Feature Reduction using a Singular Value Decomposition for the Iterative Guided Spectral Class Rejection Hybrid Classifier

Rhonda D. Phillips, Layne T. Watson, *Fellow, IEEE*, Randolph H. Wynne, *Member, IEEE*, and Christine E. Blinn

Abstract—Feature reduction in a remote sensing dataset is often desirable to decrease the processing time required to perform a classification and improve overall classification accuracy. This work introduces a feature reduction method based on the singular value decomposition (SVD). This feature reduction technique was applied to training data from two multitemporal datasets of Landsat TM/ETM+ imagery acquired over a forested area in Virginia, USA and Rondônia, Brazil. Subsequent parallel iterative guided spectral class rejection (pIGSCR) forest/non-forest classifications were performed to determine the quality of the feature reduction. The classifications of the Virginia data were five times faster using SVD-based feature reduction without affecting the classification accuracy. Feature reduction using the SVD was also compared to feature reduction using principal components analysis (PCA). The highest average accuracies for the Virginia dataset (88.34%) and for the Rondônia dataset (93.31%) were achieved using the SVD. The results presented here indicate that SVD-based feature reduction can produce statistically significantly better classifications than PCA.

I. INTRODUCTION

In the remote sensing and image processing discipline, large data volumes and slow processor speeds have necessitated feature reduction. Since the cost of classifications is dependent upon the number of discriminating features (bands) associated with each pixel in multispectral space, it is desirable to reduce the number of features in a dataset [25]. Even as processing speeds increase due to faster computers and better algorithms, such as the parallel iterative guided spectral class rejection (pIGSCR) classification algorithm [22], the need for feature reduction methods remains as modern sensors increase in sensitivity. Iterative guided spectral class rejection (IGSCR) [30][21] is a hybrid classification method, which combines the automation of the cluster (spectral) class creation of unsupervised classification with the mathematical cluster assignment procedure of supervised classification [14][4][5]. A supervised classification uses a decision rule to assign individual pixels (or objects) to a particular prelabeled spectral class (training data prepared by an analyst). An unsupervised classification uses a clustering method to mathematically group pixels (or objects) within an image and then an analyst assigns a label to each cluster. Hybrid classifiers that often exhibit characteristics of both supervised and unsupervised classification may be suitable for feature selection, feature reduction, or creative methods unique to the hybrid classifier [26][3].

Traditional methods of feature selection include the use of separability indices such as divergence, the Jeffries Matusita Distance, and transformed divergence [25]. A set of spectral classes are analyzed to determine which combinations of bands will result in the greatest separability (greatest distinction between classes), and only those bands are used for the ensuing classification. Another popular approach is that of feature reduction, where the image is transformed to a new coordinate system requiring (hopefully) fewer bands to accurately represent the image. Most feature reduction methods do not require analysis of training data, making this an attractive option for a classification method that does not require training, such as unsupervised classification methods and some hybrid classification methods. Standard feature reduction methods include principal components analysis (PCA) (also called Karhunen-Loeve analysis) [14], maximum noise fraction (MNF) [12], canonical analysis [9], and the

R. D. Phillips and L. T. Watson are with the Departments of Computer Science and Mathematics, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061.

R. H. Wynne and C. E. Blinn are with the Department of Forestry, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061.

Kauth-Thomas tasseled cap transformation [16]. The tasseled cap transformation has a fixed axis and is therefore somewhat confined in its application. The other feature reduction methods mentioned are transformations that align the data along axes of decreasing variance, and the resulting low order de-correlated bands are sufficient to perform a classification in many cases. However, Lowitz [19] and Chang [6] have shown that sometimes the higher order components resulting from such transformations are necessary to differentiate between classes in a classification. Also, the axes generated using PCA may not allow for accurate classification of the data using fewer dimensions while a different set of axes exists that will allow for class discrimination using fewer dimensions, the premise upon which canonical analysis is based [25]. Furthermore, rather than directly revealing the rank and basis of the data from the data itself, these methods attempt to reveal these attributes indirectly from a summary of the data, such as the covariance matrix. This explains why the resulting PCA data has full rank (no reduction is possible), but it is still possible that a different alignment of the data will result in a feature reduction.

A mathematical construct that directly reveals the rank and corresponding ideal basis of a dataset is the singular value decomposition (SVD). For a dataset in n-dimensional space, for any $k < n$, the SVD will show the ideal basis for representing that data using only k dimensions, as will be explained in a later section. If the SVD reveals that the dataset is full rank and no feature reduction is possible along the calculated axes, then no axes exist for which a reduction is possible.

In remote sensing applications, the SVD is a popular alternative factorization to QR factorization for solving least squares problems [20][7]. The use of the SVD as a feature reduction tool has been limited in remote sensing as the storage and processing are expensive, especially for large datasets such as entire images [8][13]. In the discipline of chemistry, van den Broek et al. [28] used the SVD to reveal the rank and reduce the data dimension of multivariate images.

This paper presents a feature reduction method using the singular value decomposition (SVD). The SVD is applied in a new way, drastically reducing the computer processing time and memory requirements and therefore making the SVD feasible for feature reduction in large datasets. This work examines various feature reduction methods for the pIGSCR classification algorithm, and demonstrates that the proposed SVD method significantly decreases classification execution times while not affecting classification accuracy, and the SVD outperforms some other commonly used feature reduction methods such as principal components analysis.

II. PIGSCR

pIGSCR [22] begins with the user inputting an image to be classified, training data specific to that image, and various parameters for the unsupervised classification and the spectral class homogeneity test. Clustering is performed on the input image, and the resulting spectral classes are analyzed for informational class homogeneity to ensure that only one informational class is present (with high probability) in each spectral class. If more than one informational class is present with significant probability, that spectral class will be rejected, and all pixels belonging to that class will have the opportunity to be reclustered. In order to conduct the homogeneity test, the training data must be analyzed to determine how many points from each informational class fall into each spectral class. For any training point t_i, the point's coordinates in the original image are used to determine to which spectral class it was assigned in the unsupervised classification. Each spectral class has an informational class count vector that has a length equal to the number of informational classes, and the component of the vector that corresponds to the informational class that was associated to t_i (by the analyst in the training phase) is incremented. After this process has been applied to each training point, the homogeneity test is conducted for each spectral class. The homogeneity test [21] is given as

$$N_k(1 - p_0) \geq 5 \text{ (spectral class must have minimum}$$
$$\text{number of pixels) and } Z > Z(\alpha) \text{ (homogeneity test)},$$

where

$Z \qquad = (\hat{p} - p_0 - (0.5/N_k))/\sqrt{p_0(1 - p_0)/N_k}$,
$\alpha \qquad$ is the type-I error rate,
$Z(\alpha) \quad$ is the value such that $P(Z \geq Z(\alpha)) = \alpha$ for a

2

$\quad\quad N(0,1)$ variable Z,

$\hat{p}\quad\quad = N_{maj}/N_k$,

$p_0\quad\quad$ is the user input threshold,

$N_k\quad\quad$ is the total number of training pixels in the spectral class, and

$N_{maj}\quad$ is the majority informational class count (count of training pixels belonging to the informational class with the highest count).

If a spectral class is determined to be pure ($Z > Z(\alpha)$), then the majority informational class is assigned to the spectral class, and each pixel that was assigned to the spectral class by clustering is recoded to the informational class value in the output classification image. Furthermore, each pixel belonging to a pure spectral class is removed from the input image for successive applications of clustering. The pure spectral class and its signature (a set of statistics for the class such as mean and covariance among bands) are kept for later processing. After all spectral classes have been processed, clustering (with the same number of clusters that was used in each previous iteration) is used again on the remaining pixels and the homogeneity test ensues. This iteration continues until no pixels belonging to impure classes remain in the input image, no pure classes were found during the previous iteration, or a maximum number of iterations has occurred. In this manner, all pixels belonging to impure spectral classes are reclustered during the subsequent iteration and each iteration operates on a proper subset of the pixels used in the previous iteration.

Once the iteration terminates, there is a set of pure spectral classes, a stacked clustering image recoded to informational class values with zeroes for background and unclassified pixels, a set of unclassified pixels remaining from the original image, and the original image. There are three options for output classification images at this stage, and any or all of them may be selected by the user. First, a decision rule may be applied to the original image using the pure spectral classes as training classes to produce a Decision Rule (DR) image. Second, all unclassified pixels may be recoded to a reserved value designating unclassified pixels (the number of informational classes plus one) and added to the stacked clustered image to produce an Iterative Stacked (IS) image. Finally a decision rule may be applied to only the unclassified pixels (using the pure spectral classes as training classes), and the resulting classification may be added to the stacked clustered image to produce an Iterative Stacked plus (IS+) image. The pIGSCR algorithm is given below.

Algorithm pIGSCR

User inputs an image, training dataset, maximum number
of iterations, α and p_0 for the homogeneity test,
and number of clusters to be created in each iteration.
The output that is produced includes a set of pure class
signatures that is used in the final decision rule,
and any of the following three classification images:
Decision Rule (DR), Iterative Stacked (IS), or
Iterative Stacked plus (IS+).
begin
   do until maximum number of iterations is reached,
   no pixels remain in original image,
   or no pure classes are found:
      Cluster remaining pixels in parallel.
      do for each point in the training data set:
         determine the spectral class assignment based on
         clustering, and increment the appropriate
         informational class count and spectral class count
      end do
      do for each cluster:

use the homogeneity test given above to
determine informational class assignment.

if a particular cluster is pure, **then**

    Add that cluster's signature to the set of pure
    class signatures and mask all vectors/pixels
    belonging to that class out of the input image.
    Recode pixels belonging to that spectral class
    in the cluster image to the value of the
    assigned informational class for IS image.

  end if

 end do

end do

Apply a parallel decision rule to the input image
using the pure signature set to create DR image.

do in parallel

  Recode all impure classes in the clustering output
  image to value reserved for impure classes and add
  this to the clustering image to produce IS image.

end do

Perform a parallel decision rule on only the impure
pixels and add them to the clustering image to produce
the IS+ image.

end

III. Feature Reduction Within pIGSCR

Previous work [22] showed that pIGSCR is significantly faster than IGSCR, and therefore input parameters may be selected using a factorial analysis to determine the specific combination of parameters that results in the highest classification accuracy. This approach has been extended to factorially determine which combination of bands would produce the best classification accuracy for a specific set of input parameters, dataset, and training dataset. Initial experimental runs on images containing six bands demonstrated that the most accurate classifications on all experimental images typically used all six bands. This approach was then further extended to experimentally determine the most accurate combination of bands for each unsupervised classification performed within each iteration of pIGSCR. Unfortunately, each iteration requires that $2^{bands} - 1$ classifications be performed to determine the most accurate combination of bands, and it is therefore not feasible for images with large numbers of bands. A classification performed on an image containing six bands would require that 63 individual unsupervised classifications be run per iteration, and an image with as few as ten bands would require over 1000 classifications be performed for each iteration of pIGSCR. Additional algorithmic improvements and additional processing power would be required to make this band selection algorithm a feasible option.

IV. Feature Reduction using the SVD

At a high level, the SVD reveals the minimum number of dimensions required to represent a matrix or linear transformation [18][27]. Often multi-dimensional data may be represented equivalently (or approximately so) in fewer dimensions due to redundancies in data. If a set of n-dimensional vectors all lie in a k-dimensional subspace, $k < n$, then each n-vector effectively has only k degrees of freedom, and can be uniquely described by k numbers. The natural correlations that occur in nature make the SVD a good candidate for feature reduction. Furthermore, if no reduction is possible, this will be shown by the magnitudes of the singular values revealed by the SVD.

The SVD of a linear transformation $A : R^n \longmapsto R^m$ is

$$A = U\Sigma V^t,$$

4

where $U : R^m \longmapsto R^m$ is orthogonal ($U^t U = I$), $\Sigma : R^n \longmapsto R^m$ is a diagonal matrix whose diagonal elements are called the singular values of A, and $V^t : R^n \longmapsto R^n$ is orthogonal as well. Each linear transformation in the SVD is expanded and shown as follows, assuming $m < n$:

$$
\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}
$$

$$
= \begin{pmatrix} u_{11} & \cdots & u_{1m} \\ \vdots & \ddots & \vdots \\ u_{m1} & \cdots & u_{mm} \end{pmatrix} \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{m1} & \sigma_{m2} & \cdots & \sigma_{mn} \end{pmatrix}
$$

$$
\times \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & v_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ v_{n1} & v_{n2} & \cdots & v_{nn} \end{pmatrix}.
$$

Since entries σ_{ij} when $j > m$ are all zero, the product ΣV^T will produce entries of zero for rows $m+1$ through n. The SVD may be rewritten as

$$
\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}
$$

$$
= \begin{pmatrix} u_{11} & \cdots & u_{1m} \\ \vdots & \ddots & \vdots \\ u_{m1} & \cdots & u_{mm} \end{pmatrix} \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_m \end{pmatrix}
$$

$$
\times \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ v_{m1} & v_{m2} & \cdots & v_{mn} \end{pmatrix}.
$$

This shows that a column $A_{\cdot i}$ of A, an m vector, can be expressed as a linear combination of the m basis vectors in U ($U_{\cdot 1}, U_{\cdot 2}, \ldots, U_{\cdot m}$), using the singular values in Σ ($\sigma_1, \sigma_2, \ldots, \sigma_m$), and the ith column $V_{\cdot i}^t$ in V^t. The diagonal elements of Σ are nonnegative, and can be ordered such that $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_m$. If some entries on the diagonal of Σ are zero, then for some k, $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_k > \sigma_{k+1} = \ldots = \sigma_m = 0$. Using the above logic that allowed a reduction in the number of rows in V^t from n to m, the number of columns in U can be reduced to k, the number of rows and columns of Σ can be reduced to k, and the number of rows in V^t can be reduced to k, yielding

$$
\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}
$$

$$
= \begin{pmatrix} u_{11} & \cdots & u_{1k} \\ \vdots & \ddots & \vdots \\ u_{m1} & \cdots & u_{mk} \end{pmatrix} \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_k \end{pmatrix}
$$

$$
\times \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ v_{k1} & v_{k2} & \cdots & v_{kn} \end{pmatrix}.
$$

An operation such as a classification that would be performed on the entire $m \times n$ matrix A can now be equivalently performed on the entire $k \times n$ matrix ΣV^t where $k < m$, resulting in a reduction in the number of bands present in each vector. For practical purposes, singular values may in fact be nonzero yet be sufficiently

close to zero to reduce the dimension of the data. The singular values $\sigma_{k+1}, \ldots, \sigma_m$ represent distances from the subspace spanned by $U_{\cdot 1}, \ldots, U_{\cdot k}$, and very small distances may not affect the operation that will be performed on the reduced data, such as classification. If none of the singular values on the diagonal are close to zero, then the data is already represented using as few dimensions as possible. However, this seems unlikely in the application of remote sensing as geographic data is naturally redundant.

Practically speaking, it would be necessary to think of a three-dimensional (pixel row, pixel column, data bands) image in two dimensions in order to take advantage of the feature reduction made possible by using the SVD. This would also be an expensive computation as the leading dimension of the matrix would be the number of bands, but the second dimension would be equal to the number of pixels in the image. One of the features of an SVD is that it reveals the basis vectors U that can be used to transform any vector from the original vector space (range of A) to the new vector space (range of $U\Sigma$). If a training dataset (columns of T) is truly representative of a particular image (columns of A), then the resulting SVD $T = \tilde{U}\tilde{\Sigma}\tilde{V}^t$ will also reveal a set of basis vectors for the range of A. Using this result, it is possible to perform the SVD on a training data matrix that is $m \times p$ in dimension, where p is the number of points in the training data set, and use the resulting SVD to transform A. In order to do this, it is necessary to project the columns of the matrix A onto the subspace spanned by the first k columns of $\tilde{U}$. This is accomplished by simply computing $(\tilde{U}_{\cdot 1}, \tilde{U}_{\cdot 2}, \ldots, \tilde{U}_{\cdot k})^t A$. This result can be arrived at algebraically. Assume that the $m \times p$ matrix T is a submatrix of the $m \times n$ matrix A, and that dim range A = dim range $T = k$. Then range $(U_{\cdot 1}, \ldots, U_{\cdot k})$ = range A = range T = range $(\tilde{U}_{\cdot 1}, \ldots, \tilde{U}_{\cdot k})$, which means each column of A can be reduced to its k coordinates with respect to the orthogonal basis $\tilde{U}_{\cdot 1}, \ldots, \tilde{U}_{\cdot k}$ of range $(\tilde{U}_{\cdot 1} \ldots, \tilde{U}_{\cdot k})$. These coordinates are given by multiplying by $(\tilde{U}_{\cdot 1}, \ldots, \tilde{U}_{\cdot k})^t$. After the original image has been transformed and the number of bands has been reduced, the pIGSCR classification algorithm is run with no modifications.

V. Data Description and Preparation

pIGSCR with the SVD or PCA was tested using Landsat Thematic Mapper (TM) and Enhanced Mapper (TM/ETM+) images (path 17/row 34) acquired on December 1, 1999, March 6, 2000, June 10, 2000. The images from 1999 and 2000 were layered together to form a composite multi seasonal image, and all images were registered to a rectified 2003 image with an RMSE of 1/5 a pixel or less. Using Leica Geosystems Erdas Imagine 8.7 software, registration was performed with 24 control points for each image pair, eight of which were randomly selected as check points. A first order transformation was used, and resampling was performed using nearest neighbor. Each of the three images used for the composite image was converted to reflectance in accordance with the Landsat 7 user's guide [17], and dark object subtraction was performed in ITT ENVI 4.2 using the band minimum method. The composite image, which will be referred to as VA17, contained 18 total bands. Figures 1, 2, and 3 show representative subsets of the VA17 image (bands 4, 3, and 2), the VA17 PCA (first three bands) and VA17 SVD (first three bands), respectively. Figure 4 shows the locations of the subset and VA17 in relation to each other and Virginia.

The training data for these images was created by the interpretation of point locations from a systematic, hexagonal grid over Virginia Base Mapping Program (VBMP) true color digital orthophotographs [29]. The data were collected in the Spring of 2002 at a scale of 1:4,800, and were subsequently resampled to a 1 meter spatial resolution. The data were used to perform a two-class classification, forest or non-forest. For the purpose of accuracy assessment, validation data in the form of point locations at the center of USDA Forest Service Forest Inventory and Analysis (FIA) ground plots [24][2] were used to validate the classifications. Only homogeneous FIA plots were used (either 100 percent forest or non-forest), and these plots were visited between 1997 and 2001.

A second dataset consisting of a multitemporal series of TM and ETM+ images from path 232/row 67 (1995-2002) acquired over Rondônia were used, each image registered to the 2001 image. The 2001 image was rectified by the US National Center for Earth Resources Observation & Science, and the remaining registrations were performed using Leica Geosystems Erdas Imagine 8.5 with at least 50 control points and 25 randomly selected check points. The root mean squared error was less than 1/3 of a pixel for each image registration, and nearest neighbor resampling was used [15][31]. This multitemporal image will be referred to as AM232. Figures 5, 6, and 7 show representative subsets of AM232 (bands 4, 3, and 2 from the 1998 image, corresponding to
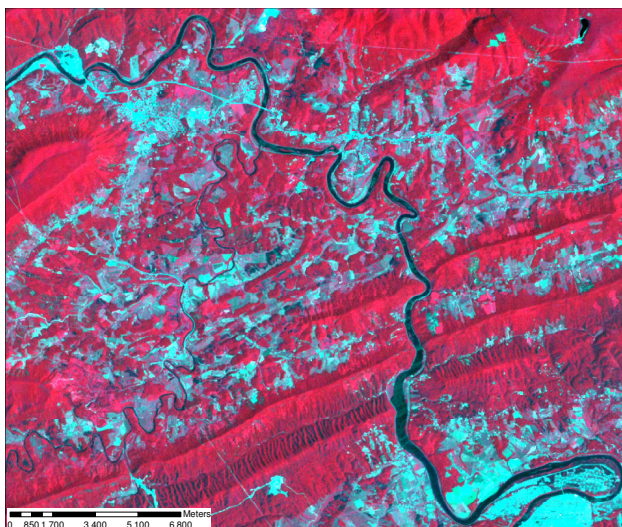
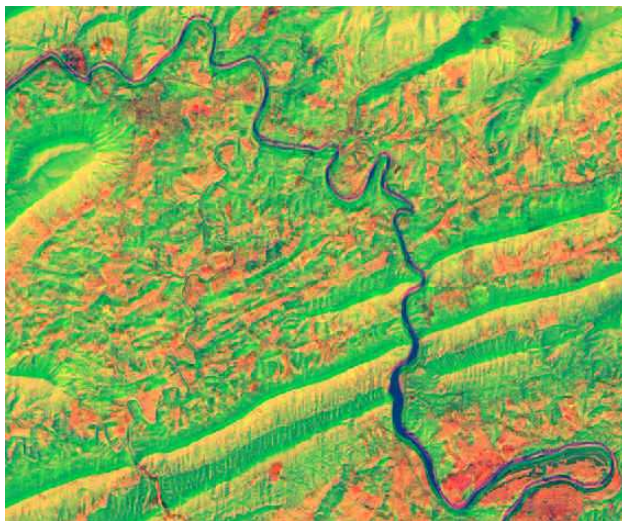Fig. 1. VA17 (bands 4, 3, and 2 shown) zoomed to a subset of interest.

Fig. 2. VA17 PCA (bands 1, 2, and 3 shown) zoomed to a subset of interest.

training data used for classification), AM232 PCA (first three bands), and AM232 SVD (first three bands). Figure 8 shows the locations of the subset and AM232 in relation to each other and Rondônia, Brazil.

Collecting training data for this region was a challenge, as described by [15][31]. The training data used for a forest/non-forest pIGSCR classification was collected using detailed interviews, Landsat imagery, and detailed maps of various farms shown in the images. At least 67 pixels for each of the two classes (forest and non-forest) were present, and these points were used as seed pixels for a region growing algorithm (Erdas Imagine 8.5). Because of the difficulties in acquiring training data, a total of 200 validation pixels were randomly selected from the training data. This ensured a representation of edge and mixed pixels, which is necessary in order to avoid inflated classification accuracies [23].

VI. Experimental Results and Discussion

A two class classification (forest, non-forest) using pIGSCR with 80 initial classes and a homogeneity
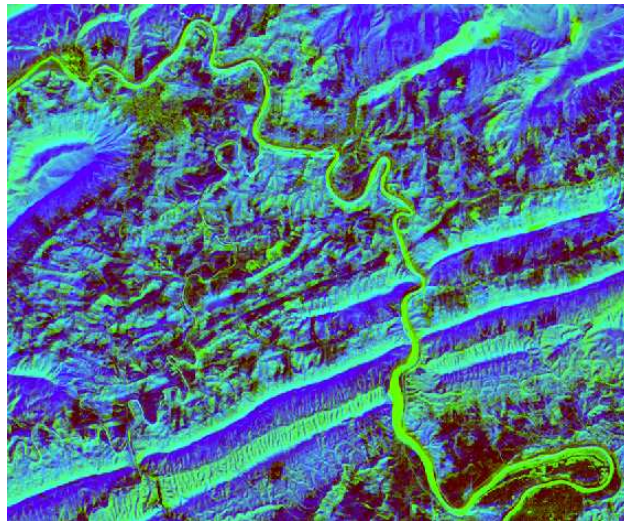
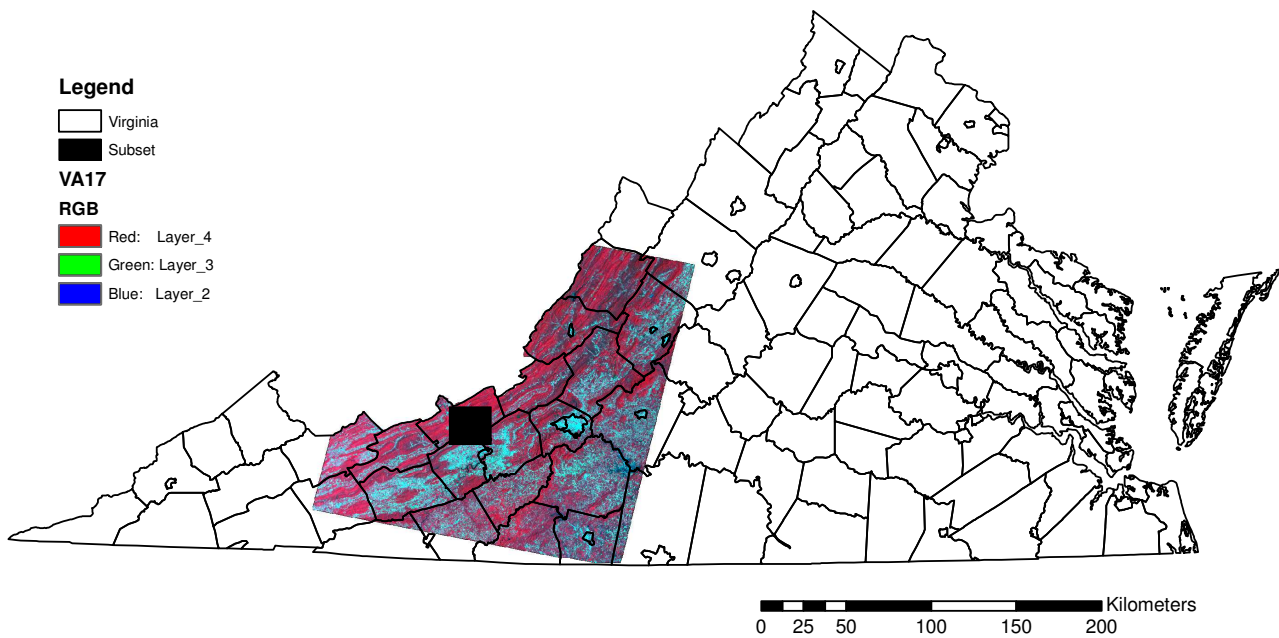Fig. 3. VA17 SVD (bands 1, 2, and 3 shown) zoomed to a subset of interest.

Fig. 4. Locations of Virginia, VA17 (Landsat TM path 17/row 34 acquired December 1, 1999), and zoomed subset shown in Figures 1–3 in relation to each other.

threshold of 80 percent was run on VA17 and AM232 and both of the above feature reduction methods were used. pIGSCR was applied to the composite VA17 image using all 18 bands, but applying pIGSCR to AM232 without
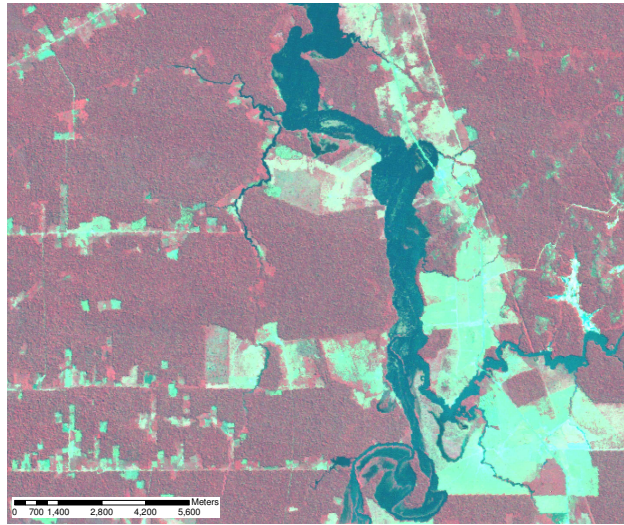
8

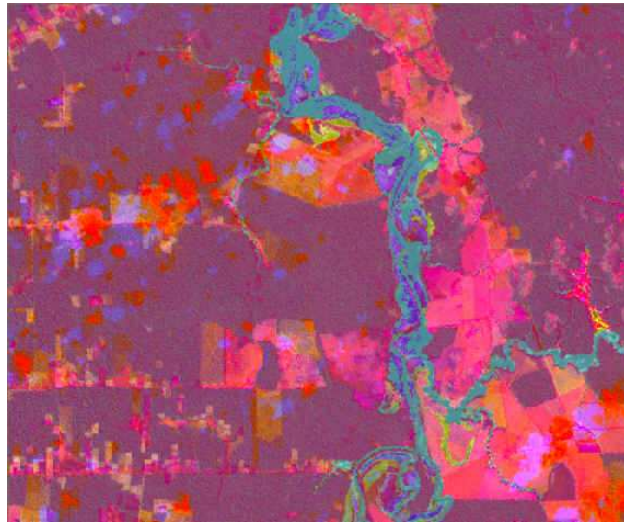Fig. 5. AM232 (bands 4, 3, and 2 of image acquired in 1998 shown) zoomed to a subset of interest.

Fig. 6. AM232 PCA (bands 1, 2, and 3 shown) zoomed to a subset of interest.

feature reduction was infeasible considering the number of bands (52) in AM232. The above algorithms were implemented using Fortran 95 [11] and LAPACK [1]. These classifications were run using an SGI Altix 3300 with 24GB of RAM and twelve 1.3 GHz Itanium processors.

pIGSCR was applied to the VA17 SVD image using as few as four and as many as fourteen bands. Each resulting SVD classification was compared to the corresponding classification of the entire eighteen band VA17 image using McNemar's test for statistical significance of the difference between the two classifications as described by Foody [10]. A chi-square distribution with one degree of freedom and $\alpha = .05$ (3.841) was used where

$$\chi^2 = \frac{(x_1 - x_2)^2}{x_1 + x_2},$$

with x_1 being the number of pixels correctly classified by the first classification but incorrectly classified by the second, and vice versa for x_2. Using this test, all classifications using five or more bands from the SVD image
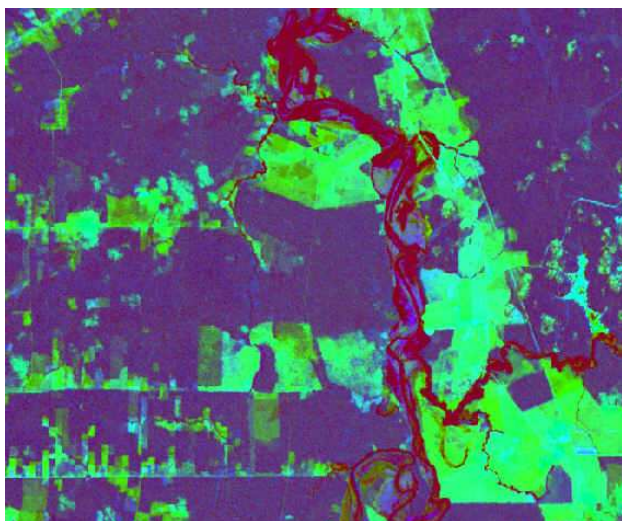
9

Fig. 7. AM232 SVD (bands 1, 2, and 3 shown) zoomed to a subset of interest.

were not statistically different from the classifications using all eighteen bands of VA17. Out of eleven tests run (four through fourteen bands) four tests using the SVD of VA17 were actually marginally higher in accuracy than the classification accuracy of 88.62% using all eighteen bands. Figures 9 and 10 show the resulting classification images of VA17 and VA17 SVD (using six bands), respectively, where green pixels were determined to be forest and tan pixels were determined to be non-forest. Although these two classifications are statistically the same classification (as determined by McNemar's test), it is clear that using the SVD image resulted in the correct classification of the large river that is prominent in this scene, while the classification using all eighteen bands of VA17 determined most of the river to be forest.
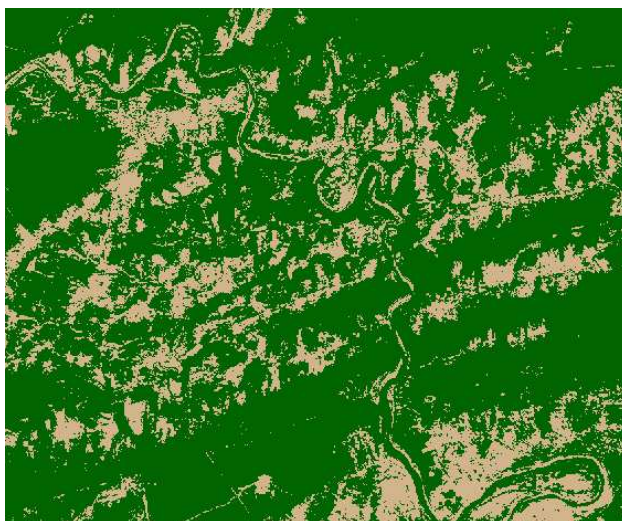

Fig. 9. VA17 DR classification, green = forest, tan = non-forest.

Using all eighteen bands resulted in a pIGSCR execution time of roughly 1650 seconds, more than five times as long as the quickest run of just over 300 seconds using five bands of the SVD image. In general, using more bands increases the execution time, however, execution time is also dependent on the number of iterations
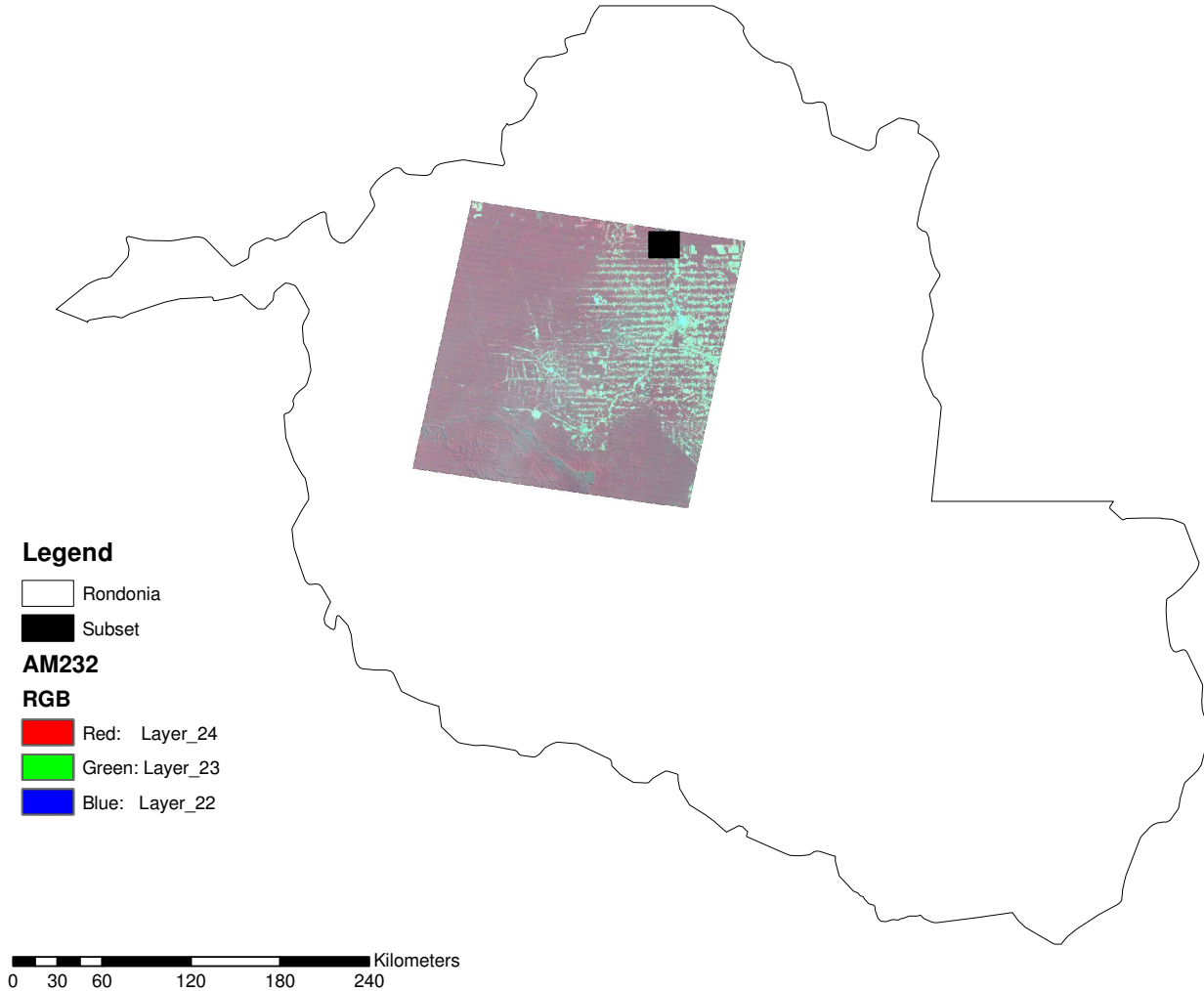
Fig. 8. Locations of Rondônia, AM232 (Landsat TM path 232/row 67 acquired 1998), and zoomed subset shown in Figures 5–7 in relation to each other.

performed during the iterative clustering and the number of pure classes used for the decision rule. It is therefore advantageous to use as few bands as possible when execution time is a great concern, and the accuracies resulting from using five through fourteen bands of the SVD image are consistently around 88%, both above and below the classification accuracy of VA17. These consistent accuracies indicate that the choice of k, the singular value cutoff, is not crucial past five bands.

A. SVD versus PCA

Table 1 summarizes all of the experimental runs of pIGSCR on both datasets, using the SVD or PCA as a means of feature reduction. The values in the table are computed by averaging the results of 11 runs for the Virginia data (four through fourteen bands used) and 18 runs for the Amazon data (three through twenty). Classification accuracies are reported for both pIGSCR output images that contain only two classes (the DR and IS+ images), corresponding to the validation set containing only two classes. Consider the classification accuracies
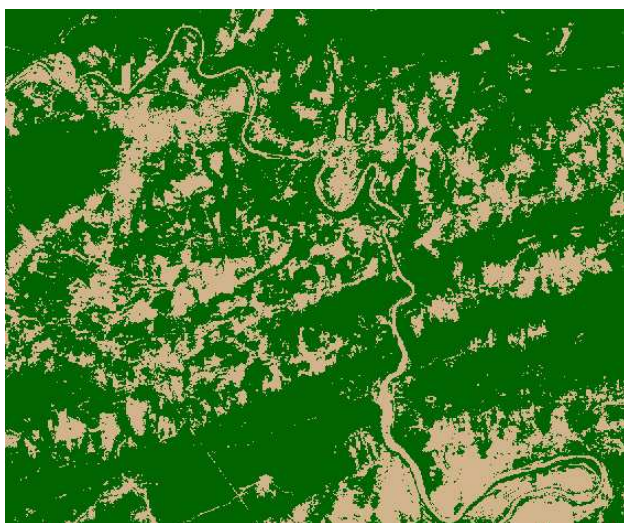
11

Fig. 10. VA17 SVD DR classification, green = forest, tan = non-forest.

TABLE 1
COMPARISON OF AVERAGES FOR
PCA AND SVD PIGSCR CLASSIFICATIONS

	SVD	PCA
VA DR accuracy	**88.18**	88.01
VA IS+ accuracy	**88.34**	88.16
VA execution time (secs.)	817	818
VA iterations	6.5	6.27
VA classes	78.2	77
AM DR accuracy	**92.75**	92.5
AM IS+ accuracy	**93.31**	92.94
AM execution time (secs.)	528	643
AM iterations	2.33	2.5
AM classes	18.33	20.11

listed, and notice that using the SVD to reduce the data dimension resulted in higher accuracies than using the PCA for both classification images (DR and IS+) using both datasets.

Considering now just the VA dataset, a comparison between the SVD and PCA using McNemar's test for statistically different classifications as described above showed that the classifications were consistently the same. All of the classification accuracies of the Virginia data are similar, but the accuracies for the SVD are more often higher (seven out of eleven cases). For this particular dataset and classification (forest/non-forest), the subspace spanned by the first few axes of decreasing variance is likely similar to the subspace revealed by the SVD, accounting for the lack of distinction between the two feature reduction methods.

Applying the same feature reduction methods and classification to the data from the Amazon region, however, demonstrated more separation between the two feature reduction methods. Table 2 lists all classification accuracies for the SVD and PCA AM232 images where the SVD and PCA reduction to the same number of bands resulted in statistically different classifications. In seven out of ten cases, the SVD feature reduction resulted in a higher classification accuracy. Figure 11 contains a graph comparing the classification accuracies of the two feature reduction methods for different numbers of bands kept (singular value or eigenvalue cutoff). Notice that the SVD accuracies are consistently higher than the PCA accuracies, and increasing the number of bands does not usually result in increased accuracy. Of particular interest are the cases where the number of bands is six or nine,

TABLE 2
CLASSIFICATION ACCURACIES OF STATISTICALLY
DIFFERENT CLASSIFICATIONS

bands	class	SVD accuracy	PCA accuracy
6	DR	**95**	91.5
6	IS+	**95.5**	92.5
7	DR	**93.5**	91
7	IS+	**94.5**	92
9	DR	92.5	**95.5**
9	IS+	92.5	**95.5**
10	IS+	**94**	91.5
12	DR	**94**	92
12	IS+	**94.5**	92.5
19	IS+	92	**94**

where one method drastically outperformed the other regarding accuracy. Figures 12–15 contain subsets of the classification results of AM232 for the SVD and PCA images, using either six or nine bands. Observing Figures 12 and 15, it appears that these images contain more misclassified forested regions due to cloud cover. Also notice that in Figure 14, the most accurate PCA classification, the river running vertically through the image has been incorrectly classified as forest, showing that this classification is not necessarily better than the corresponding SVD classification.

Fig. 12. AM232 PCA IS+ classification (6 bands), green = forest, tan = non-forest.

A more subtle advantage of using the SVD over PCA for feature reduction prior to pIGSCR classification is the unexpected execution time savings. Although the time savings in the Virginia classifications was minimal, the execution times for the classification applied to the SVD AM232 image are consistently shorter than execution times for the classification applied to the PCA AM232 image, as shown in Figure 16. An analysis of pIGSCR reveals that execution time can be affected by the size of the image, the number of iterations required, and the number of pure classes. Since the PCA and SVD produce images of the exact same size, the differences in execution time can be attributed to differences in the number of iterations and classes. Notice that the peaks in the execution time graph (Figure 16), correspond to peaks in Figure 17 (number of iterations) and peaks in Figure 18 (number of classes). A small number of iterations and classes does not appear to negatively impact

Fig. 13. AM232 SVD IS+ classification (6 bands), green = forest, tan = non-forest.

Fig. 14. AM232 PCA IS+ classification (9 bands), green = forest, tan = non-forest.

overall classification accuracy. For example, at six bands, classification accuracies were higher using the SVD for feature reduction (and were statistically different classifications than when using PCA), yet the classification on the PCA data required more iterations and resulted in more pure classes. Although this advantage seems tailored to pIGSCR, the implications extend to classification in general, both supervised and unsupervised, as pIGSCR is a hybrid classification algorithm and exhibits characteristics of both.

VII. Conclusion and Future Work

This paper presents a feature reduction method for remotely-sensed data using the singular value decomposition. This new feature reduction technique was applied to training data from two multitemporal datasets of Landsat TM/ETM+ imagery acquired over a forested area in Virginia, USA and Rondônia, Brazil. pIGSCR forest/non-forest classifications of the Virginia data were five times faster using SVD reduction without affecting

Fig. 15. AM232 SVD IS+ classification (9 bands), green = forest, tan = non-forest.
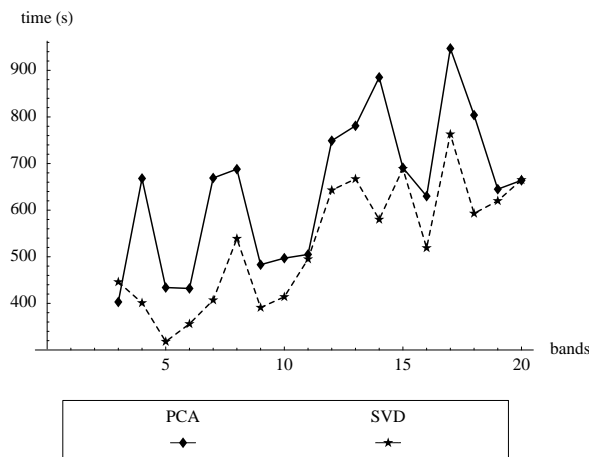

Fig. 16. Execution time for AM232.

the classification accuracy. Feature reduction using the SVD was also compared to feature reduction using principal components analysis (PCA) for both datasets. The highest average accuracies for the Virginia dataset (88.34%) and for the Rondônia dataset (93.31%) were achieved using the SVD. SVD-based feature reduction can yield statistically significantly better classifications than PCA.

This paper demonstrated the utility of SVD-based feature reduction on images containing fewer than 100 bands, however, SVD-based feature reduction of higher dimensional data (i.e., hyperspectral) should be evaluated in the future. Another area for future exploration with SVD-based feature reduction is classification with increased categorical specificity. With less variability between informational classes, feature reduction based on variance (such as with PCA) is likely to produce a lower quality classification on a reduced dimension image than feature reduction based on SVD. Finally, this paper demonstrates that applying the SVD to training data in order to reduce the dimension in an entire image produces good classification results. An ideal implementation would apply the SVD to the entire image to reveal the exact basis vectors, not an approximation derived from the training data. A parallel SVD implementation on multiple processors would allow the SVD to be performed on a large image in a reasonable amount of time, likely resulting in greater classification accuracy.
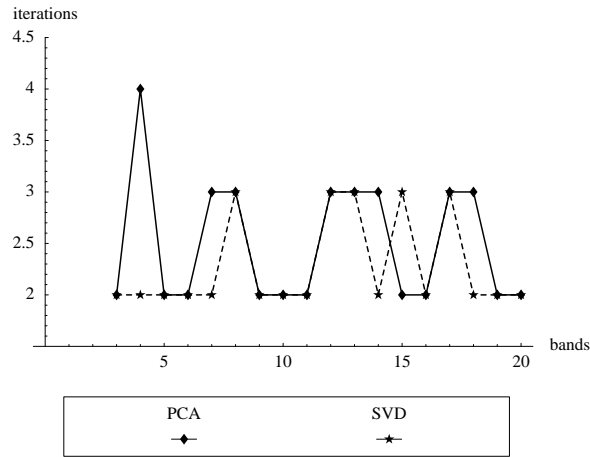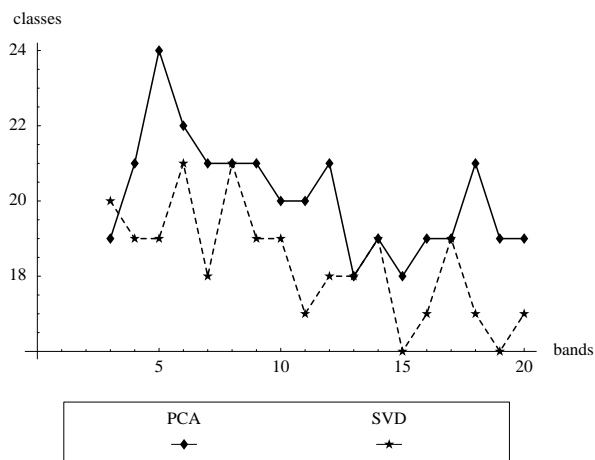
Fig. 17. Number of iterations for AM232.

Fig. 18. Number of pure classes for AM232.

REFERENCES

[1] E. Anderson, Z. Bai, C. Bischof, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, S. Ostrouchov, and D. Sorensen, *LAPACK Users' Guide*, SIAM, 1992.

[2] W. A. Bechtold and C. T. Scott, "The Forest Inventory and Analysis Plot Design," In The Enhanced Forest Inventory and Analysis Program – National Sampling Design and Estimation Procedures, W. A. Bechtold and P. L. Patterson, editors. USDA Forest Service Southern Research Station, General Technical Report SRS-80, pp. 27-42.

[3] J. Bigün, "Unsupervised feature reduction in image segmentation by local transforms," *Pattern Recognition Letters*, 14, 573-583, 1993.

[4] L. Bruzzone and D. F. Prieto, "Unsupervised Retraining of a Maximum Likelihood Classifier for the Analysis of Multitemporal Remote Sensing Images," *IEEE Trans. on Geosciences and Remote Sensing*, 39(2), 456-460, 2001.

[5] J. Byeungwoo and D. Landgrebe, "Partially Supervised Classification Using Weighted Unsupervised Clustering," *IEEE Trans. on Geosciences and Remote Sensing*, 37(2), 1073-1079, 1999.

[6] W. C. Chang, "On using principal components before separating a mixture of two multivariate normal distributions," *Applied Statistics*, 32, 267-275, 1983.

[7] C. Clark, A. F. Clark, "Spectral identification by singular value decomposition," *International Journal of Remote Sensing*, 19(12), 2317-2329, 1998.

[8] S. Danaher, E. O'Mongain, "Singular value decomposition in multispectral radiometry," *International Journal of Remote Sensing*, 13(9), 1771-1777, 1992.

[9] W. Eppler, "Canonical analysis for increased classification speed and channel selection," *IEEE Transactions on Geoscience Electronics*, 14, 26-33, 1976.

[10] G. M. Foody, "Thematic map comparison: evaluating the statistical significance of differences in classification accuracy," *Photogrammetric Engineering & Remote Sensing*, 705, 627-633, 2004.

[11] Fortran, *Information technology — Programming languages — Fortran —*, ISO/IEC 1539-1, 1997.

[12] A. A. Green, M. Berman, P. Switzer, M. D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Transactions on Geoscience and Remote Sensing*, 26(1), 65-74, 1988.

[13] S. Hattori, Y. Myint, "Automatic estimation of initial approximations of parameters for bundle adjustment," *Photogrammetric Engineering and Remote Sensing*, 61(7), 909-915, 1995.

[14] J. R. Jensen, *Introductory Digital Image Processing, A Remote Sensing Perspective*, Prentice Hall, 2005.

[15] K. A. Joseph, R. H. Wynne, J. O Browder, and J. B. Campbell, "Comparison of segment and pixel-based non-parametric land cover classification in the Brazilian Amazon using multitemporal Landsat TM/ETM+ imagery," *Photogrammetric Engineering & Remote Sensing*, in press.

[16] R. J. Kauth, G. S. Thomas, "The tasseled cap—a graphic description of the spectral-temporal development of agricultural crops, as seen by landsat," in *Proceedings, Symposium on Machine Processing of Remote Sensed Data*, West Lafayette, IN, 41-51, 1976.

[17] Landsat 7 Science Data Users Handbook (May 2006), *NASA. [ONLINE]*, Available: http://landsathandbook.gsfc.nasa.gov/handbook.html.

[18] C. Lawson, *Solving Least Squares Problems*, Prentice Hall, 1974.

[19] G. E. Lowitz, "Stability and dimensionality of Karhunen–Loeve multispectral image expansions," *Pattern Recognition*, 10, 359-363, 1978.

[20] M. Migliaccio, A. Gambardella, "Microwave radiometer spatial resolution enhancement," *IEEE Transactions on Geoscience and Remote Sensing*, 43(5), 1159-1169, 2005.

[21] R. F. Musy, R. H. Wynne, C. E. Blinn, J. A. Scrivani, and R. E. McRoberts, "Automated forest area estimation via iterative guided spectral class rejection," *Photogrammetric Engineering & Remote Sensing*, 72(8),949-960, 2006.

[22] R. D. Phillips, L. T. Watson, and R. H. Wynne, "Hybrid image classification and parameter selection using a shared memory parallel algorithm," *Computers & Geosciences*, doi:10.1016/j.cageo.2006.10.014, 2007.

[23] R. L. Powell, N. Matzke, C. De Souza, M. Clark, I. Numata, L. L. Hess, D. A. Roberts, and M. Clark, "Sources of error in accuracy assessment of thematic land cover maps in the Brazilian Amazon," *Remote Sensing of Environment*, 90, 221-234, 2004.

[24] G. A. Reams, W. D. Smith, M. H. Hansen, W. A. Bechtold, R. A. Roesch, and G. G. Molsen, "The Forest Inventory and Analysis Sampling Frame," In The Enhanced Forest Inventory and Analysis Program – National Sampling Design and Estimation Procedures, W. A. Bechtold and P. L. Patterson, editors. USDA Forest Service Southern Research Station, General Technical Report SRS-80, pp 11-26.

[25] J. A. Richards and X. Jia, *Remote Sensing Digital Image Analysis*, Springer, 1999.

[26] J. San Miguel-Ayanz and G. S. Biging, "Comparison of single-stage and multi-stage classification approaches for cover type mapping with TM and SPOT data," *Remote Sensing of Environment*, 59, 92-104, 1997.

[27] G. Strang, *Linear Algebra and Its Applications*, Academic Press, 1976.

[28] W. H. A. M. van den Broek, D. Wienke, W. J. Meissen, C. W. A. de Crom, L. Buydens, "Identification of plastics among nonplastics in mixed waste by remote sensing near–infrared imaging spectroscopy. 1. Image improvement and analysis by singular value decomposition," *Analytical Chemistry*, 67, 3753-3759, 1995.

[29] VBMP, *Virginia Base Mapping Program (VBMP) Digital Orthophotography Project*, Virginia Geographic Information Network (VGIN) Advisory Board, Richmond, Virginia, 2003.

[30] J. P. Wayman, R. H. Wynne, J. A. Scrivani, and G. A. Reams, "Landsat TM-based forest area estimation using Iterative Guided Spectral Class Rejection," *Photogrammetric Engineering & Remote Sensing*, 67(10), 1155-1166, 2001.

[31] R. H. Wynne, K. A. Joseph, J. O Browder, P. M. Summers, "Comparing farmer–based and satellite–derived deforestation estimates in the Amazon basin using a hybrid classifier," *International Journal of Remote Sensing*, in press.