Organ Analysis and Classification using Principal Component and Linear Discriminant Analysis

William H. Horsthemke
Daniela S. Raicu
DePaul University

ABSTRACT

Texture analysis and classification of soft tissues in Computed Tomography (CT) images recently advanced with a new approach that disambiguates the checkboard problem where two distinctly different patterns produce identical co-occurrence matrices, but this method quadruples the size of the feature space. The feature space size problem is exacerbated by the use of varying sized texture operators for improving boundary segmentation. Dimensionality reduction motivates this investigation into systematic analysis of the power of feature categories (Haralick descriptors, distance, and direction) to differentiate between soft tissues.

The within-organ variance explained by the individual components of feature categories offers a ranking of their potential power for between-organ discrimination. This paper introduces a technique for combining the Principal Component Analysis (PCA) results to compare and visualize the explanatory power of features with varying window sizes. We found that 1) the two Haralick features Cluster Tendency and Contrast contribute the most; 2) as distance increases, its contribution to overall variance decreases; and 3) direction is unimportant.

We also evaluated the proposed technique with respect to its classification power. Linear Discriminant Analysis (LDA) and Decision Tree (DT) were used to produce two classification models based on the reduced data set. We found that using PCA either fails to improve or markedly degrades the classification performance of LDA as well as of the DT model. Though feature extraction for classification shows no promise, the proposed technique offers a systematic mechanism to compare feature reduction strategies for varying window sizes as well as other measurement techniques.

Introduction

Texture analysis continues to evolve as a feature measurement technique to analyze and classify image data. Recent advances in the analysis of co-occurrence matrices has produced a technique to disambiguate the checkerboard problem where two distinctly different patterns produce identical co-occurrence matrices [1]. However, this technique along with traditional texture analysis techniques generates an abundance of features thereby limiting the scope of further analysis due to computational constraints. Methods to reduce the number of features becomes a goal of this research and this paper examines the role of principal components analysis in achieving feature reduction in texture analysis.

Feature Reduction

Two main strategies serve to reduce the number of attributes used in classifying, separating, or clustering data and generally correspond to prior knowledge about the class (type, identity) of the instances in the dataset. Without class labels, only unsupervised techniques such as principal components analysis (PCA) can be applied. If prior knowledge is available about the class of each instance, numerous feature selection techniques are available.

Feature Selection

Evaluation and search form the main tasks of feature selection algorithms. These methods use the class label of the dataset instances to assess the role or effectiveness of the attribute (or set of attributes) in separating or distinguishing the classes.

Evaluation defines a criteria and decision threshold for inclusion or rejection of a feature. These methods measure the usefulness of a feature in classification such as information gain in decision trees or R^2 in multiple regression prediction and reject or accept based upon some prior or adaptive criteria. Some methods perform a full classification algorithm to assess the feature (the wrapper approach [6]). Search methods form the basis for identifying the next feature for evaluation and range from best fit (greedy) to best-first [6] which retains a priority queue of unsearched features. In addition these approaches consider forward (bottom-up) aggregation strategies where features are added one by one or by groups of features as well as backward selection (top-down) where non-useful features are removed.

Feature Extraction using Principal Components Analysis

Principal Components Analysis (PCA) transforms the dataset and offers a method to identify and rank the attributes according to the amount of variation within the data explained by each attribute. PCA uses the covariance between the attributes to transform the attribute space of the dataset to produce a new space where the attributes are uncorrelated. After first computing the covariance matrix for the dataset, the eigenvectors (principal components) are extracted to form a new linear transformation of the original attribute space. After ranking each eigenvector (principal component) for the amount of dataset variation they explain, the top ranking eigenvectors are selected to represent the entire dataset.

The resulting eigenvector representation forms an attribute loadings (weighting) vector which can be used for modeling the data. The attributes with large loadings (weights) contribute more to the principal component of the data, the attributes with lower loadings (weights) can be considered noise. Using both the loadings (weights of the original features) and the amount of variance explained by the principal components, the importance of the individual features can be compared and ranked.

Methodology

This paper examines four (4) datasets representing four (4) window sizes for the texture measurement operator as described in a previous paper [1]. Their paper applied the novel DDP (direction vs. displacement pairs) approach to manually segmented computed tomography images of eight organ types (aorta, trabecular bone, fat, kidney, liver, lung,

muscle, and spleen) from a set of five (5) patients. Prior to applying the DDP co-occurrence texture analysis methodology, [1] enhanced the contrast of the soft tissues by applying a histogram-based clipping/stretching technique described in [2]. This adaptive approach offers more resolution on the smaller intensity band of soft tissues while permitting an overall reduced set of intensity bins.

A pseudo-stratified random sample selected a set of 20,000 instances from each window size dataset (approximately 10% of the overall data) using a uniform random sampling to extract approximately equal numbers of patients and organs for the classification datasets (or equal number of patients for the organ only datasets). Only restraints on computational memory forced the motivation for data reduction. The method has been termed pseudo since there exists slight variation in the number of instances of organ measurements per patient. Additional research is needed to verify this does not introduce a bias in the results.

Classification employed the LDA methods from the Discrim [8] toolkit using leave-one-out testing and the J48 method (a C4.5 variant) from the Weka [9] toolkit using a 67% training and 33% testing split dataset approach. Accuracy was measured as the percentage of correct classifications per number of instances of that class.

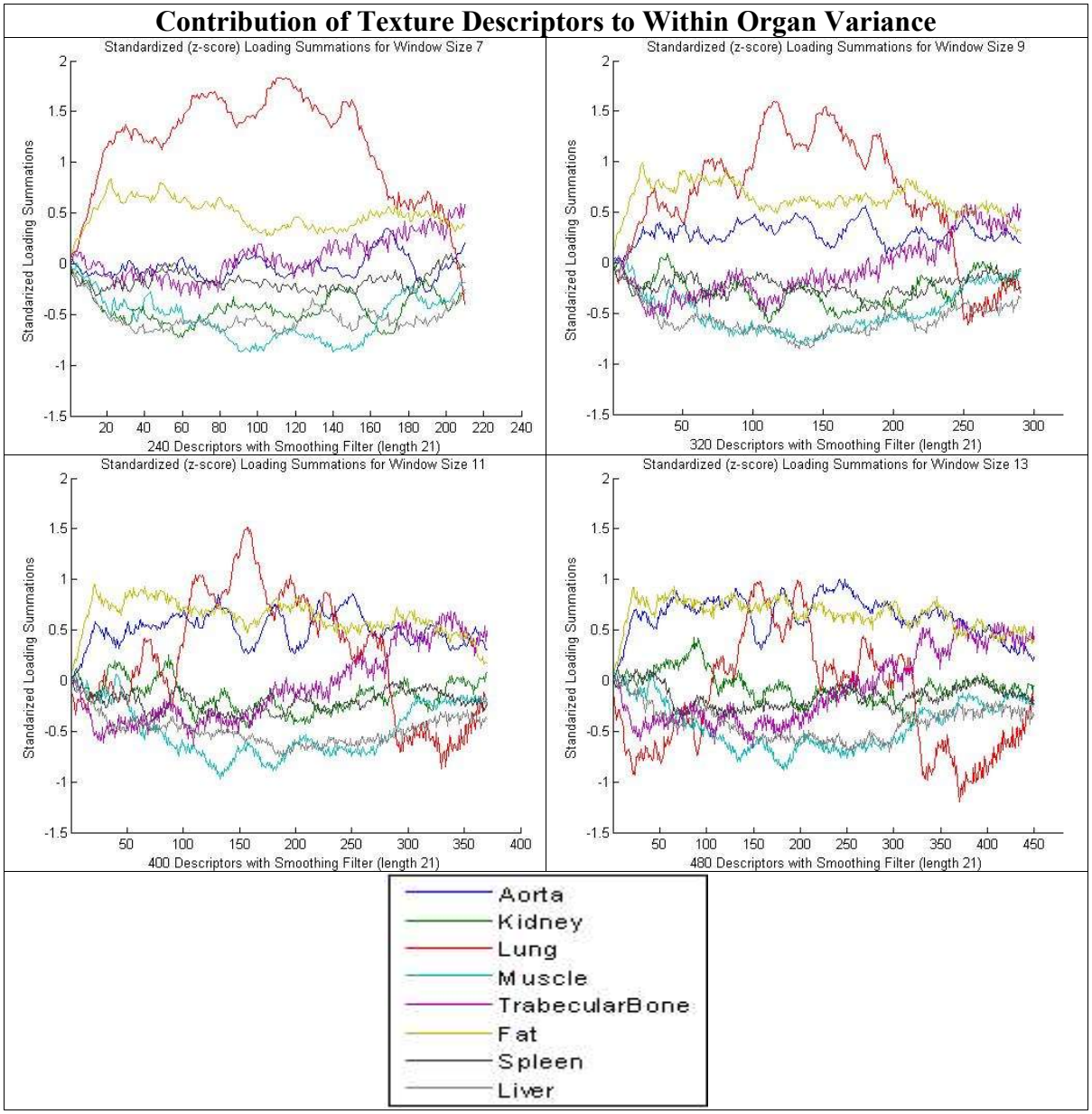
LDA class visualizations (from the Discrim toolkit) plotted the distribution of features using their mean and covariance, thus producing covariance ellipsoids in the feature space.

Principal Components Feature Analysis for Individual Organs

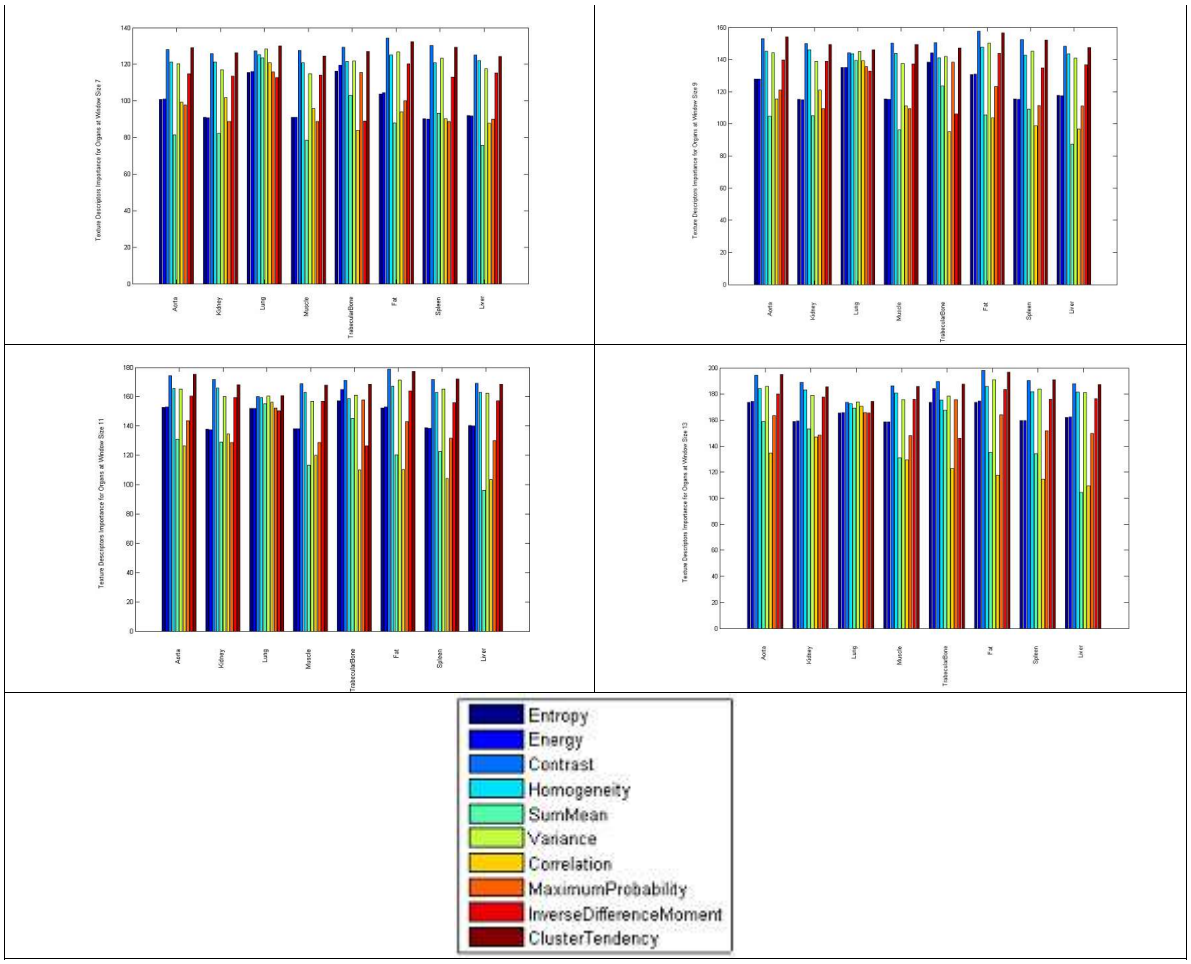
Using organ-specific partitions of the overall dataset, a principal components analysis of each organ was examined to determine the importance of the texture descriptor (e.g. SumMean or Contrast, etc.), direction (0,45,90,135), and distance (e.g. 1,2, 3 ...). Assessing and comparing these principal components (PC) presents a challenge since each PC contains as many features as the original dataset and an examination of numerous might be necessary to understand how the variance of the organ is represented by the features. Additionally, some organs might require more PCs to represent an commensurate amount of explanation of variance. A technique introduced in this paper uses only the weight of the feature coefficients in the PC (not the sign or direction), then multiplies this absolute coefficient value by the percentage of explanation represented by the entire PC. This explanation weighted coefficient represents the individual PC. This representation of an individual PC easily extends to a summation of all similarly weighted PCs to represent the PCA of the organ. Though this representation folds the contrasting weights of the individual features, it offers a visualization mechanism to understand the relative importance of the individual features.

Comparing PCA across different window sizes presents an additional challenge since the magnitude range of the explained-weighted summations of PCs differences across window sizes. To permit a cross-window size evaluation, the summations were z-score normalized. Additional filtering (21-point moving average) was applied to smooth the variation along the descriptor axis for increased clarity of the organ differences within

each window size. The following graph illustrates the differences between organs for the window sizes (7, 9, 11, & 13).

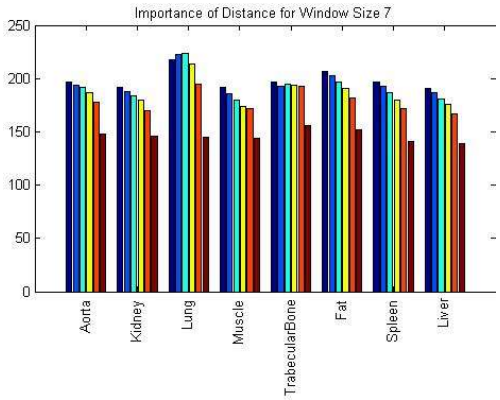


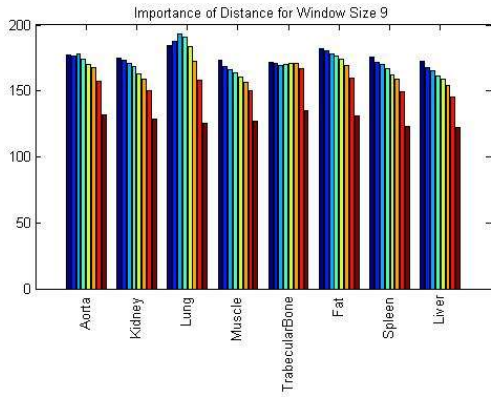
Contribution of Texture Descriptors to Within Organ Variance

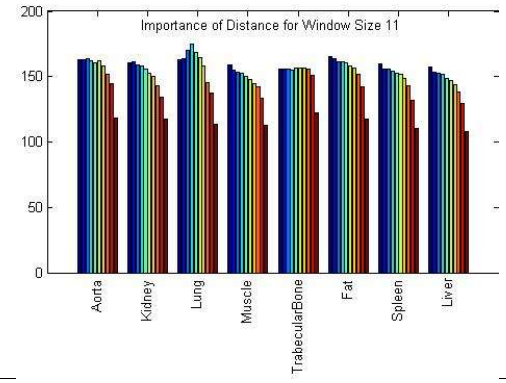


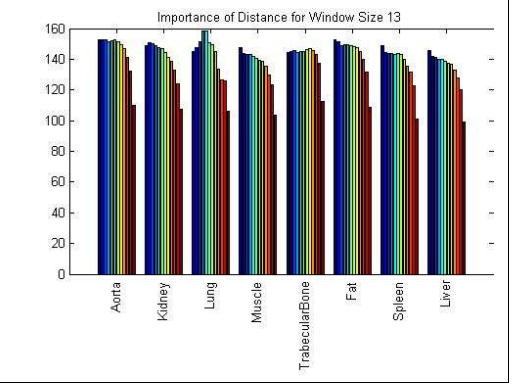
The above bar charts compare the 10 different texture descriptors for each organ and 4 window sizes to identify their contribution to variance within the organ. Two feature descriptors dominate for each organ and window size: Cluster Tendency and Contrast. Two feature descriptors tend to contribute little to the within organ variance: SumMean and Correlation. In general, the top 5 texture descriptors are Contrast, Homogeneity, Variance, Inverse Difference Moment, and Cluster Tendency.

Contribution of Distance to Within Organ Variance



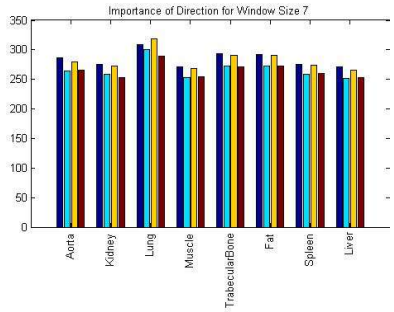


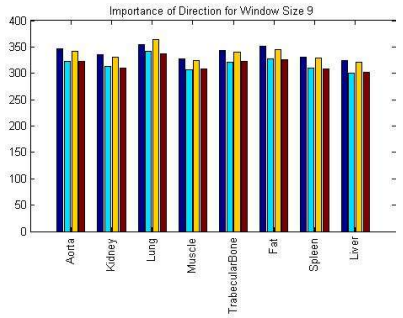


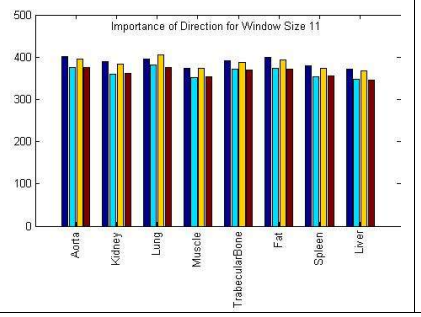


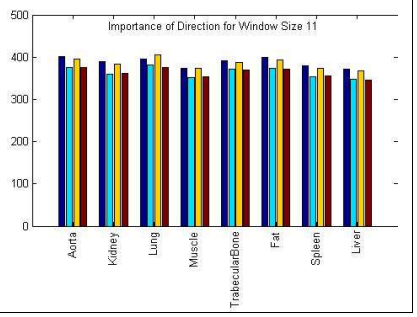
The effect of distance on the variance contribution is illustrated in the above bar charts and consistently shows a declining importance of the texture features as the distance increases (the bars decrease within each organ group for all organs and all window sizes). Only the Lung and Trabecular Bone differs from this trend; Lung increases slightly till the middle range of distances (e.g. Lung increases from 1 to 3 then decreases from 3 to 6 for window size 7). Trabecular Bond appears flat across the distances. Though the contribution effect of distance decreases with increasing distance, the rate of change for the largest distance is noticeably larger than the others. This decline in explanation at the largest distance merits further study.

Contribution of Directions to Within Organ Variance



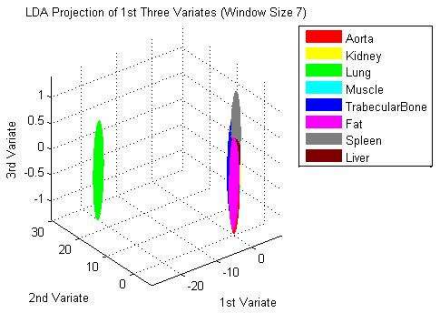


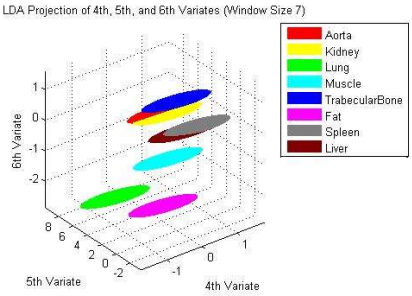




The directions of 0 and 90 degrees contribute approximately 10% more to the within organ variance. Though consistent across organs and window size, this small difference does not appear significant enough to exploit for organ analysis or classification.

Visualization of Between Organ Variance using LDA





The above visualization of the organs within the first six (6) LDA variates shows marked separation of the lung from the other organs but marked overlap between the spleen and liver as well as the aorta, kidney, and trabecular bone. This degree of separation between the organs using only 6 LDA variates suggests LDA classification will perform well.

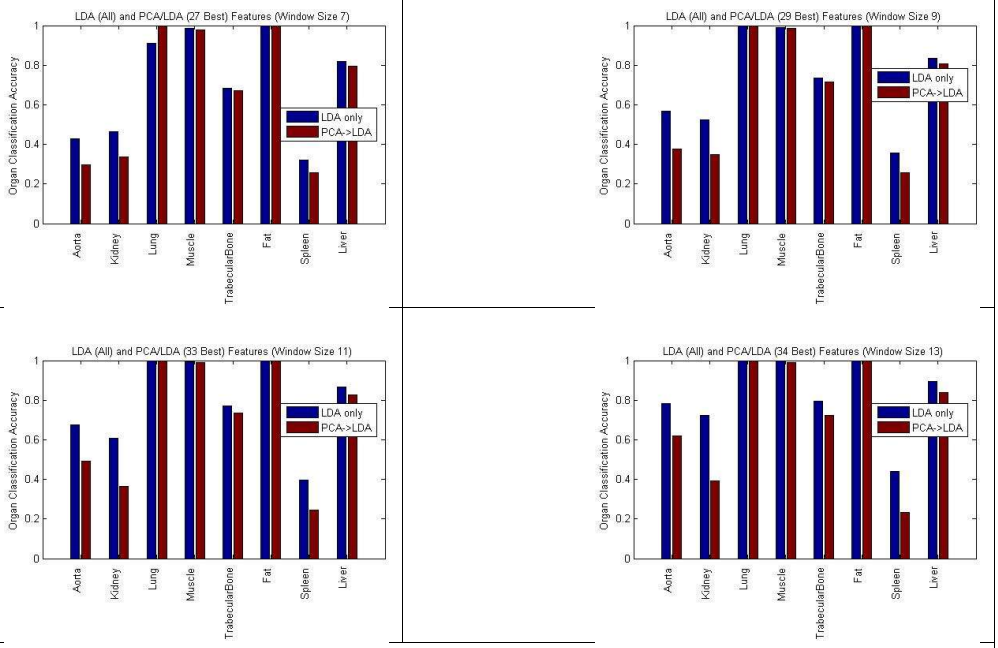
Classification of Organs

Evaluating the effect of PCA feature extraction on organ classification performance represents the major goal of this paper and two classification algorithms were chosen. The first, Linear Discriminant Analysis (LDA), attempts to separate the organs using a linear combination of the original features; this method closely relates to PCA which also finds a linear combination of the features but not to separate, since it does not use (or know) the actual class label; PCA only attempts to explain dataset variance using dimensions formed by linear combinations of the original features. The second technique employs a methodology unrelated to PCA but selects features according to their discriminatory power.

Decision trees (CART, C4.5) offers the other classification methodology for comparison of the pre- and post-treatment effect of PCA. Decision trees evaluate features

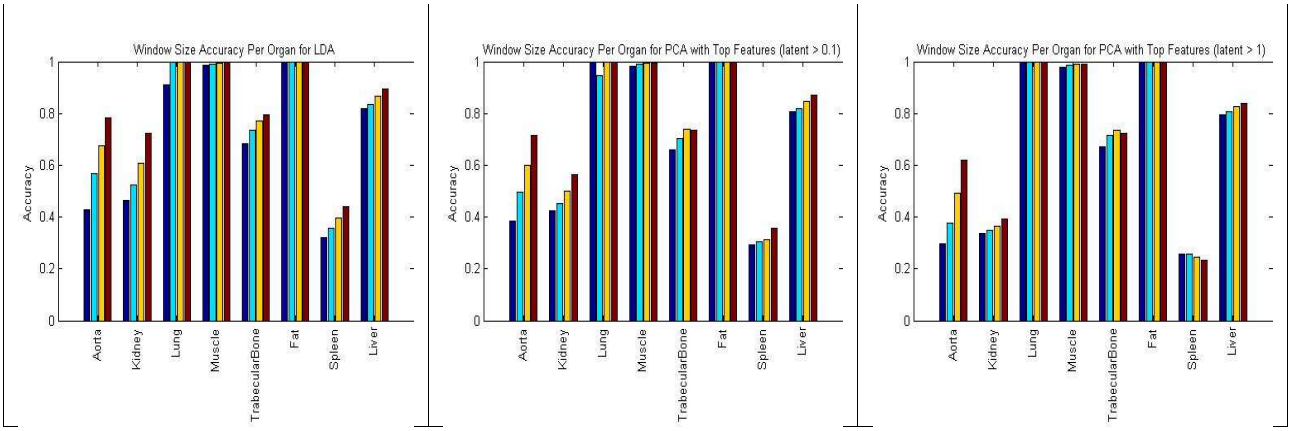
individually, not in combination though dependent upon the prior selection, according to their effectiveness in separating classes. Successful classification of organs using decision trees was recently reported [5].

LDA with all features compared to LDA using PCA extracted "best" features

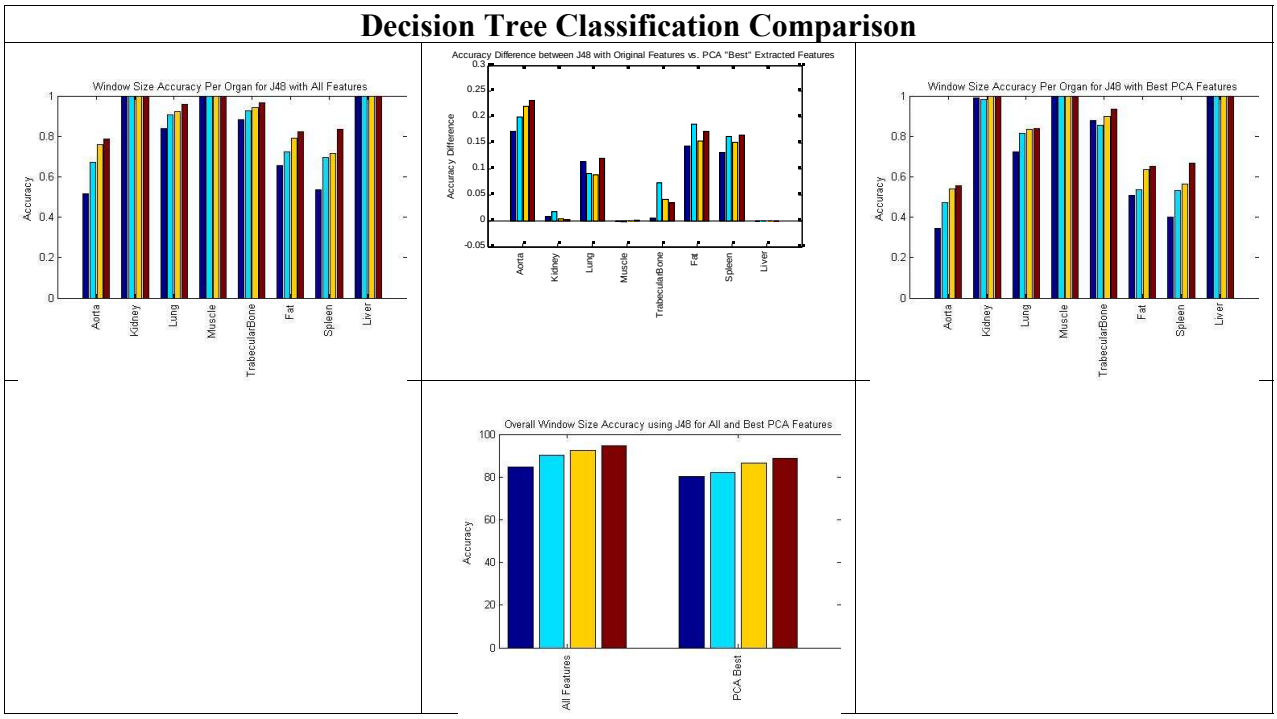


The above bar charts illustrate the classification performance of LDA using all texture features compared to LDA using features extracted from the "best" principal components. The "Kaiser" criteria selects only those PCs which explain more that one standard deviation of the variance in the original data and have an explained value greater than 1 (using z-score normalized data). The LDA classification performance decreased consistently across all organs and window sizes when the features were extracted using the best PCs versus using the original features. A further analysis was performed using effectively all PCs (any PC with an explanation above 0) and follows in the next bar charts.

LDA and PCA->LDA Classification Comparison



The above bar charts compare the performance of LDA using all original features, all PCs extracted by PCA, and only the Kaiser "best" PCs across all window sizes. Two major points can be observed using this comparative analysis. The most salient observation illustrates the improvement in classification performance with increasing window size. This is consistent across all feature selection (extraction) methods with only one exception for Spleen where only the "best" features are chosen (the Spleen performance actually decreases with increasing window size). A second observation compares the all versus "best" PCs selection criteria and illustrates a slight performance reduction if only the "best" PCs are chosen. Though this performance reduction is small, it indicates that PCA does not improve the features by its extraction technique.



The performance of LDA does not improve with PCA feature extraction but the close relation between LDA and PCA suggests a need to employ a classification technique unrelated to PCA. Classification and regression trees, as implemented in the C4.5

derivative (J48), offers such an unrelated classification approach. As shown in the above bar charts, the performance of J48 degrades with the use of PCA extracted features for most organs and window sizes. The classification accuracy of Aorta, Fat, and Spleen are reduced by over 15%, though Liver, Kidney, and Muscle are generally unchanged. This offers a confirmation of the ineffectiveness of PCA to improve the classification performance of organs using DPP texture analysis.

Conclusion

Principal components analysis offers insight into how much variance is captured by various aspects of the texture descriptors, but using the PCA extracted features does not improve the classification performance of either LDA or J48 (decision tree). The insight obtained by PCA suggest some texture descriptors might not capture enough variance to be justify their computation and the surprising result of the reduced importance of the texture features measured across the entire window where the distance is one less than the size of the window.

Future Work

Further feature selection approaches shall be investigated, especially the direct selection of features based upon the results of per organ feature analysis using the explained-weighted loadings. Classification performance using various reduced feature sets will be examined using both LDA and J48 classifiers. Selection of feature sets will also be investigated such as Correlation-based Feature Selection [12].

Supervised PCA approaches have been proposed [10,11] and will be investigated.

Acknowledgment of Support

This material is based upon work supported by the National Science Foundation under Grant No. 0453456.

References

[1] R. Susomboon, D.S. Raicu, and J.D. Furst, "Pixel-Based Texture Classification of Tissues in Computed Tomography", CTI Research Symposium, April 2006
[2] R. Lerman, D. S. Raicu, and J. D. Furst, “Contrast enhancement of soft tissues in Computed Tomography images”, SPIE Medical Imaging, 2006.
[3] R.M. Haralick, K. Shanmugam, and I. Dinstein, “Textural Features for Image Classification”, IEEE Trans. on Systems, Man, and Cybernetics, vol. Smc-3, no.6, pp. 610-621, 1973.
[4] M. Kalinin, D.S. Raicu, J. Furst, and D.S. Channin, “A classification Approach for anatomical regions segmentation”, IEEE Int. Conf. on Image Processing, 2005.
[5] D. Channin, D. S. Raicu, J. D. Furst, D. H. Xu, L. Lilly, and C. Limpsangsri, "Classification of Tissues in Computed Tomography using Decision Trees", The 90th Scientific Assembly and Annual Meeting of Radiology Society of North America (RSNA04), 2004.

[6] Kohavi, R., & John, G. (1997). Wrappers for feature subset selection. Artificial Intelligence, 97, 273-324.
[7] Kiefte, Michael. Discrim. (Matlab Toolkit) 1999. MATLAB File Exchange
[8] B. D. Ripley (1996) Pattern Classification and Neural Networks. Cambridge
[9] Ian H. Witten and Eibe Frank (2005) "Data Mining: Practical machine learning tools and techniques", 2nd Edition, Morgan Kaufmann, San Francisco, 2005.
[10] Danijel Skocaj, Martina Uray, Ales Leonardis, Horst Bischof (2006) “Why to Combine Reconstructive and Discriminative Information for Incremental Subspace Learning”, Proceedings of the Computer Vision Winter Workshop 2006
[11] Myoung Soo Park; Jin Hee Na; Jin Young Choi, "PCA-based feature extraction using class information", *IEEE International Conference on Systems, Man and Cybernetics*, pp.341-345, Jun. 2005
[12] Hall, M. (1999). Correlation based feature selection for machine learning. Doctoral dissertation, University of Waikato, Dept. of Computer Science