Feature Extraction and Classification for EO-1 Hyperspectral Imagery

Becky Castano   Steve Chien

Contributors: Ben Cichy, Sam Prentice, Nghia Tang

Jet Propulsion Laboratory, California Institute of Technology

September 11, 2003

Hyperspectral data compression via onboard data prioritization

- Opportunity
  - Hyperspectral instruments collect high volumes of data
  - The instruments are capable of collecting more data than available downlink bandwidth permits returning to Earth
- Response
  - **Better utilize limited downlink resource by enabling the return of the most important science data**
- Example criteria for determining important science data
  - Data quality control (e.g. do not downlink cloud covered images)
  - Feature detection
  - Change detection
  - Summary

Classification

Goal: Develop an effective classifier suitable for onboard processing

EO-1 hyperspectal image classification problem:

- Discriminate between multiple classes of land cover in hyperspectral images of ground scenes of the Earth
- Primary example presented in this work consists of distinguishing between clouds, ice, land & water in Hyperion imagery gathered by EO-1

Outline

- Background
  - Hyperspectral imagery and the Hyperion instrument
- Feature extraction
  - Dimension reduction
  - Spatial features
- Classification
  - Support Vector Machines (SVM)
  - Linear Discriminant Analysis (LDA)
  - Manually developed classifier
- Results
- Conclusions and Future Work

Hyperion Instrument

One of three advanced land imaging instruments on EO-1

Parameters	Hyperion
# Spectral Bands	220
Spectral Range	0.4-2.4 μm
Spatial Resolution	30 m
Image Size	256 x 6926 pixels ~ 7 x 200 km

Feature Extraction

- Objective: extract features from the data that can be used to discriminate between classes
- Feature Components:
  - Pixel based (purely spectral)
    - Use all 220 bands
    - Reduce dimensionality of data
  - Region Based (spatial and spectral)
    - use pixel neighborhood regions
- Considerations:
  - Classifier performance can degrade with excessive redundant or irrelevant information
  - Limited onboard processing capabilities restrict access to only 12 of the 220 bands (any 12 bands may be selected)

Subset of Bands

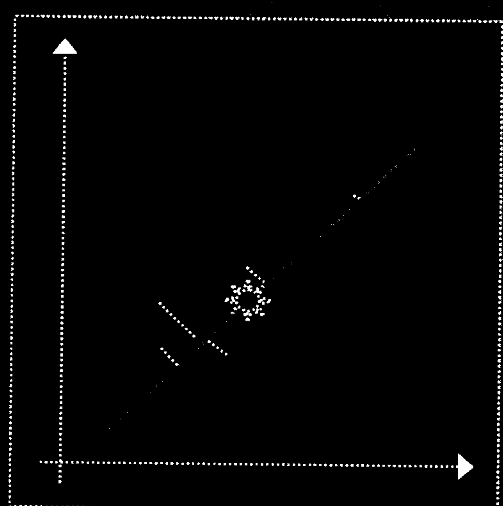
Reduce dimensionality by using bands manually selected by scientists

- **Hand-6** – 6 bands chosen to discriminate snow, water, ice and land
- **Hand-11** – 11 bands, consisting of Hand-6 and 5 additional bands chosen for cloud detection purposes

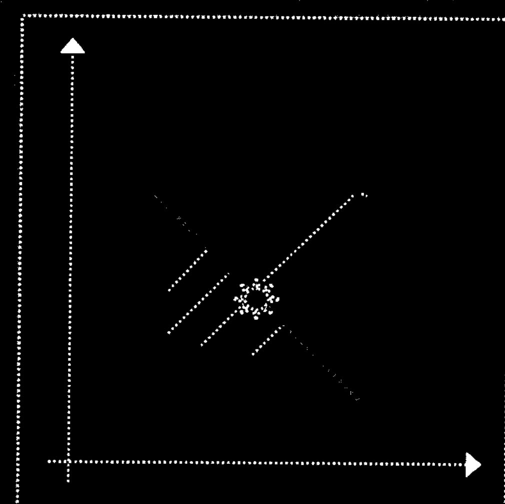
		Band #
Hand-11	Hand-6	15
		25
		85
		90
		110
		145
		20
		31
		51
		123
		150

Dimension Reduction

- Principal Component Analysis (PCA)
  - finds a linear subspace that results in smallest mean-square error between the feature vectors and their projections
  - identifies subspace that is efficient for representation



Best 1-D Dimension for representing the data



Worst 1-D Dimension for representing the data

Simple Spatial Features

- 3x3 Pixel Neighborhood
  - X is the current pixel
  - A, B, ..., H are the surrounding pixels
  - Raw pixels are vectors representing 220 bands. In experiments we use the **Hand-6** and **Hand-11** values of the pixel.

	x coordinate		
y coordinate	A	B	C
	D	X	E
	F	G	H

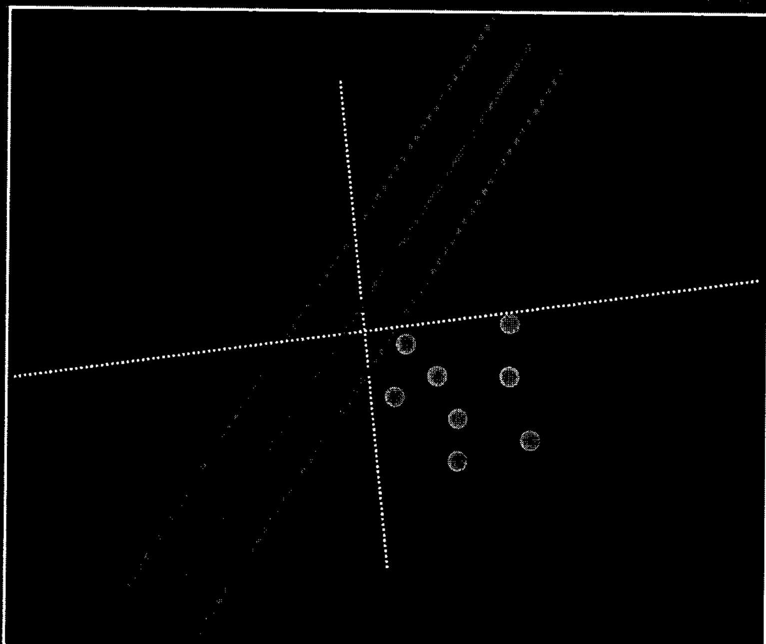
Feature Vector:

A	B	C	D		E	F	G	H

Classification

- Binary Classification
  - Support Vector Machine (SVM)
  - Use SVM to train land cover classifiers (e.g. cloud classifier identifies clouds vs. [ice land water])
- Multiple Class Classification
  - Linear Discriminant Analysis
  - Manually developed classifier

Support Vector Machines (SVM)

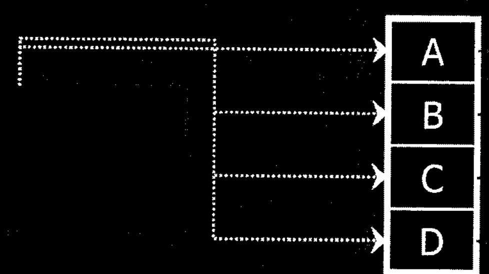


The turquoise lines represent the optimal hyperplane and its corresponding margin for these data. White lines are non-optimal hyperplanes.

- Creates classifier that separates two distinct classes
- Maps the data into a high dimensional space and finds a hyperplane that separates data from two classes
- The optimal hyperplane maximizes the margin (the distance between the hyperplane and nearest points from the two classes)
- Kernels used:
  - linear
  - Gaussian radial basis function (rbf)
  - normalized polynomial (npoly)

Merging Pairwise Classifications

- Run each image pixel through 4 binary classifiers (cloud vs. all, ice vs. all, ...) to obtain result
- More than 95% of pixels are uniquely classified
- However, in some cases a pixel may be classified as more than one class (4.6% have 2 conflicts, 0.0006% have 3 conflicts)



A	A	A	B
A	A	B	B
A / D	D	D	B

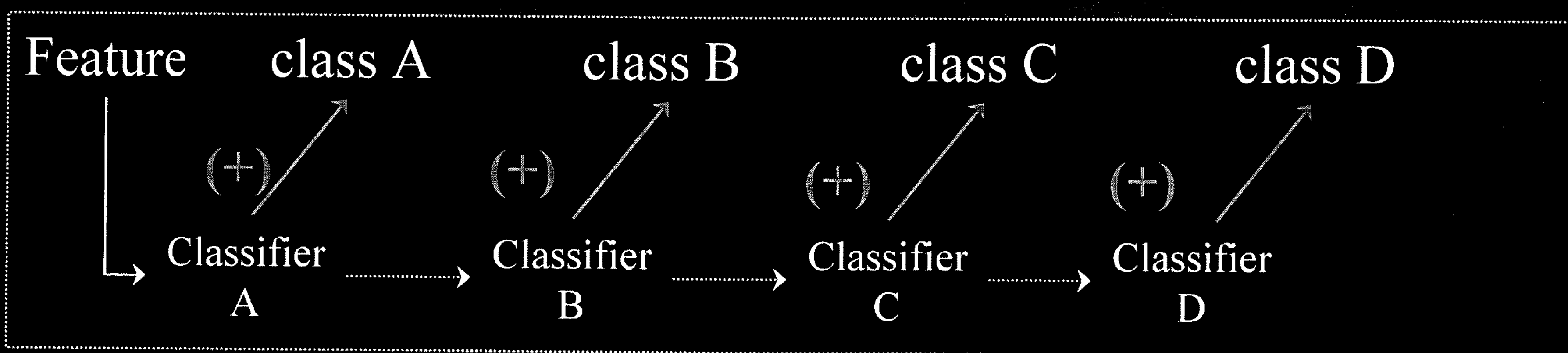
Image pixels   Binary classifiers   Classification results

How do we resolve conflicts?

Solution 1: Ordered Merge

- Create a pipeline of pairwise classifications with the most accurate being first.
- Go through the pipeline in the following manner:

1. Run classifier A
2. Remove all points classified as A
3. Run classifier B on remaining test data
4. Repeat steps 2 and 3 for classifiers C and D



= negative classification (+) = positive classification

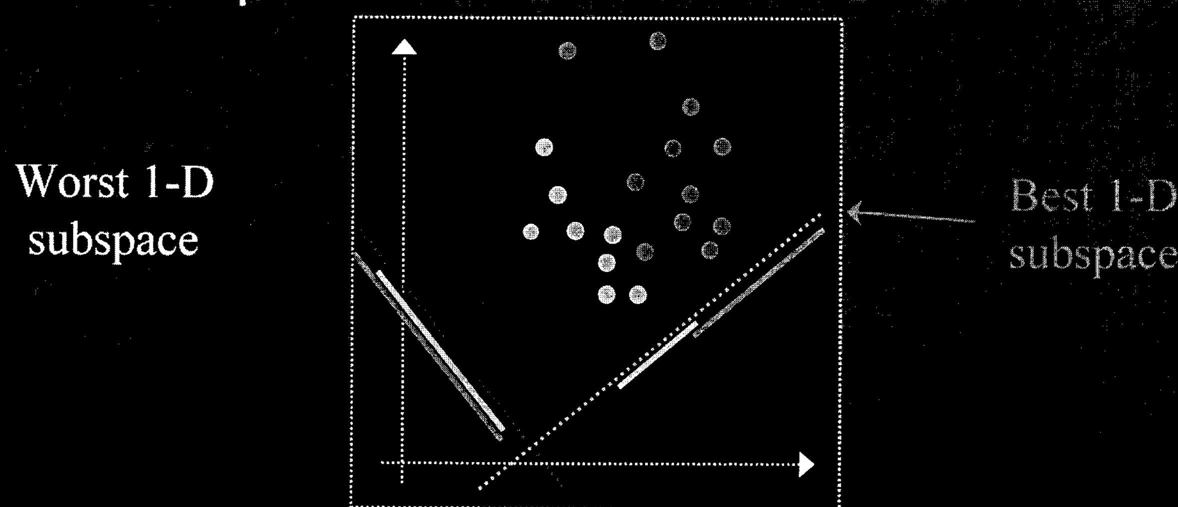
Solution 2: Pairwise 1-on-1

- If a pixel is classified as 2 different classes, reclassify with a pairwise classifier trained specifically for those 2 classes.

- Example:
  - Pixel is classified as *ice* and *water* (the pixel is given a (+) classification by the *ice vs. all* and *water vs. all* classifiers)
  - Run pixel through an *ice vs. water* classifier and use the result

Linear Discriminant Analysis (LDA)

- Finds a linear subspace that maximizes class separability among the projections of feature vectors
- Separability criterion: ratio of between-class scatter and within-class scatter
- Seeks subspace that is efficient for discrimination



SWIL Classifier

- **S**now, **W**ater, **I**ce, and **L**and classifier
- Features are band values and ratios of bands
- Features and thresholds hand-selected by scientists
- Sequential evaluation similar to ordered merge
- Computationally inexpensive

Thermal Detection

- Detects volcanoes and fires
- Manually selected bands and thresholds
- Feature is difference of band values
- Scheduled as first classifier to fly
  - January 2004

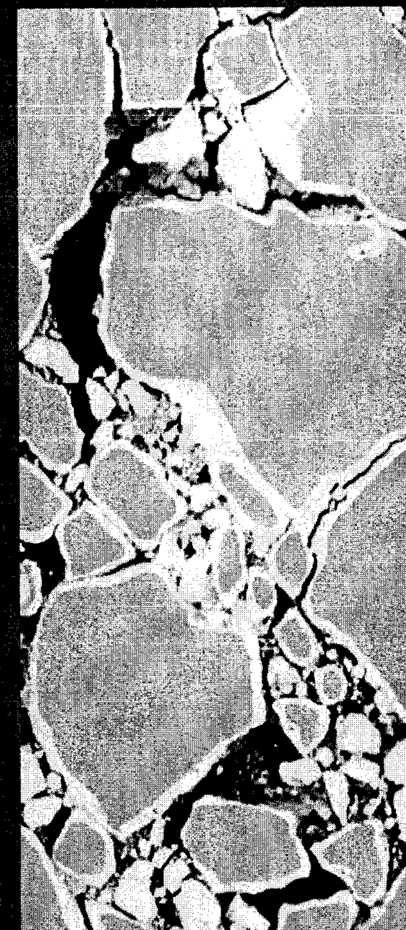
Training Data

- Manually label data by looking at a visible band
- Extract random samplings of labeled images for each class (cloud, ice, land & water)
- Combine each class into merged set, taking equal #'s of samples from each image

Each SVM classifier is trained on 3000 positive and 3000 negative examples

Ex: training a *cloud* classifier

- use 3000 cloud, 1000 each of ice, land, water

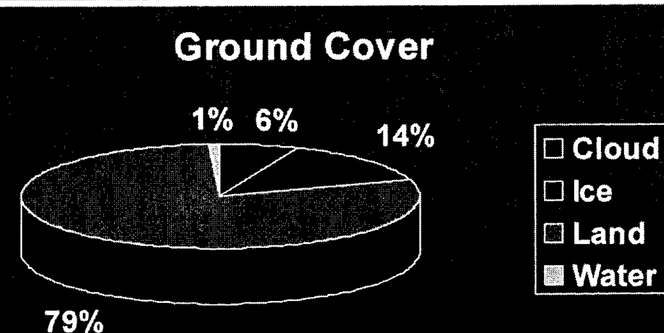


Example training image. Green regions are part of the ice mask for this image.

Test Data

Labeled portions of 14 Level 1 Hyperion images.

	cloud	ice	land	water
# pixels	655,445	1,389,939	8,132,354	90,383
	10,268,121 test pixels (~40% of total pixels in image set)			



Feature Type Results

Accuracy of one vs. all classifiers trained with different features. Best results were achieved with the radial basis function (rbf) kernel. The rbf is used in all classifiers shown below.

% of pixels correctly classified

Feature Type

Classifier	All	Hand-6	Hand-11	PCA	3x3 Hand-6
cloud v. all	95.8	95.3	96.2	95.2	**96.5**
ice v. all	99.0	99.0	98.8	**99.5**	97.8
land v. all	96.9	97.0	97.2	96.6	**97.6**
water v. all	97.5	96.1	95.9	**97.9**	97.2

Classifier Comparison

Comparison of SVM, LDA and SWIL classifiers

Accuracy

Classifier

	Overall Accuracy % (including cloud)	Only Water, Ice, and Land %
Ordered Merge Hand-6 rbf	95.8	97.8
1-on-1 Pairwise Merge Hand-6 rbf	96.2	98.1
1-on-1 Pairwise Merge All rbf	94.4	98.2
LDA using All bands	88.0	89.0
SWIL	92.6	98.9

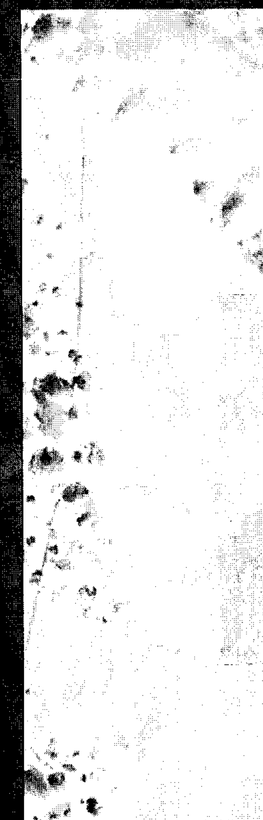
Examples of Failure

Example of cloud & ice

Very difficult to discriminate *cloud* from *ice*

	All	PCA
cloud classifier		
EO1H...PP_cloud	99.8	99.8
EO1H...PP_ice	**3.6**	**2.7**
EO1H...PP_water	100.0	100.0
EO1H...PY_ice	100.0	100.0

PY – has no problem finding ice



PP - hard to discriminate cloud from ice

Summary

- Automated SVM system and manually developed SWIL classifier have similar accuracy for land, water and ice on current test data set
- The SVM outperforms the SWIL classifier for cloud classification
- Comparisons to stand-alone manually developed cloud classifier not yet performed.
- Issues and tradeoffs:
  - development effort (hand vs. automated)
  - number of computations at run time
  - level of accuracy
  - simple & specific vs. more complex and general

Plans for FY 04

- Test classifier performance on L0 (uncalibrated) data which will be available onboard
- Automate feature selection
- Spatial/spectral features
  - Use more information than just neighbor pixel values

Backup

Classification for Prioritization

- Classification
  - Detect events
  - Change detection
- Content-driven compression modes
  - Data selection
    - Data frame
    - Region within data frame
  - Summarization