Image classification using Householder Transform

P. V Nishana Rasheed [1] R. Shreej [2]

MTech Student [1], Assistant Professor [2]

Department of Computer Science and Engineering

MES College of Engineering

Kuttippuram - India

ABSTRACT

The problem of image classification has aroused considerable research interest in the field of image processing. Classification algorithms are based on the as assumption that image depicts one or more features and each of these features belong to one of the several distinct and exclusive *classes. Different* classification techniques have been analysed both traditional vector base method as well as Tensor based method. A novel classification method using HHT (Householder Transform) for matrix data is implemented. Unlike MRR (Multiple Rank Regression) in which computational complexity is more for uncorrelated data, In this method complexity is reduced.MRR was trial and error method. Multiple left projecting vectors and right projecting vectors are employed to regress each matrix data set to its label for each category.

This document gives formatting instructions for authors preparing papers for publication in the Proceedings of an IEEE conference. The authors must follow the instructions given in the document for the papers to be published. You can use this document as both an instruction set and as a template into which you can type your own text

Keywords:- Multiple Rank Regression; Tensor ;Supervised Learning; Principal Component Analysis; Regularization; Eigen Vectors; Eigen Faces.

I. INTRODUCTION

Classification algorithms are based on the as-sumption that image depicts one or more features and each of these features belong to one of the several distinct and exclusive classes. Image such as face images, palm images, or MRI [8] data are usually represented in the form of data matrices. Additionally, in video data mining, the data in each time frame is also a matrix. How to classify this kind of data is one of the most important topics for both image processing and machine learning. Most classification methods require that an image be represented by a vector, which is normally obtained by concatenating each row (or column) of an image matrix. Although the performances of traditional classification are prominent in many cases, they may be lack of *efficiency* in managing matrix data. The reasons mainly: When we reformulate an image matrix as a vector, the dimensionality of this vector is often very high. For example, for a small image of resolution 100×100, the reformulated vector is 10,000 dimensional. The performances of these methods will degrade due to the increase of dimensionality. With the increase of dimensionality, the computational time will increase drastically. If the matrix scale is a little larger, traditional approaches cannot be implemented in this scenario. When a matrix is expanded as a vector, we would lose the correlations of the matrix data. Aiming to preserve the correlation within the image matrix while reducing the computation complex- ity, researchers have proposed two-dimensional based analysing methods for images that are better represented as matrices.

A well- known approach within this paradigm is the two-dimensional subspace learning based classification. This approach is normally achieved by a two-step process. First, it eliminates noise and redundancy from the original data by projecting the data into a lower dimensional subspace. Then it applies *classifiers* on the low *dimensional* data for classification. A merit is that both computational *efficiency* and classification accuracy can be obtained. Classical works include the two-dimensional LDA. The aforementioned methods are able to preserve the spatial correlation of an image and to avoid the curse of dimensionality. Nonetheless, for classification they require a non-convenient two-step process, i.e., subspace learning followed by different classifiers. Although the first step processes image matrices directly, the classifying step still requires the data to be vectored. Be- sides, the separation of subspace learning and classification does not guarantee the *classifiers benefit* the most from the learned subspace.

SVM classifier which is able to classify image matrices in an integrated framework and a regression model for matrix data classification are encouraging, however, they need many labelled training data but labelled data are expensive to acquire. The over-fitting problem is likely to occur when the number of training data remains small. It would be more appealing if a classifier classifies image matrices with good performance by using only limited labelled training samples. A suitable classification system and *sufficient* number of training samples are prerequisites for meaningful classification. In literature survey several classification approaches have been proposed such as KNN, SVM, 1DREG, LDA,

2DLDA, GBR and MRR.

II. STEPS IN IMAGE CLASSIFICATION

Image classification is a complex process that may be affected by many factors. Non-parametric classifiers such as neural network, decision tree classifier, and knowledge-based classification have increasingly be- come important approaches for multi source data classification. Integration of remote sensing, geographical information systems (GIS), and expert system emerges as a new research frontier

The major steps of image classification may include determination of a suitable classification system, selection of training samples, image pre-processing, feature extraction, selection of suitable classification approaches, post-classification processing, and accuracy assessment. This section focuses on the description of the major steps that may be involved in image classification.

Another important factor influencing the selection of data is the type of image taken. Different natural images belonging to different classes may contain identical features. selecting images is therefore a tedious task . In this Project two sets of database is considered one be- longing to natural scenes image and another consists of images with specific characteristics.

• Selection of a classification system and training samples

A suitable classification system and a sufficient number of training samples are prerequisites for a successful classification. In general, a classification system is designed based on the users need

• Data pre-processing

Image pre-processing may include the detection and restoration of bad lines, geometric rectification or image registration, radiometric calibration and atmospheric correction, and topographic correction. If different ancillary data are used, data conversion among different sources or formats and quality evaluation of these data are also necessary be- fore they can be incorporated into a classification procedure. Accurate geometric rectification or image registration of remotely sensed data is a pre- requisite for a combination of different source data in a classification process.

• Feature extraction and selection

Selecting suitable variables is a critical step for successfully implementing an image classification. Many potential variables may be used in image classification, including spectral signatures, vegetation indices, transformed images, textural or con- textual information, multi temporal images, multi sensor images, and ancillary data. Due to different capabilities in land-cover separability, the use of too many variables in a classification procedure may decrease classification accuracy

It is important to select only the variables that are most useful for separating land-cover or vegetation classes, especially when hyper spectral or multi source data are employed. Many approaches, such as principal component analysis, minimum noise fraction transform, discriminant analysis, decision boundary feature extraction, non-parametric weighted feature extraction, wavelet transform, and spectral mixture analysis

• Selection of a suitable classification method

Many factors, such as spatial resolution of the remotely sensed data, different sources of data, a classification system, and availability of classification software must be taken into account when selecting a classification method for use. Differ- ent classification methods have their own merits. The question of which classification approach is suitable for a specific study is not easy to answer. Different classification results may be obtained de- pending on the classifier(s) chosen.

• Post-classification processing

Traditional per-pixel classifiers may lead to salt and pepper.. A majority filter is often applied to reduce the noises. Most image classification is based on remotely sensed spectral responses. Due to the complexity of bio- physical environments, spectral confusion is common among land-cover classes. Thus, ancillary data are often used to modify the classification image based on established expert rules. For example, forest distribution in mountainous areas is related to elevation, slope, and aspects. Data describing terrain characteristics can therefore be used to modify classification results based on the knowledge of specific vegetation classes and topographic factors.

In urban areas, housing or population density is related to urban land-use distribution patterns, and such data can be used to correct some classification confusions between commercial and high-intensity residential areas or between recreational grass and crops. Although commercial and high-intensity residential areas have similar spectral signatures, their population densities are considerably different. Similarly, recreational grass is often found in residential areas, but pasture and crops are largely located away from residential areas, with sparse houses and a low population density. Thus, expert knowledge can be developed based on the relation- ships between housing or population densities and urban land-use classes to help separate recreational grass from pasture and crops.

• Evaluation of classification performance Evaluation of classification results is an important process in the classification procedure. Different approaches may be employed, ranging from a qualitative evaluation based on expert knowledge to a quantitative accuracy assessment based on sampling strategies. To evaluate the performance of a classification method, six criteria are : accuracy, reproducibility, robustness, ability to fully use the information content of the data, uniform applicability, and objectiveness. In reality, no classification algorithm can satisfy all these requirements nor be applicable to all studies, due to different environmental settings and datasets used.

• Classification accuracy assessment

Before implementing a classification

accuracy assessment, one needs to know the sources of errors
. In addition to errors from the classification itself, other sources of errors, interpretation errors, and poor quality of training or test samples, all affect classification accuracy. In the process of accuracy assessment, it is commonly assumed that the difference between an image classification result and the reference data is due to the classification error.

III. CLASSIFICATION APPROACHES

In recent years, many advanced classification approaches, such as artificial neural networks, fuzzy-sets, and expert systems, have been widely applied for im- age classification. In general, image classification approaches can be grouped as supervised and unsupervised, or parametric and non-parametric, or hard and soft (fuzzy) classification, or per-pixel, sub pixel.
Per-pixel classification approaches

Traditional per-pixel classifiers typically develop a sig- nature by combining the spectra of all training-set pixels for a given feature. The resulting signature contains the contributions of all materials present in the training pixels, but ignores the impact of the mixed pixels. Per-pixel classification algorithms can be parametric or non-parametric. The parametric classifiers assume that a normally distributed dataset exists, and that the statistical parameters (e.g. mean vector and covariance)

• Whether training samples are used or not

1. Supervised
Land cover classes are defined. Sufficient reference data are available and used as

training samples. The signatures generated from the training samples are then used to train the classifier to classify the spectral data into a thematic map.

2. Unsupervised classification
Clustering-based algorithms are used to partition the spectral image into a number of spectral classes based on the statistical information inherent in the image. No prior definitions of the classes are used. The analyst is responsible for labelling and merging the spectral classes into meaningful classes.
• Whether parameters such as mean vector and co- variance matrix are used or not

1. Parametric classifiers
Gaussian distribution is assumed. The parameters (e.g. mean vector and covariance matrix) are often generated from training samples. When landscape is complex, parametric classifiers often produce noisy results. Another major drawback is that it is difficult to integrate ancillary data, spatial and con- textual attributes, and non-statistical information into a classification procedure

Non Parametric classifiers

No assumption about the data is required. Non-parametric classifiers do not employ statistical parameters to calculate class separation and are especially suitable for incorporation of non-remote-sensing data into a classi- fication procedure

• Which kind of pixel information is used

1. Per-pixel classifiers

Traditional classifiers typically develop a signature by combining the spectra of all training-set pixels from a given feature. The resulting signature contains the contributions of all materials present in the training-set pixels, ignoring the mixed pixel problems

2. Sub pixel classifiers

The spectral value of each pixel is assumed to be a linear or non-linear combination of defined pure materials (or end members), providing proportional membership of each pixel to each end member

• Output is a definitive decision about land cover class or not

1. Hard classification

Making a definitive decision about the land cover class that each pixel is allocated to a single class. The area estimation by hard classification may produce large errors, especially from coarse spatial resolution data due to the mixed pixel problem

2. Soft (fuzzy) classification

Providing for each pixel a measure of the degree of similarity for every class. Soft classification provides more information and potentially a more accurate result, especially for coarse spatial resolution data classification.

IV. ADVANCED CLASSIFICATION APPROACHES

In recent years, many advanced classification approaches, such as artificial neural networks, fuzzy-sets, and expert systems, have been widely applied for image classification.

1. Per-pixel classification approaches
   Traditional per-pixel classifiers typically develop a signature by combining the spectra of all training- set pixels for a given feature. The resulting sig- nature contains the contributions of all materials present in the training pixels, but ignores the im- pact of the mixed pixels. Per-pixel classification algorithms can be parametric or non-parametric

2. The parametric classifiers
   It assume that a normally distributed dataset ex- ists, and that the statistical parameters (e.g. mean vector and covariance matrix) generated from the training samples are representative. The maximum likelihood may be the most commonly used para- metric classifier in practice, because of its robust- ness and its easy availability in almost any image-processing software

Drawback
The assumption of normal spectral distribution is often violated, especially in complex landscapes.
In addition, insufficient, non-representative, or multi mode distributed training samples can further introduce uncertainty to the image classification procedure.
Another major drawback of the parametric classifiers lies in the difficulty of integrating spectral data with ancillary data.

3. Non-parametric classifiers For this, the assumption of a normal distribution of the dataset is not required. No statistical parameters are needed to separate image classes. Non-parametric classifiers are thus especially suitable for the incorporation of non-spectral data into a classification procedure. Much previous research has indicated that non-parametric classifiers may provide better classification results than parametric classifiers in complex landscapes.
Among the most commonly used non-parametric classification approaches are neural networks, decision trees, support vector machines, and expert systems. In particular, the neural network approach has been widely adopted in recent years. The neural network has several advantages, including its non-parametric nature, arbitrary decision boundary capability, easy adaptation to different types of data and input structures, fuzzy output values, and generalization for use with multiple images, making it a promising technique for land-cover classification
The multilayer perception is the most popular type of neural network in image classification.

V. .LITERATURE SURVEY

In the literature survey , a lot of classification approaches have been proposed, such as K-Nearest Neighbourhoods classifier (KNN) , Support Vector Ma- chine (SVM) and Regression methods . Some of them are similarity based, such as KNN. Some of them are margin based, such as SVM. Among these approaches, due to their simplicity, effectiveness, and inductive nature, regression methods have widely been used in many real applications . This chapter briefly presents some of such approaches to various classification methods.

5.1. K- Nearest Neighbour Classifier (KNN)

K-Nearest-Neighbour classifier (KNN)[1] by G. Shakhnarovich which is similarity based. In pattern recognition, the k-nearest neighbour algorithm (KNN) is a non-parametric method for classification and regression, that predicts objects' "values" or class member- ships based on the k closest training examples in the feature space. KNN is a type of instance-based learning, or lazy learning where the function is only approximated locally and all computation is deferred until classification. The k-nearest neighbour algorithm is amongst the simplest of all machine learning algorithms. An object is classified by a majority vote of its neighbour, with the object being assigned to the class most common amongst its k nearest neighbours (k is a positive integer, typically small).If k = 1, then the object is simply assigned to the class of that single nearest
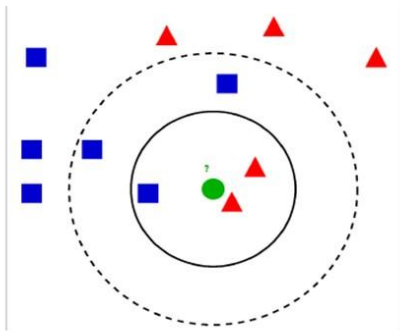
neighbour.

Figure 1: K-Nearest Neighbor Classifier

5.2. Support Vector Machine

Support Vector Machine (SVM) [2] by V. N. is viewed as a p-dimensional vector and separate such points with a (p - 1)-dimensional hyperplane. This is called a linear classifier. There are many hyperplanes that might classify the data reasonable choice as the best hyperplane is the one that represents the largest separa- tion, or margin, between the two classes .
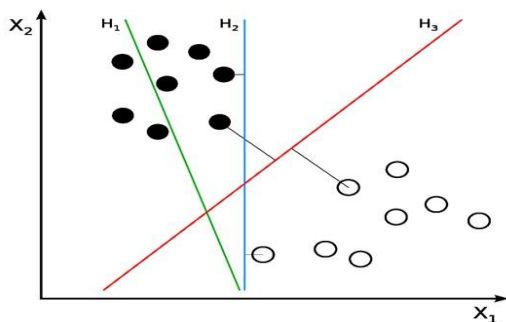

Figure 2: Support Vector Machine

The fig.2 shows 3 Hyperplanes in 2-Dimensional space. H3 does not separate the two classes, H1 does,

with a small margin and H2 with the maximum margin. The goal of SVM is trying to find H2.

5.3. Linear Discriminant Analysis(LDA)

Curse of dimensionality is that higher the dimension of the feature vectors leads to data sparsity and under trained classifier. It is important to try to reduce dimension of feature vectors without loss of informa- tion.LDA tries to optimize class separability It is also known as Fishers discriminant analysis. When the training data set are labelled for each identity, supervised training techniques like

LDA are more profitable for feature extraction compared with methods of unsupervised learning. By applying the supervised learning, illumination variation and pose variation will be removed and retaining the identity information. The LDA provides a procedure to determine a set of axes whose projections of different groups have the maximum separation Linear Discriminant Analysis projects data on to a lower dimensional vector space such that the ratio of between-class distance to within class distance is maximized thus achieving maximum discrimination between classes. It suffers from singularity problem. It is based on maximizing the distance means of the classes.

The between class scatter matrix is given as

Where

$$S_b = \sum_{k=1}^{k} p_k (\vec{\mu_k} - \vec{\mu})(\vec{\mu_k} - \vec{\mu})^t$$

k=number of classes.

$p_k = N_k / \sum_{l=1}^{k} N_l$ (where p_k is the fraction of data belonging to class k)

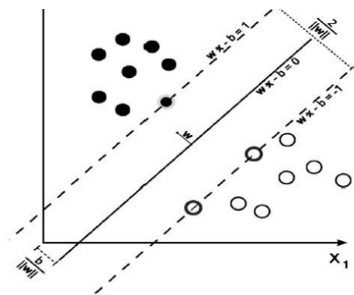$\vec{\mu_k} = 1/N_k \sum_{i=1}^{N_k} \vec{X_i}, k$ (where μ_k is mean of class k.)

μ is mean of all vectors

- The within class scatter matrix is given as

where $S_w = \sum_{k=1}^{k} p_k \sum_{k}$ (2)

$\sum_k = 1/N_k \sum_{i=1}^{N_k} (\vec{X_{i,k}} - \vec{\mu_k})$

$(\vec{X_{i,k}} - \vec{\mu_k})$(which is covariance matrix of class$k$)

In fig.3 Direction W is taken such that both differences between the class means projected on to these directions 1 and 2 is large and variance(s 1 and s 1) around these mean is small.
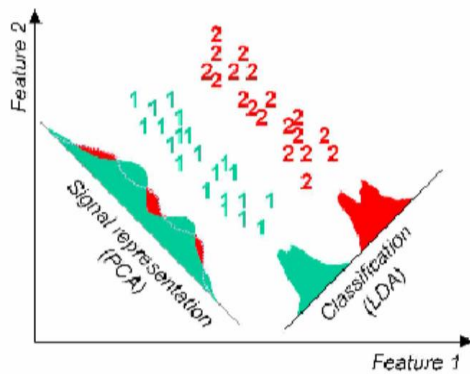
Figure 4: Linear Discrimnant Analysis

5.4. Two-dimensional Linear Discriminant Analysis

Two-dimensional Linear Discriminant Analysis (2DLDA) [4] by J. Ye, R. Janardan is a popular supervised tensor based approach. 2DLDA aims to find two transformation matrices L and R, which map X_i to its low dimensional embedding, i.e., Z_i, by the equa- tion

$$Z_{i_i} = L^T X_i R \quad (3)$$

Since it is difficult to derive the optimal L and R simultaneously, 2DLDA solves the above problem in Eq.(3) in an alternative way. Briefly, it fixes L in computing R and fixes R in computing L. The within and between class distance is given by the the Eq(4) and Eq(5) respectively.

$$D_w = Tr(\sum_{j=1}^{c} \sum_{x_i \in M_j} L^T(X_i - \overline{X_j})RR^T(X_i - \overline{X_j})L) \quad (4)$$

It tries to minimize the within-class distance D_w and maximize the between-class distances D_b .

$$D_b = Tr(\sum_{j=1}^{c} l_j L^T(\overline{X_j} - \overline{X})RR^T(\overline{X_j} - \overline{X})^T)L \quad (5)$$

5.5. One Dimensional Regression(1DREG)

Among these approaches, due to their simplicity, effectiveness and inductive nature one dimensional Regression methods (denoted as 1DREG) [3] by C. M. Bishop regression methods have been widely used in many real applications. 1DREG is a representative method in vector-based regression works. It is also a famous model for classification. Denote the matrix data X_i (ith training matrix data) as an mn-

dimensional vector data x_i by connecting each row (or column). 1DREG aims to regress each data to its label vector by computing c transformation vectors and constant denoted as W = [w1, w2, ..., c] where $\varepsilon R^{mn \times c}$ and b = [b1, b2, ..., bc].In order to avoid over fitting, we of- ten add a regularizer. The most commonly used one is the Tikhonov regularization. Briefly, the objective func- tion of 1DREG with Tikhonov regularization is given by the Eq(6).

$$L(W,b) = \sum_{i=1}^{l} \| W^T X_i + b - Y_i \|_F^2 + \alpha \| W \|^F \quad (6)$$

where k is Frobenius norm of a matrix 1DREG converts the matrix data into a vector. Thus, it will losethe correlation of matrix data and its computational time consuming is unacceptable if the matrix scale is large

5.6. General Bilinear Regression(GBR)

As mentioned in [6], GBR is the two- dimensional counterpart of 1DREG. It replaces the regression function of 1DREG by a bilinear regression function. More concretely, in two class scenario, it is assumed that the left and right projection vectors are u and v and its objective function is given by the Eq.(7). Besides, it only uses one left projecting vector together with one right projecting vector. Its fit- ting error is too large for some real regression problem.

5.7. Multiple Rank Regressions (MRR)

Multiple Rank Regression Model is meant for matrix data classification . Unlike traditional vector-based methods, multiple-rank left projecting vectors and right projecting v $L(u,v,b) = \sum_{i=1}^{l} \| u^T X_i v + b - y_i \|^2$ ta set to its la er accuracy and has lower computational complexity. Compared with traditional supervised tensor-based methods, MRR performs better for matrix data classification. Computational complexity is more for uncorrelated data in this method.MRR can be extended for unsupervised and semi supervised cases. Eq(8) is reduced to Eq.(9).

$$L(w,b) = \sum_{i=1}^{l} \| W^T x_i + b - y_i \|^2 + \alpha \| W^2 \|_F \quad (8)$$

$$\sum_{r=1}^{c} \sum_{i=1}^{l} W_r^T x_i + b_r - y_{ir}^2 + \alpha w_r^T w_r \quad (9)$$

As seen from Eq. (9) it is clear that this regression model is a combination of multiple two-category classifiers via one versus rest strategy. More concretely, in training the classifier for the rth category, the labels for the points who belong to the rth category are one. If a point does not belong to this class, its label is zero. Moreover, this training process is separate. We can regard it as c independent procedures, in which we only compute the corresponding w_r and b_r for r = 1, 2, . . ., c. In other words, the formulation in Eq.(9) can be regarded as training c classifiers for c categories separately. One direct way in constructing the loss function is to replace the traditional projection term, i.e., $w^T x_i$ in Eq.(9), by its tensor counterpart, such as $u^T X_i v_r$ where u_r and v_r are the left and right transformation vectors for the rth category. By doing the replacement regression error increases.

To solve this problem, instead of using merely one couple of projecting vectors, i.e., the left projecting vector u_r and right projecting vector v_r for the rth classifier, using k couples of left projecting vectors and right projecting vectors is proposed. They are denoted as

$$\{u_j^{(r)}\}_{j=1}^k \quad \{v_j^{(r)}\}_{j=1}^k$$

The intuition is shown in the bottom of Fig(5). Compared with the employment of only one couple of projecting vectors, there are several advantages of this method. Since we have multiple rank projecting vectors, the above mentioned constraints will be relaxed to some extent and consequently, the joint effects of these projections will decrease the regression error. k is the parameter to balance the capacity of learning and generalization. GBR is the special case of MRR when k = 1.

Formally, the first loss function is to train the rth classifier is
Where b_r is the unknown constant for the rth category.

$$\sum_{i=1}^{l}(\sum_{j=1}^{k}(u_j^{(r)})^T \times V_j^{(r)} + b_r - y_{ir})^2$$

VI. ALGORITHM

Alg 1 : Training Step in Classification using

 MRR

 Input:

X = x_1, x_2, . . ., x_l // set of n Input images belonging to C classes in Training Set.

C // Number of desired Labels

Output:

Optimised Right Regression Vector for i=1,2,...l

Steps:
1.Find the correlation matrix XX^T .
2.Diagonalise the correlation matrix by finding the eigen vector.
3.Fix left vector and find the projection of left vector on correlation matrix.
4.Find right vector so as to get the required label in such a way that dot product of 2 label vector is 0.
5.find the mean value of the right vector to fix a single right vector for all the images.

Alg 1 : Testing Step in Classification using

MRR

Input:
        set of n Testing images X_i /i = l + 1, l + 2, . . ., l + t
.
Output:
        Labels for testing data y_i /i = l + 1, l + 2, . . ., l + t
Steps:
1.Read the test image X_i
2.Find the projection of left vector on X_i .
3.Find the resultant projection on right vector.
4.Find label.
5.Find the minimum of l_2 norm ie-
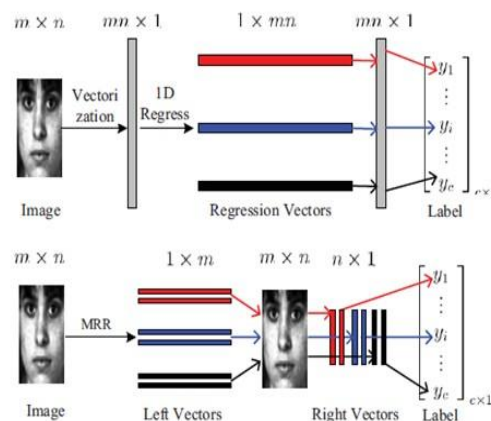
$$:\| l_{training} - l_{testing} \|.$$

Figure 5: Intuition of multiple rank regression. The top procedure is traditional regression and the bottom is multiple rank regression.

6.1. Observation and Analysis

In this section the performance of various classification methods are analysed w.r.t accuracy and

computational complexity. Since one of the motivations is to reduce the computational complexity of traditional regression methods, hence different the compu- rational complexity of related methods like 1DREG, LDA, 2DLDA, GBR and MRR are analysed.

- The first group of methods is LDA and 2DLDA. As seen from the procedure of LDA and 2DLDA, the most time consuming step is to solve the Eigen decomposition problem. Its computational complexity is about $O(D^3)$, where D is the dimensionality of original data. Thus, traditional LDA has the computational complexity $O(m^3 n^3)$. 2DLDA solves two Eigen-decomposition problem with the sizes m and n respectively. Thus, its computational complexity is $O(s(m^3 + n^3))$, where s is the time of iteration.

- The second group of methods is 1DREG, GBR and MRR. The most time consuming step of 1DREG, GBR and MRR is to solve the regularized least square regression problem. It has the computational complexity about $O(D^2)$, where D is also the dimensionality. Since 1DREG treats a m   n matrix as a mn-dimensional vector, its computational complexity is $O(c(m \, n))$. In each iteration, GBR solves two regularized least square regression problems with the data dimensionality m and n respectively. Assume that there are mainly s iterations, the computational complexity of GBR is $O(sc(m^2 + n^2))$. Similar to GBR, MRR solves two regularized least square regression problems with the dimensionality mk and nk respectively. Thus, its computational complexity is $O(sc(m^2 + n^2)k^2)$. Commonly s is less than 10 and k is far less than min m, n. Thus, the computational complexity of MRR is similar to GBR and much smaller than 1DREG.

- In summary, the computational complexities of four methods have the following relationships. GBR $\leq$ MRR $\leq$ 2DLDA $\leq$ 1DREG $\leq$ LDA. There are several observations from the performance comparisons as follows.

  - Among different methods and different data sets, MRR performs best. It achieves the highest accuracy in most cases. This is mainly due to the fact that MRR has smaller fitting error and stronger capacity for generalization.

  - With the increase of training points number, all methods achieve higher accuracies. This is consistent with intuition since we have more information for training.
  - For classification, 2D based methods do not al- ways perform better than 1D based methods. LDA achieves higher accuracy than 2DLDA in most cases. The reason may be that the adding

constraints in 2DLDA will degrade the performances.

6.2. Experiments

There are also some observations from the results shown in the table. With each fixed number of training points training points for 50 runs were selected randomly. mist and Ar are face images. The data size ranges from about 500 to 11000 and the image resolu- tion ranges from 16×16 to 64×64. The calculations are made with a naive MATLAB implementation on a 3.2-GHz Windows machine

- It can be seen that among *different* methods on different data sets, GBR consumes the least time. Al- though MRR costs a little more time than GBR, it still consumes much less time than other one dimensional methods. Among different methods on *different* data sets, GBR consumes the least time.

- Comparing the results on different image resolutions, we can see that dimensionality is the key factor in dominating the computational complex- ity. Certainly, with the increase of training points, all methods need more time. Nevertheless, com- pared with the influence of dimensionality, its ef- fect is not so significant.

- Computational time on AR data with different number of training points are shown in table. The scale of 32×32 and 64×64 are considered to show how varying resolution effects computational time. Details of the computation time is summarized in Table Fig(6)

- Also results from the fig(7) reveal that With the increase of training points, all methods achieve higher accuracies. This is consistent with intuition since we have more information for training.

- For classification, 2D based methods do not al- ways perform better than 1D based methods. Take the results in fig(7) as an example, LDA achieves higher accuracy than 2DLDA in most cases. The reason may be that the adding constraints in 2DLDA will degrade the performances.

	LDA	2DLDA	1DREG	GBR	MRR
20	215.79±2.38	6.71±1.73	63.78±0.89	0.04±0.02	0.12±0.02
40	217.76±2.40	8.31±1.62	64.42±0.89	0.07±0.02	0.15±0.03
60	219.96±2.59	10.75±1.94	65.60±0.72	0.14±0.19	0.15±0.06
80	225.12±2.80	13.67±1.49	65.92±0.46	0.15±0.17	0.25±0.16
100	240.75±2.90	14.82±1.02	66.82±0.28	0.20±0.09	0.28±0.11

Figure 6: Computational Time of Different Classification Methods on AR Data With *Different* Number of Training Points

	SVM	LDA	2DLDA	1DREG	GBR	MRR
40	59.82±1.47	63.66±1.87	61.74±1.69	63.85±1.84	65.59±1.54	**69.20±1.30**
60	66.65±1.74	75.33±1.85	67.83±1.66	74.34±1.74	75.73±1.81	**78.81±1.40**
80	72.73±1.71	82.76±1.58	82.17±1.15	82.46±1.64	82.64±1.37	**85.19±1.14**
100	77.38±1.44	86.37±0.95	**87.66±1.68**	85.63±1.34	86.41±1.44	88.63±1.12
120	81.01±1.38	88.10±0.96	87.86±0.96	87.76±1.00	87.73±1.26	**90.56±0.98**
140	85.03±1.42	**91.33±0.95**	90.03±0.99	90.29±0.97	89.40±0.90	92.40±0.83
160	87.78±0.81	91.58±0.79	**92.70±0.89**	91.90±0.82	91.07±0.98	93.78±0.76
180	89.81±1.07	92.59±0.69	93.03±0.63	92.83±0.73	92.33±0.94	**94.73±0.74**
200	91.25±0.90	92.49±0.82	93.62±0.68	93.52±0.69	92.81±0.91	**95.51±0.72**
220	92.39±0.76	93.23±0.87	93.49±0.75	94.23±0.77	93.45±0.86	**96.21±0.69**

Figure 7: Classification Accuracy of *Different* Methods on UMIST Data

VII. PROPOSED METHOD

The proposed method is named as classification using Householder transform. It is not a trial and error method as MRR. Householder transformations are orthogonal transformation (reflections) that is used to introduce zeroes into lower triangle of a matrix. In our transformation A is the matrix representing mean of all images of a class.

let $A^{(0)}$ = A Reflection across orthogonal to unit normal vector V can be expressed in matrix form as

$$H = I - 2V_1 V_1^T / V_1^T V_1 \qquad (12)$$

The reflector V is computed as

7.2. Algorithm

Algorithm 1 : Training Step in Classification using HHT

Input:

X = x_1, x_2, . . . , x_l // set of n Input images belonging to C classes in Training Set.
C // Number of desired Labels. Output:
Projection of mean image of each class on basis vector
Steps:
1. Find mean of images of each class;
2.Triangularize mean image of each class by using HouseHolder Transform to get the C basis vectors(OPi).
3. Get the projection of mean of each class of images on C basis vector.

Algorithm 2 : Testing step in Classification using HHT

Input: X = x_1 , x_2 , . . . , x_l // set of n Input images in Test Set
C //Projections of mean of each class of images on C basis vector.

Output:

Projection of image to the class to which it belongs.
Steps:1. Get the projection of input image on the each of the C basis vectors (OPi) ;
Step 2. Find Euclidean Distance between inner product taken in step 3 of alg1 with inner product taken in step1 of alg2.

7.3. steps to compute householder transform

- $LetA^{(0)} = A$
- compute $V_1 = a_1^{(0)} - sign(a_{11})\|a_1^0\| * e_1$
- compute Householder Transform matrix $H_1 = I - 2V_1 V_1^T / V_1^T V_1$
- compute $A^{(1)} = H_1 * A^{(0)}$

- Next consider the sub matrix of $A^{(1)}$(removing the first row and column)
- compute $V_2 = a_1^1 - sign(a_{11}\|a_1^{(1)}\| * e_1$
- $H_2 = I - 2V_2 V_2^T / V_2^T V_2$
- compute $A^{(2)} = H_2 * A^{(1)}$

VIII. EXPERIMENTS AND RESULTS

8.1. Implementation

The proposed algorithm were implemented, tested and compared. Implementations were done in Matlab 7.7.0(R2008b). The data sets available in the UCI data repository were used for testing. Details of the data sets used are summarized in fig(5.1).The database consists of 48 trained images belonging to 8 classes. There are a no: of images in test database. The images of that exhibits variations in terms of illumination are normalised by the algorithm. The input and test
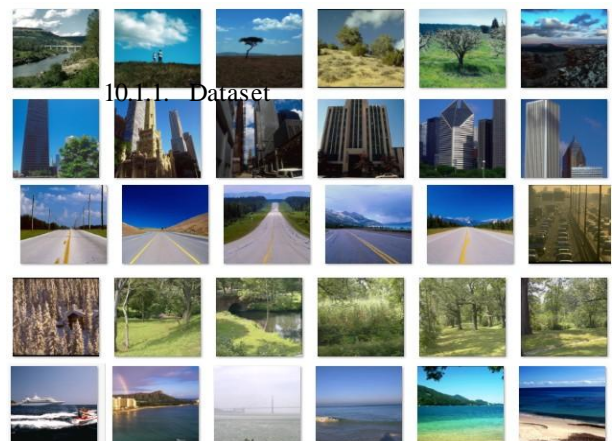images are resized to 300 $\times$300.

10.1.1. Dataset

Figure 8: Images belonging to five different dataset

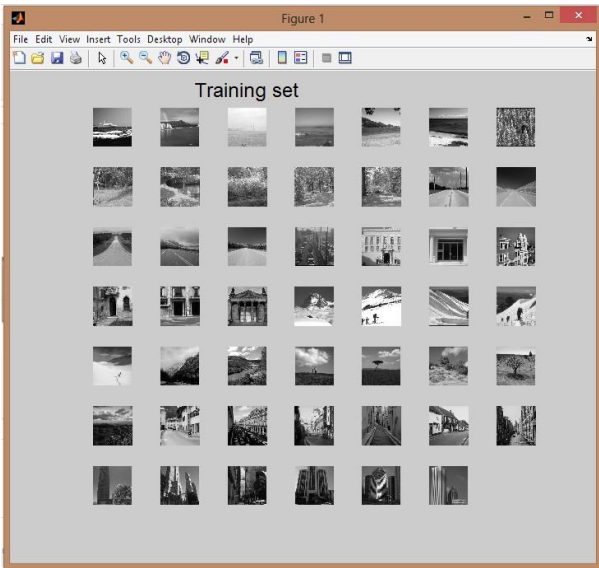8.1.2.   Result of MRR:Training set data

Figure 9: Training set data
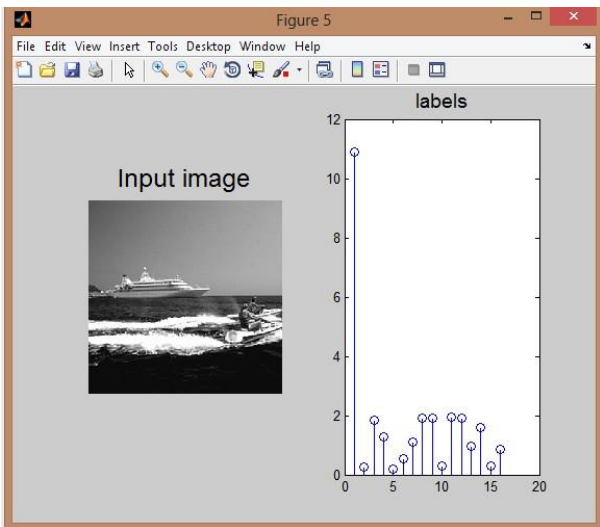
8.1.3.   Result of MRR :Classified output 1

Figure 10: Result showing after giving input from set of test images

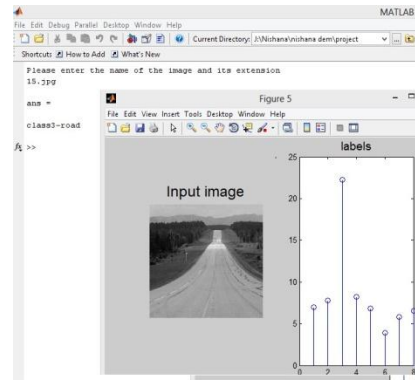8.1.4.   Result of MRR cont..:Classified output 2

Figure 11: Result showing after giving input from set of test images
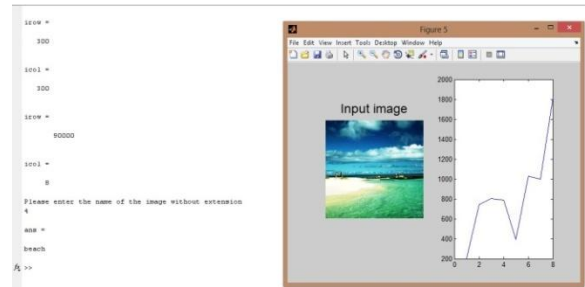
8.1.5.   Result of HHT

Figure 12: Result showing after giving input from set of test images

8.1.6. Discussion

Classification Accuracy using Householder Trans- form was found to be better than classification using Multiple Rank Regression.30 test images were taken out of which 26 images were classified correctly using HHT as against 20 images using MRR. Also HHT was found to have higher noise tolerance over MRR. Different levels of noise were input for different classes of images HHT was found to be more tolerant to noise. The Table1 gives the details of the experiment

Table 1: Performance analysis showing noise tolerance

Image1	(MRR)	(HHT)

Beach image	0.01-0.358	0 .01-0.8
Forest image	0.01-0.655	0.01-0.795
Building image	0.01-0.540	0.01-0.565

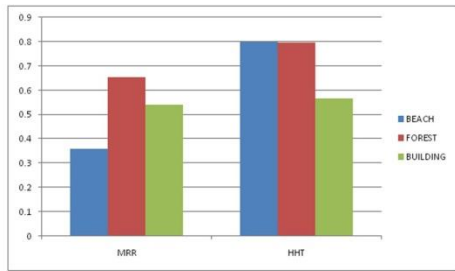10.1.7. Graph showing comparison of MRR with HHT

Figure 13: Noise tolerance of MRR vs HHT

IX. CONCLUSION

- Classification using Householder Transform is used for achieving lower computational time.
- Found computation time of MRR higher compared to Householder Transform.
- Noise is added to the test image with various noise variance.
- Proposed method is found to have better noise tolerance which can be computed based on accuracy.
- In future instead of applying classification using HHT on image matrix,it may be applied on image features to have more accurate result.

REFERENCES

[1] G. Shakhnarovich , T. Darrell, and P. Indyk, Nearest-Neighbor Methods in Learning and Vision: Theory and Practice (Neural Information Processing)Cambridge, MA: MIT Press,2006.

[2] V. N. Vapnik, The Nature of Statistical Learning Theory,New York:Springer-Verlag, 1995.

[3] C. M. Bishop, Pattern Recognition and Machine Learning ,Se- caucus, NJ: Springer-Verlag, 2006.

[4] J. Ye, R. Janardan, and Q. Li, Two-dimensional linear discrimi- nant analysis , Advances in Neural Information Processing Sys- tems, Cambridge, MA: MIT Press, 2004.

[5] Chenping Hou, Member, IEEE, Feiping Nie, Dongyun Yi, and Yi Wu , Efficient Image Classification via Multiple Rank Re- gression , IEEE Transaction On Image Processing , January 2013.

[6] K. R. Gabriel, Member, IEEE, Feiping Nie, Dongyun Yi, and Yi Wu , Generalised bilinear regression , Biometrika, vol. 85,no. 3, pp. 689700, 1998.

[7] A. W.-K. Kong, D. D. Zhang, and M. S. Kamel, Member, IEEE, Feiping Nie, Dongyun Yi, and Yi Wu , A survey of palmprint recognition , Pattern Recognition , vol. 42, no. 7, pp. 14081418, 2009.

[8] J. J. Koo, A. C. Evans, W. J. GrossH. GOVIL , M. KUMAR and S. FAROOQ , 3-D brain MRI tissue classification on FPGAs , IEEE Trans. Image Process, vol. 18, no. 12,pp. 27352746, Dec. 2009.