| 1 | 0.2092 |
| 2 | 0.2021 |
| 3 | 0.2128 |
| Average error rate=0.2080 | |

For PCA:

|  |  |
| --- | --- |
| Fold | Error rate |
| 1 | 0.4539 |
| 2 | 0.4291 |
| 3 | 0. 5177 |
| Average error rate=0.4669 | |

For FLD:

|  |  |
| --- | --- |
| Fold | Error rate |
| 1 | 0.2553 |
| 2 | 0.2695 |
| 3 | 0.2624 |
| Average error rate=0.2624 | |

From the classification accuracy results, back propagation seemed to have higher performance for normalized dataset.

**Decision Tree**

For normalized data set:

|  |  |
| --- | --- |
| Fold | Error rate |
| 1 | 0.2872 |
| 2 | 0. 2943 |
| 3 | 0. 3050 |
| Average error rate=0.2955 | |

For PCA:

|  |  |
| --- | --- |
| Fold | Error rate |
| 1 | 0.3865 |
| 2 | 0.3865 |
| 3 | 0.3794 |
| Average error rate=0.3841 | |

For FLD:

|  |  |
| --- | --- |
| Fold | Error rate |
| 1 | 0.3298 |
| 2 | 0.3546 |
| 3 | 0.2908 |
| Average error rate=0.3251 | |

For decision tree, the best accuracy is 0.2955 and is for normalized dataset.

**SVM**

After applying SVM for normalized data set and reduced dimension data set we get the following results:

1. SVM for normalized data set:

|  |  |
| --- | --- |
| Fold | Error rate |
| 1 | 0.2589 |
| 2 | 0.2943 |
| 3 | 0.2518 |
| Average error rate=0.26 | |

Table X: Error rate from SVM

1. SVM for dimension reduced data set by using PCA

|  |  |
| --- | --- |
| Fold | Error rate |
| 1 | 0.44 |
| 2 | 0.42 |
| 3 | 0.41 |
| Average error rate=0.42 | |

Table X: Error rate from SVMfor reduced data set by using PCA

1. SVM for dimension reduced data set by using FLD

|  |  |
| --- | --- |
| Fold | Error rate |
| 1 | 0.27 |
| 2 | 0.33 |
| 3 | 0.29 |
| Average error rate=0.29 | |

Table X: Error rate from SVM for reduced data set by using FLD

Comments:

We got the best result for SVM from normalized data set by using FLD, and it’s high accuracy compared to other classifiers.

**k-means:**

It was tested on a single whole sample set.

For normalized data set:

Accuracy=0.6501

For FLD:

Accuracy=0.6478

For PCA:

Accuracy=0.6158

**Winner Takes All**

For normalized data set:

Accuracy=0.6454

For FLD:

Accuracy=0.6477

For PCA:

Accuracy=0. 6158

**Kohonen Map**

For normalized data set:

Accuracy= 0.6489

For FLD:

Accuracy=0. 5284

For PCA:

Accuracy=0. 6478

**For all unsupervised clustering methods, performance error rate was found way higher.**

**Classifier Fusion**

After experimenting and classifying with all classifiers, the best three classifiers for the project were MPP case II, back propagation and kNN with k=3, and for normalized dataset. So we opted to do fusion using output labels from those classifiers on normalized datasets.

**Majority Voting (Plurality)**

It used only simple algorithm that chooses label with highest occurrence in out of three classifier output labels. Using this, accuracy of integrated classifier was measured to be 0.2399.

**Naïve Bayes**

After we selected the best classifier, we constructed the confusion matrix for each of them:

1. Confusion Matrix from MPP case 2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Class | 0 | 1 | 2 | 3 |
| 0 | 200 | 2 | 7 | 4 |
| 1 | 3 | 203 | 1 | 5 |
| 2 | 12 | 10 | 129 | 62 |
| 3 | 11 | 4 | 58 | 135 |

Confusion Matrix from KNN with K=3

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Class | 0 | 1 | 2 | 3 |
| 0 | 195 | 10 | 3 | 5 |
| 1 | 5 | 204 | 0 | 3 |
| 2 | 21 | 11 | 109 | 72 |
| 3 | 21 | 6 | 90 | 91 |

Confusion Matrix from Back Propagation

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Class | 0 | 1 | 2 | 3 |
| 0 | 192 | 2 | 3 | 2 |
| 1 | 3 | 210 | 2 | 3 |
| 2 | 6 | 8 | 150 | 53 |
| 3 | 8 | 10 | 76 | 118 |

Based on these results, we constructed the following lookup table:

|  |
| --- |
| c1 c2 c3 result |
| 0 0 0 0 |
| 0 0 1 0 |
| 0 0 2 0 |
| 0 0 3 0 |
| 0 1 0 0 |
| 0 1 1 1 |
| 0 1 2 2 |
| 0 1 3 3 |
| 0 2 0 0 |
| 0 2 1 2 |
| 0 2 2 2 |
| 0 2 3 3 |
| 0 3 0 0 |
| 0 3 1 3 |
| 0 3 2 2 |
| 0 3 3 3 |
| 1 0 0 0 |
| 1 0 1 1 |
| 1 0 2 2 |
| 1 0 3 2 |
| 1 1 0 1 |
| 1 1 1 1 |
| 1 1 2 1 |
| 1 1 3 1 |
| 1 2 0 2 |
| 1 2 1 2 |
| 1 2 2 2 |
| 1 2 3 2 |
| 1 3 0 2 |
| 1 3 1 1 |
| 1 3 2 2 |
| 1 3 3 3 |
| 2 0 0 0 |
| 2 0 1 2 |
| 2 0 2 2 |
| 2 0 3 3 |
| 2 1 0 0 |
| 2 1 1 1 |
| 2 1 2 2 |
| 2 1 3 2 |
| 2 2 0 2 |
| 2 2 1 2 |
| 2 2 2 2 |
| 2 2 3 2 |
| 2 3 0 2 |
| 2 3 1 2 |
| 2 3 2 2 |
| 2 3 3 3 |
| 3 0 0 0 |
| 3 0 1 3 |
| 3 0 2 3 |
| 3 0 3 3 |
| 3 1 0 0 |
| 3 1 1 1 |
| 3 1 2 2 |
| 3 1 3 3 |
| 3 2 0 3 |
| 3 2 1 3 |
| 3 2 2 2 |
| 3 2 3 3 |
| 3 3 0 3 |
| 3 3 1 3 |
| 3 3 2 3 |
| 3 3 3 3 |

By using this lookup table we tested the Naïve Bayes performance and got the following results

|  |  |
| --- | --- |
| Fold | Error rate |
| 1 | 0.24822 |
| 2 | 0.2375 |
| 3 | 0.2269 |
| Average error rate=0.237 | |

Comments:

From the previous table, we can see the overall performance is improved that compared by using only one classifier.

**Behavior Knowledge Space (BKS)**

Classification output labels from all kNN, MPP case II and BP were used to generate lookup table for BKS.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| c1 | c2 | c3 | t1(0) | t2(1) | t3(2) | t4(3) | result |
| 0 | 0 | 0 | 166 | 0 | 7 | 5 | 0 |
| 0 | 0 | 1 | 6 | 1 | 0 | 0 | 0 |
| 0 | 0 | 2 | 3 | 0 | 1 | 1 | 0 |
| 0 | 0 | 3 | 3 | 1 | 0 | 1 | 0 |
| 0 | 1 | 0 | 1 | 3 | 0 | 0 | 1 |
| 0 | 1 | 1 | 1 | 34 | 1 | 0 | 1 |
| 0 | 1 | 2 | 0 | 0 | 1 | 0 | 2 |
| 0 | 1 | 3 | 0 | 0 | 0 | 0 | 3 |
| 0 | 2 | 0 | 4 | 0 | 4 | 3 | 2 |
| 0 | 2 | 1 | 1 | 0 | 1 | 0 | 2 |
| 0 | 2 | 2 | 0 | 0 | 3 | 3 | 3 |
| 0 | 2 | 3 | 1 | 0 | 3 | 3 | 3 |
| 0 | 3 | 0 | 3 | 1 | 2 | 2 | 0 |
| 0 | 3 | 1 | 1 | 1 | 2 | 1 | 2 |
| 0 | 3 | 2 | 0 | 0 | 3 | 6 | 3 |
| 0 | 3 | 3 | 0 | 0 | 0 | 2 | 3 |
| 1 | 0 | 0 | 18 | 0 | 0 | 2 | 0 |
| 1 | 0 | 1 | 1 | 1 | 0 | 0 | 1 |
| 1 | 0 | 2 | 0 | 0 | 0 | 0 | 3 |
| 1 | 0 | 3 | 1 | 0 | 3 | 1 | 2 |
| 1 | 1 | 0 | 0 | 1 | 1 | 0 | 2 |
| 1 | 1 | 1 | 0 | 92 | 3 | 1 | 1 |
| 1 | 1 | 2 | 0 | 0 | 2 | 1 | 2 |
| 1 | 1 | 3 | 0 | 0 | 1 | 2 | 3 |
| 1 | 2 | 0 | 1 | 0 | 0 | 0 | 0 |
| 1 | 2 | 1 | 0 | 0 | 0 | 1 | 3 |
| 1 | 2 | 2 | 0 | 0 | 3 | 5 | 3 |
| 1 | 2 | 3 | 0 | 0 | 3 | 3 | 3 |
| 1 | 3 | 0 | 0 | 0 | 2 | 5 | 3 |
| 1 | 3 | 1 | 0 | 2 | 2 | 2 | 3 |
| 1 | 3 | 2 | 0 | 0 | 5 | 8 | 3 |
| 1 | 3 | 3 | 0 | 0 | 2 | 20 | 3 |
| 2 | 0 | 0 | 2 | 0 | 1 | 0 | 0 |
| 2 | 0 | 1 | 0 | 0 | 0 | 0 | 3 |
| 2 | 0 | 2 | 0 | 0 | 0 | 0 | 3 |
| 2 | 0 | 3 | 0 | 0 | 0 | 0 | 3 |
| 2 | 1 | 0 | 0 | 0 | 0 | 0 | 3 |
| 2 | 1 | 1 | 0 | 32 | 0 | 0 | 1 |
| 2 | 1 | 2 | 0 | 0 | 0 | 0 | 3 |
| 2 | 1 | 3 | 0 | 2 | 0 | 0 | 1 |
| 2 | 2 | 0 | 0 | 0 | 1 | 2 | 3 |
| 2 | 2 | 1 | 0 | 1 | 1 | 0 | 2 |
| 2 | 2 | 2 | 0 | 0 | 52 | 15 | 2 |
| 2 | 2 | 3 | 0 | 0 | 23 | 6 | 2 |
| 2 | 3 | 0 | 0 | 0 | 1 | 0 | 2 |
| 2 | 3 | 1 | 0 | 0 | 1 | 1 | 3 |
| 2 | 3 | 2 | 0 | 0 | 9 | 24 | 3 |
| 2 | 3 | 3 | 0 | 0 | 14 | 20 | 3 |
| 3 | 0 | 0 | 0 | 0 | 0 | 1 | 3 |
| 3 | 0 | 1 | 0 | 0 | 0 | 0 | 3 |
| 3 | 0 | 2 | 0 | 0 | 0 | 0 | 3 |
| 3 | 0 | 3 | 0 | 0 | 0 | 0 | 3 |
| 3 | 1 | 0 | 0 | 0 | 0 | 0 | 3 |
| 3 | 1 | 1 | 0 | 39 | 0 | 0 | 1 |
| 3 | 1 | 2 | 0 | 0 | 0 | 0 | 3 |
| 3 | 1 | 3 | 0 | 0 | 1 | 0 | 2 |
| 3 | 2 | 0 | 0 | 0 | 1 | 1 | 3 |
| 3 | 2 | 1 | 0 | 0 | 0 | 0 | 3 |
| 3 | 2 | 2 | 0 | 0 | 20 | 9 | 2 |
| 3 | 2 | 3 | 0 | 0 | 14 | 7 | 2 |
| 3 | 3 | 0 | 0 | 0 | 1 | 0 | 2 |
| 3 | 3 | 1 | 0 | 1 | 0 | 0 | 1 |
| 3 | 3 | 2 | 0 | 0 | 10 | 18 | 3 |
| 3 | 3 | 3 | 0 | 0 | 8 | 26 | 3 |

The accuracy using this lookup table was 0.1891. This was better than any of the performances noted for any classifier.

**Discussion**

In the project, we tried to classify vehicle silhouettes as different vehicles using different classification techniques. Supervised classification techniques were found to give better performance than unsupervised classifiers. And for improving performance of classification, we did experiment on fusion of best three classifiers. We used three different fusion algorithms plurality, Naïve Bayes and behavior knowledge space to test on fused results. The accuracy of fused classifiers on overall was found better than individual classifiers and among them behavior knowledge space fused classifier gave highest accuracy for the dataset.