

SUPPLEMENTARY MATERIALS

The following pages contain supplementary information and materials related to my application package for the for the Earth Science Tenure Track Assistant Professor in the Department of Environmental and Ocean Sciences at the University of San Diego:

- Copy of Official Graduate Transcripts from the University of California, Berkeley
- Sample lecture/lab slides from courses taught at MIT, the University of California, Berkeley, and at a Summer School on Water Research
- Copy of student teaching evaluations from two courses taught at the University of California, Berkeley in 2011 and 2009

Massachusetts Institute of Technology 77 Massachusetts Avenue, Building 54–1025 Cambridge, Massachusetts 02139–4307 Phone:(617) 253–2578, Fax: (617) 253–2578 Email: dinob@mit.edu http://eapsweb.mit.edu/people/dinob

GRADUATE TRANSCRIPT

SAMPLE LECTURE SLIDES

The following pages contain sample lecture/lab slides from courses taught at the Massachusetts Institute of Technology, the University of California, Berkeley, and from one of three shourt courses given at a Summer School on Water Research:

- Guest lecturer, MIT, Fall 2013, EAPS 12.163 "Geomorphology". Instructor: Prof. Taylor Perron. Lecture on shallow landslide prediction.
- Instructor, UCB, Spring 2011, EPS209 "Matlab Applications in Earth Science", co-developed and cotaught with Prof. Burkhard Militzer. Lecture and lab on image segmentation, lecture and lab on GIS, Mapping Toolbox, and WMS.
- Lecturer, 2nd International Summer School on Water Research, "Landslide modeling and Early Warning Systems", Summer 2013. Lecture on support Vector Machine Classification for Earth Science applications.

Shallow landslides

Outline:

Slope stability

- Mohr-Coulomb failure
- Simple theory
- Implications for failure
- Hydrological model
	- Darcy's law
- Simple model for wetness **Shalstab**
	- Theory
	- Application

If we have time:

- Debris flows potential
- Landslide size
	- 3-D stability model
	- Search algorithm

[Photo: Bill Dietrich]

Shallow landslides

[Photo: Bill Dietrich]

Shallow landslides

Shallow landslides

Shallow landslides

168号線地すべり

Deep seated landslides

Source ESRI, GOT, USGS, FEMA

The Clumnicle

Ridge and valley topography In the Oregon Coast Range

5 Key Questions

Where?

When?

How big?

How far?

Mass gain or loss?

Laguna Beach, California, 1998

What we know: when?

What we know: how big?

5 Key Questions

Where?

When?

How big?

How far?

Mass gain or loss?

Laguna Beach, California, 1998

Ridge and valley topography In the Oregon Coast Range

Example: processes shaping the Oregon Coast Range

Shallow Landslides

Mohr-Coulomb Failure

Infinite Slope

Implications:

- Soil does not need to be fully saturated for failure!
- Four cases:
	- "Unconditionally stable": failure requires h/z > 1, i.e.: $\tan \theta \leq \tan \phi (1 (\rho_{s} \rho_{w}))$
	- "Unconditionally unstable": slope > than friction angle $(h/z < 0)$
	- "Stable": h/z < *R.H.S.*
	- "Unstable": h/z > *R.H.S.*

Instability Condition

- Mapped landslides

- LiDAR data

Coos Bay Field Site

- Piezometers
- Rain gauges
- Weirs
- Sprinkling experiments
- LiDAR data
- Mapped landslides

Coos Bay Instrumentation

Coos Bay Rain

[Photos: Dietrich]

h/z for Slope Instability

Hydrology: Darcy's Law

Hydrology: Darcy's Law

Steady State Subsurface Flow

Saturation Subsurface Flow

Topographic Index

Drainage Area

Simple model for relative saturation

Relative Saturation

Shalstab

Shalstab: a compact simple model

Controls on instability: Slope vs. Area

Relative instability potential

Controls on instability: Slope vs. Area

Drainage area dominates, until the slopes become steep!

[Dietrich and Montgomery, 1998]

Shalstab: Performance

Controls on instability: Resolution

High resolution vs. Inghitesolution vs.
low resolution data: Number of captured landslides Percent of landscape affected

[Dietrich et al., 2001]

5 Key Questions

Where? How big? How far? When? Mass gain or loss?

Laguna Beach, California, 1998

Infinite Slope Framework

Infinite Slope Framework

Landslide Size

Lateral reinforcement matters

3-D Slope stability model

3-D Slope stability model

A discrete landscape model

A discrete landscape model

- Discretize landscape into *grid* of cells
- Associate each cell with a node in a *graph*
- *Nodes:* landscape cells annotated by driving forces
- *Edges:* resistive forces between neighboring grid cells

Results

density function

a.

Captures observed distributions!

10

Homework

Final Project

Think about it soon:

- **If will help us fine-tune the class**
- We can point you towards useful resources
- You will run out of time

Individual:

- Exception: collaboration (2 people max) for a harder problem
- Proposal (< 1 page) stating objectives and methods by March $9th$
- 10 minute presentation (plus questions) on April 13th
- Note: a little more expected from a group effort

Current proposals:

- Quantify constituents in concrete from electron microscopy images
- **Extract vegetation characteristics from satellite imagery**
- Model hillslope hydrology

Review

- Transforming images to binary (thresholding, indexing, filtering)
- Regions from binary images (**bwconncomp()**)
- Measuring regions (**regionprops()**)

Today: Image Segmentation (part 1)

Algae Patches:

- Selecting a Region of Interest (ROI) in an image
- **Alternative metric for measuring similarity**
- Pixel-based segmentation

Victoria Crater (Mars):

- **Images as watersheds**
- The Watershed Transform
- Median filtering

San Pablo Reservoir:

- **"** "Texture" Filters
- **Morphological operations**

Algae Identification and rgb2ind()

- Using **rgb2ind()** with 4 colors (recolored for more contrast)
- Our task: we only want to label *coherent* regions that are *similar* to a few user-defined algae types and discard the rest.

Algae Identification and rgb2ind()

- Note: **rgb2ind()** assigns *every* pixel to one of *n* colors
- Our task: we only want to label *coherent* regions that are *similar* to a few user-defined algae types and discard the rest.
- How do we *define* "similar"?
- How do we *select* a "color representation" for each type?

Selecting a Region of Interest (ROI)

- The function **roipoly()** allows for manual region selection
- If returns a mask containing 1 inside the region and 0 outside
- **How do we describe this region?**
- What is representative of this region?

Picking a Representative Color

- Simplest: pick the *mean* color
- Then *similarity* implies being within a certain (Euclidean) distance in RGB space
- The pixels that are within a sphere centered at the mean color are *similar*
- **Is this too simplified?**

Picking a Representative Color

Smoothing

- **Some of the variance is noise, and we'd like coherent regions**
- We can remove it by averaging neighboring pixels: **myFilter = fspecial('average', 10); smoothImg = imfilter(myImage, myFilter);**

Simple Supervised Learning Algorithm

- Read the image
- Smooth the image
- Pick the representative color patches
- Learn mean and covariance for each patch

Simple Supervised Learning Algorithm

- Compute the mininum *Mahalanobis distance* of all pixels to any *mean* color
- Assign each pixel to index of closest color

- Label each pixel that is too distant as belonging to no class (index $= 0$)
- **Function of two or three parameters:**
	-
	- Number of colors Threshold of distance Filter size (optional)
-

Result

 A reasonable and simple start, but still very crude **How could we improve the procedure?**

Regions from "continuous" images

So far:

 Pixel-based segmentation methods

Other approaches:

- Global: "Topography" "Morphology"
- **Local:** Edge detection Topology

Cardiac MRI image Gradient image

Victoria Crater (Mars)

Dunes in Victoria Crater

Watersheds

"That area of land, a bounded hydrologic system, within which all living things are inextricably linked by their common water course and where, as humans settled, simple logic demanded that they become part of a community."

John Wesley Powell

Director of the USGS (1881-1894)

Watershed Transformation to Group Pixels into Regions

•**Interpret grey level as elevation** •**Decide which pixels are connected** •**Drop water on every pixel and see which local minimum it drains to** •**Identify regions that drain to the same minimum, identify borders also**

Images and Watershed Transformation

Intensity as elevation:

Watershed Transform

Marker-Controlled Watershed Transform

Dunes in Victoria Crater

% clear and close all clear all; close all

```
% read image
I = 
imread('victoria_crater4.png');
imshow(I); 
title('original image');
```


Grayscale

% grayscale image G = rgb2gray(I); figure; imshow(G); title('grayscale');

Watershed

% now the watershed L = watershed(G); G2 = G; G2(L == 0) = 255; figure; imshow(G2); title('result'); impixelinfo

What happened?

Smoothed (Mean)

```
% smooth a little
fSize = 24;
h = ones(fSize,fSize) / fSize^2;
Sm = imfilter(G, h);
figure; imshow(Sm);
title('mean smoothing');
```


Watershed on Smoothed Image

```
% now the watershed
L = watershed(Sm);
L0 = (L==0);
G2 = Sm;
G2(L0) = 255;
figure; imshow(G2);
title('result');
```


Overlay Watershed Boundaries and Original Color Image

% color image

I1=I(:,:,1); I1(L0)=0; % red I2=I(:,:,2); I2(L0)=0; % green I3=I(:,:,3); I3(L0)=0; % blue II=cat(3, I1, I2, I3); % rgb figure; imshow(II); title('color image with watershed boundaries');

Detail View

Good enough?

Grayscale Image vs. Smoothed (Mean)

A Different Way to Smooth: Median Filter

- Removes "salt and pepper" noise
- Preserves edges

```
% smooth a little (median)
fSize = 24;
Sm = medfilt2(G, [fSize fSize]);
figure; imshow(Sm);
title('median smoothing');
```


Smoothed Image: Mean vs. Median

Watershed and Median Smoothing

% now the watershed L = watershed(Sm); L0 = (L==0); G2 = Sm; G2(L0) = 255; figure; imshow(G2); title('result');

Overlay Watershed Boundaries and Original Color Image

```
% color image
I1=I(:,:,1); I1(L0)=0; % red
I2=I(:,:,2); I2(L0)=0; % green
I3=I(:,:,3); I3(L0)=0; % blue
II=cat(3, I1, I2, I3); % rgb
figure; imshow(II); 
title('color image with 
watershed boundaries');
```


Comments, ideas?

Guiding the Watershed Transform

- What do the depressions have in common?
- Can we inform the watershed transform?

Regional Extended Minima

```
% get regional minima "deeper" 
% than a specified threshold
thresh = 3;
Imin = imextendedmin(Sm,thresh);
figure; imshow(Imin);
title('extended minima');
```
Note: similar function for maxima: **imextendedmax()**

Imposed Extended Minima

- Impose regional minima on the original grayscale image
- The function **imimposemin(I, BW)** modifies an intensity image I so that it only has the regional minima contained in BW

```
% now we have internal markers, 
% impose them on original image:
G2 = imimposemin(G, Imin);
figure; imshow(G2);
title('markers');
```


Marker-Controlled Watershed

```
% now the watershed
L2 = watershed(G2);
L0 = (L2 == 0);G3 = Sm;
G3(L0) = 255;
figure; imshow(G3./255);
title('result');
```


Overlay Watershed Boundaries and Original Color Image

% color image I1=I(:,:,1); I1(L0)=0; % red I2=I(:,:,2); I2(L0)=0; % green I3=I(:,:,3); I3(L0)=0; % blue II=cat(3, I1, I2, I3); % rgb figure; imshow(II); title('color image with watershed boundaries');

Better!

Detail View

Watershed Function also Outputs Labels

Output of Watershed function:

- Matrix the size of the image
- Zero on the boundaries
- Region number inside the regions

```
% label and show
map = lines(max(L2(:)));
G4 = label2rgb(L2, map);
figure; imshow(G4);
title('result labels');
```


Applications of Watershed Transform

- **Directly on an image**
- On the gradient of the image
- On the Distance Transform

Figure 8.63 EDM for touching features: (a) binary image of two touching circular features; (b) EDM with pixels color-coded to show distance from boundary; (c) rendered display showing pixel heights. Note the boundary between the two cones.

From: "The Image Processing Handbook"

In Matlab: **bwdist()** *Also called: the Euclidean Distance Map (EDM)*

San Pablo Reservoir

Identified lake area: 190146 pixels (last lecture)

Review: Entropy Filter

Entropy Filter

 0.9

fSize = 9; nHood = true(fSize); E = entropyfilt(G,nhood); figure; imshow(E);

New: Range Filter

New: Standard Deviation Filter

STD Filte

0.9

 0.8

 0.7

 0.6

 0.5

 0.4

 0.3

 $_{0.2}$

'n.

fSize = 9; nHood = true(fSize); E = stdfilt(G,nhood); figure; imshow(E);

New: Gradients

Morphological operations: Erosion, Dilation, Opening Closing

Example of Erosion

Note: Matlab likes to operate on foreground objects!

Image Erosion

- Does B "fit" in the set X?
- Retain all points (i,j) in X such that when B is centered at (i,j) , B is contained in X
- Mathematically:

$\varepsilon_B(X) = \{ \mathbf{x} \mid B_{\mathbf{x}} \subseteq X \}$

In Matlab: **imerode(Img,Elt)**

Image Dilation

In Matlab: **imopen(Img,Elt)**

Image Closing

- **Does B** "fit" in the background of the set X?
- When the previous is true, all of B belongs to the background
- The complement of the new background define X
- Mathematically:

$$
\phi_B(X) = \left[\bigcup_{\mathbf{x}} \{ B_{\mathbf{x}} \mid B_{\mathbf{x}} \subseteq X^c \} \right]^c
$$

In Matlab: **imclose(Img,Elt)**

Structuring Elements

Next time: Segmentation from Edges

The Covariance Matrix (Review)

- Consider a 6-sided *fair* die where *outcomes* all have probability 1/6
- **The** *mean* (the expected value) is: 6 6 $X(X) = \sum_{i=1}^{6} X_i P(X) = \sum_{i=1}^{6} x_i p_i = \frac{1}{6} \sum_{i=1}^{6} x_i = (1 + 2 + 3 + 4 + 5 + 6) / 6 = 3.5$ $\mu = E(X) = \sum_{i} X P(X) = \sum_{i} x_{i} p_{i} = \frac{1}{2} \sum_{i} x_{i}$ $E(X) = \sum_{i=1}^{n} XP(X) = \sum_{i=1}^{n} x_i p_i = \frac{1}{6} \sum_{i=1}^{n} x_i = (1 + 2 + 3 + 4 + 5 + 6) / 6 =$
- The *variance* is the squared deviation from the mean: $e^{2} = Var(X) = E\left[\left(X - \mu\right)^{2}\right] = \sum_{i=1}^{6} p_{i} \left(x_{i} - \mu\right)^{2} = \frac{1}{2} \sum_{i=1}^{6} \left(i - 3.5\right)^{2}$ $I(X) = E\left[\left(X - \mu\right)^2\right] = \sum_{i=1}^{6} p_i \left(x_i - \mu\right)^2 = \frac{1}{6} \sum_{i=1}^{6} \left(i - 3.5\right)^2 \approx 2.9$ $\sigma^2 = Var(X) = E[(X - \mu)^2] = \sum p_i (x_i - \mu)^2 = \frac{1}{Z} \sum (i$ $=Var(X) = E\left[\left(X-\mu\right)^2\right] = \sum_{i=1}^{N} p_i (x_i - \mu)^2 = \frac{1}{6} \sum_{i=1}^{N} (i-3.5)^2 \approx$
- The *covariance matrix* is a generalization of the variance:

$$
\sum = Cov(X) = E\bigg[\big(X - E[X]\big) \big(X - E[X]\big)^{T}\bigg]
$$

■ i.e., the matrix whose *(i,j)* entry is: \sum_{ij} = $Cov(x_i, x_j)$ = $E\left[\left(x_i - \mu_i\right) \left(x_j - \mu_j\right)^T\right]$

For the die this is an identical definition to the variance:

 $\left(\begin{bmatrix} -2.5 & -1.5 & -0.5 & 0.5 & 1.5 & 2.5 \end{bmatrix} \begin{bmatrix} -2.5 & -1.5 & -0.5 & 0.5 & 1.5 & 2.5 \end{bmatrix}^T \right) / 6 \approx 2.9$

But for the color pixels it is a 3x3 matrix

In Matlab: function **cov()**

Computer Lab Assignment 3 Image Segmentation 1

This lab has two parts. First we will apply the *watershed transform* to identify dune regions on Mars. This is an application of the *region-based image segmentation methods* that were discussed in the lecture. In the second part, we will learn how to use the *pixel-based image segmentation* and the *Mahalanobis* distance measure. We will apply this approach to aerial imagery of Marin County, CA and measure the extent of forestation.

Part 1 – Dunes in Craters on Mars

Here we will apply the *watershed* transformation to identify dune regions on Mars. Different filtering techniques will be discussed to obtain good results.

1) Download the file craterdunes $2.jpg$ $\sqrt{11}$ from bSpace. Then start writing a new Matlab script that displays it.

```
 clf ; clear ; close all
I = imread('craterdunes2.jpg');<br>G = rgb2gray(I);figure; imshow(G); title('grey scale image'); impixelinfo
```
Apply the *watershed* transformation

```
L = watershed(G, 8); figure, imshow(L); title('watershed');
```
and superimpose the region boundaries on the original image

```
I1=I (:,:,1); I1(L==0)=255;
I2=I (:,:,2); I2(L=-0)=0;<br>I3=I (:,:,3); I3(L=-0)=0;II=I; II(:,:,1)=I1; II(:,:,2)=I2; II(:,:,3)=I3;<br>imshow(II); title('color image with watershed boundaries');
```
You obtain an over-segmented image that we shall now improve upon.

2) Smooth the image before the *watershed* transform by applying an averaging filter. Insert the following lines into your code:

```
n=3; h = ones(n,n) / (n*n);<br>G = imfilter(G,h);
figure, imshow(G); title('filtered image (averaged)');
```
Increase the filter size *n* until you obtain the best result. You should now have obtained an image with region of reasonable size but with boundaries shifted towards the setting sun.

3) Rather than applying an averaging, or *mean*, filter, try a *median* filter:

```
G2 = double(G);<br>G3 = medfilt2(G2, [n n])./255;
figure; imshow(G3);title('median smoothing');impixelinfo
```
Are the results any better? Decide for yourself and keep using your preferred filtering technique with the optimal filter size for part 4.

4) Now we want to apply the *marker controlled watershed transform* where we first identify the shadiest regions in the filtered image that are darker then their surrounding by a certain threshold. Then we use the Matlab command imimposemin to manipulate the original image in such a way that it only has minima where GMin has been set to zero by the imextendedmin command.

```
 % get regional minima "deeper" than a threshold min_thresh = 3/255;
GMin = imextendedmin(G3, min_thresh);
figure; imshow(GMin);title('extended minima');
 % now we have "markers", impose them on original image
figure; imshow(G4);<br>title('original grey scale with minima imposed (markers for watershed)');impixelinfo
L = watershed(G4, 8);
```
Try this method for different threshold values. You should see how the minima identified by imextendedmin grow or shrink. Characterize how this affects the resulting dune region. Determine the best threshold value and save your final code. Well done!!

Part 2 – Telling the Forest from the Trees

For this part, we will use images from the National Agriculture Imagery Program (NAIP). NAIP acquires aerial imagery during the agricultural growing seasons in the continental U.S.A. and makes digital ortho-photography available to the public through their web site: <http://www.fsa.usda.gov/FSA/apfoapp?area=home&subject=prog&topic=landing> For California, these and other geo-spatial data are available at [http://www.atlas.ca.gov.](http://www.atlas.ca.gov/) These images have a spatial resolution of 1m, meaning that each pixel corresponds to 1 m^2 . For use in our lab, these images were converted to PNG format, hopefully preserving this resolution.

Our goal will be to select a sample forest patch and automatically classify all the forest pixels in these images.

Sample code will be provided for you to cut and paste into your Matlab session.

As the size of these images is rather large, some operations such as smoothing might be slow. For this reason, we recommend you save the code samples in separate files so that you don't have to repeat unnecessary operations and only execute those that you will be modifying. Also, you will need most of this code for your next homework assignment.

2) Read one of the two images into Matlab and display it:

```
% set image name 
imName = 'naip1.png'; 
% read and check image, get size 
rgbImg = imread(imName); 
[nRows nCols nLyrs] = size(rgbImg); 
if (nLyrs \sim = 3)error('Image is not RGB!'); 
end 
nPix = nRows * nCols; 
% display original image 
figure; 
imshow(rgbImg); 
title(['RGB image ' imName]);
```
Note how in the last line the string 'RGB image' is concatenated with the variable "imName".

If you wish you could save this section as readNaip.m

1) Download the two images available on bSpace, named *naip1.png* and *naip2.png*:

3) Smooth the image you selected and display it:

```
% set filter size 
filtSize = 25;% smooth image 
fprintf('Smoothing image ...\n'); 
myFilt = ones(filtSize) ./ filtSize^2; 
smoothImg = imfilter(rgbImg, myFilt);
% display smoothed image 
figure; 
imshow(smoothImg); 
title('smoothed RGB image');
```
Keep in mind that results may be dependent on the filter size chosen here, you may decide to vary the filter size after you see the results. Again, you may wish to save this code in a file such as smoothNaip.m

Region selection is performed with the function roipoly(). This function enables the user to click around a region and create a polygon. To complete the function, right click on the polygon and select the option "Create Mask". The process should look something like this:

4) Select a representative sample forest patch and display it:

```
% print some instructions with the function fprintf() 
fprintf('Use mouse to delineate region for color sample\n'); 
fprintf('Click to place a vertex\n'); 
fprintf('Right-click and select Create Mask to close polygon\n'); 
% open a new figure for the region selection and call roipoly() 
figure; 
title('Select region'); 
roiMask = roipoly(rgbImg); 
% display the mask 
figure; 
imshow(roiMask); 
title('Mask'); 
% use the mask to extract the color region 
fprintf('Masking region ...\n'); 
red = \text{immultiply}(roiMask, smoothImg(:,:,1));green = immultiply(roiMask, smoothImg(:,:,2));blue = immultiply(roiMask, smoothImg(:,:,3));
roiImg = cat(3, red, green, blue);
% display the color region 
figure; 
imshow(roiImg); 
title('Color sample');
```
You may want to save this code in a separate file (such as selectNaip.m) in order to easily repeat the selection process roipoly() produces a black and white mask. The color region is extracted from the smoothed image using this mask: multiplying the image's individual color planes by this mask preserves only the pixels which correspond to the value of 1 in the mask. The function cat() can then concatenate the resulting color planes back together into a new image. The result should look something like this:

To compute statistics and distances, we need to reshape both the image and the color patch into vectors. This is done by using the function reshape(), you are encouraged to look it up in the help. Also note that these vectors are converted from integers to doubles.

5) Compute mean patch color and standard deviation, and print their values:

```
% reshape image into vector of size nPix by 3 
imgFix = double(reshape(smoothImg, npix, 3));% reshape roi region into vector of size nPix by 3 
roipix = double(reshape(roilmq, npix, 3));% find the nonzero entries in the mask 
roitdx = find(roimark);% keep only the RGB pixels corresponding to the mask 
roipix = roipix(roiIdx, 1:3);% get statistics 
fprintf('Computing region statistics \ldots\n');
% mean color (could be used for drawing the forest in a "truer" color) 
myMean = mean(roipix)% standard deviation (use the max value for R, G, or B) 
myStd = max(std(rootpix))
```
Don't forget to save your code. Note that printing the mean and std can be done by simply omitting the semi-colon at the end of the line. Also note the above code only stores the maximum standard deviation (from red, green, or blue). Finally, notice the use of the function find() which returns the indices of the nonzero values. These indices are then used to select the masked color information. Look at the help for this function.

The Mahalanobis distance (see lecture notes) is computed using the function mahal(). This function computes the distance for each row of a (large) vector to a smaller sample, by considering the covariance of the smaller sample. The resulting distance needs to then be reshaped into the size of the image.

6) Compute Mahalanobis distance from all image pixels to region pixels:

```
% compute distance vector 
mahDist = mahal(imqPix, roiPix);
% reshape distance and store in matrix 
myDist = reshape(mahDist, nRows, nCols); 
% display the distance 
figure; 
imagesc(myDist); 
title('Mahalanobis distance from region color'); 
colorbar; 
impixelinfo
```
What are the sizes of imgPix, roiPix, and mahDist? What is the size of myDist after the reshaping?

Enlarge the figure with the Mahalanobis distance and hover over it with your cursor (use the command impixelinfo if you didn't include it in the code). What value ranges do you see? Can you use them to discriminate forest from non-forest?

Segmentation is performed by assigning each pixel to either forest or not forest, based on this distance. Rather than picking a completely arbitrary value, we will use the region's standard deviation value as a threshold of distance. If the distance is less than one or two standard deviations (you get to play with these values), then the pixel is classified as forest.

7) Segment the image based on the Mahalanobis distance and display the result:

```
% define how many standard deviations are allowed 
kThresh = 1;% myDist contains the distances of each pixel to mean color 
% if minimum is more than k*std label it 0, else label 1 
fprintf('Thresholding pixels with distance from region color ...\n'); 
mySegm = false(nRows, nCols); 
mySegm(myDist <= kThresh * myStd) = true; 
% get the red, green and blue parts of the image 
Ir = rqbImq(:,:,1);Ig = rgbImg(:,:,2);Ib = rgbImg(:,:,3);% label the segmented pixels in red 
Ir(mySegm==1) = 255;Ib(mySeqm==1) = 0;Iq(mySeqm==1) = 0;% put the red, green, and blue parts back together 
I = cat(3, Ir, Ig, Ib);% display the labeled image 
figure; imshow(I);
title('Classified forest on original image');
```
Look at the line mysegm(myDist \leq kThresh \star myStd) = true; What does it do? How many operations is it performing? Can you think of a way to do the same using loops? Replacing loops in this fashion is called "code vectorization".

Your result should look something like this:

Compare the original image (the very first figure) with your result. Are you identifying most of the forested areas? Are you getting too much or too little forest? Which parameters affect this outcome?

8) As the region you selected is rather small compared to the entire image, one standard deviation might not be sufficient to accommodate the variation in the forest colors. In your homework you will learn how to sample from several regions, but for now we can simply relax the distance threshold. Increase the standard deviation threshold (kThresh) to another value (perhaps to 1.5 or 2) and re-run the code from section 7. Again, compare with the original image. Are you getting too much or too little forest? How does the smoothing affect this result?

In this exercise we restricted classification to separating forest from not-forest. In general, more land cover classes may be desirable. One way to visualize such results may be with the use of a normalized histogram.

9) Build a normalized histogram of the forest and non-forest pixel counts:

```
% count the foreground pixels 
myCount = sum(mySeqm(:));
% make a normalized distribution 
myDistr = [myCount nPix-myCount] ./ nPix; 
% display histogram 
figure; 
bar(myDistr); 
axis([0.5 2.5 0 1]); 
set(gca, 'XTickLabel', {'Forest', 'Other'}); 
title('Forest percentage');
```
Note that because mySegm contained only 0 and 1 values, it was sufficient to sum the mySegm matrix to count the forest pixels. This will not be true in general, you will have to select the specific class values and count how many times they occur. You have done a similar operation in previous exercises using loops. This operation can also be done in a "vectorized fashion". For example, consider these two statements: $class2 = (mySegm == 2)$; $class2count = sum(class2(:))$; What do they do? Also note in the above code how one can set axis labels on a figure by using the functions set() and gca() (get current axis).

Perhaps the forest was not our only interest, and we wished to also measure how much plowed land we see in the image. We can select a different patch and run the code again without any changes. Note that you only have to start from step 4), no need to run from the beginning unless you wish to use a different smoothing filter.

10) Sample a patch from the image that looks like it has been plowed or farmed (like in the example below) and re-run the code starting from step 4.

How does your code perform in identifying the farmed pixels? Did you have to change the standard deviation threshold or the smoothing filter size? How different are the distance values from the ones you examined during the forest exercise? Your result should look something like this (except perhaps better!):

Today: GIS and the Mapping Toolbox

Geographic Information Systems

- Data abstraction and data types
- Datums and projections
- Data sources
- LiDAR data
- Web Map Service

GIS and Mapping Toolbox

- Features
- Data
- Analysis
- Pros and Cons

Course Evaluations

- Fill out forms
- Deliver to McCone

Lab Exercise

- Making a map
- Retrieving Web-based data
- Real-time weather
- Ortho-photos
- Digital Elevation Model

"Blue Marble Earth", NASA

Geographic Information Systems

Figure: Konecny, "Geoinformation: Remote Sensing, Photogrammetry and Geographical Information Systems"

A *geographic information system,* as defined in the Environmental Systems Research Institute's (ESRI's) *Dictionary of GIS Terminology*, is a collection of computer hardware, software, and geographic data for capturing, storing, updating, manipulating, analyzing, and displaying all forms of geographically referenced information.

Common GIS Systems

State of the art:

Knocking on the door:

The future (?):

Google earth engine

Commercial Commercial Commercial Commercial Communication Communica

Figure: Sherman, "Desktop GIS"

GIS Data Abstraction

Figure: NCDDC/NOAA

GIS Data Types

 \mathbf{r}

Figures: Brimicombe, "GIS, environmental modeling and engineering"

Raster Data

Vector vs. Raster Example

FIGURE 2.6

Four possible methods of representing landslides in GIS: (a) as points, (b) as lines, (c) as polygons, (d) as a tessellation (raster).
Vector and Raster Formats

Raster (Image)

Vector

Automated Mapping System (AMS) **ESRI** Coverage Computer Graphics Metafile (CGM) Digital Feature Analysis Data (DFAD) Encapsulated Postscript (EPS) Microstation drawing file format (DGN) Dual Independent Map Encoding (DIME)

Digital line Graph (DLG) AutoCAD Drawing Exchange Format (DXF) AutoCAD Drawing (DWG) MapBase file (ETAK) **ESRI** Geodatabase Land Use and Land Cover Data (GIRAS) Interactive Graphic Design Software (IGDS) Initial Graphics Exchange Standard (IGES) Map Information Assembly Display System (MIADS) MOSS Export File (MOSS) TIGER/line file: Topologically Integrated Geographic Encoding and Referencing (TIGER) Spatial Data Transfer Standard/Topological

Vector Profile (SDTS/TVP)

Arc Digitized Raster Graphics (ADRG) Band Interleaved by line (BIL) Band Interleaved by Pixel (BIP) Band Sequential (BSQ) Windows Bitmap (BMP) Device-Independent Bitmap (DIB) Compressed Arc Digitized Raster Graphics (CADRG) Controlled Image Base (CIB) Digital Terrain Elevation Data (DTED) ERMapper Graphics Interchange Format (GIF) ERDAS IMAGINE (IMG) ERDAS 7.5 (GIS) ESRI GRID file (GRID) JPEG File Interchange Format (JFIF) Multi-resolution Seamless Image Database (MrSID) Tag Image File Format (TIFF; GeoTIFF) Portable Network Graphics (PNG)

Figure: Fazal, "GIS Basics"

Georeferencing: Datums

Georeferencing: Projections

Figures: Clarke, "Getting Started with GIS"

UTM Projection

GIS Operations

GIS Operations

Entire world at 30 arc seconds (~1km) resolution

to

NUSGS NASA

ESAID

KINEGI (GSI)

GIS Data Sources: USGS

GIS Data Sources: NASA

GIS Data Sources: NOAA

GIS Data Sources: HydroSHEDS

GIS Data Sources: Visible Earth

GIS Data Sources: EarthScope

GIS Data Sources: geodata.gov

GIS Data Sources: BARD

GIS Data Sources: Google

Google code a "adwords" of "open source"

Search

Google Maps API Family

Google Maps has a wide array of APIs that let you embed the robust functionality and everyday usefulness of Google Maps into your own
website and applications, and overlay your own data on top of them:

Maps JavaScript API

Embed a Google Map in your webpage
using JavaScript. Manipulate the map and
add content through many services.
Version 3 - Version 2

Static Maps API Embed a fast and simple Google Maps
Image in your web page or mobile site
without requiring JavaScript or any
dynamic page loading.
Learn more

Maps API for Flash

Use this ActionScript API to embed a
Google Map in your Flash-based web
page or app. Manipulate the Map in three
dimensions and add content through
many services. Learn more

Web Services

Use URL requests to access geocoding, directions, elevation, and places
information from client applications, and
manipulate the results in JSON or XML.
Learn more

Google Earth API Embed a true 3D digital globe into your
web page. Take your visitors anywhere on
the Earth (even below the ocean) without
leaving your web page.
Learn more

Custom Panoramas in the Javascript $\frac{\Delta \mathbf{B}}{\Delta \mathbf{P1}}$ July 21, 2010

Create custom panoramas and display them
using the Street View service in the Maps
Javascript API V3!

FusionTables in the Javascript API

Documentation on Fusion Table Layers,
allowing you to display geographic data on
your map!

Home FAQ Articles Blog Forum Terms

About Google Maps API

The Maps API is a free service, available for
any web site that is free to consumers.
Please see the terms of service for more
information.

Businesses that charge fees for access, businesses untuitaige tees tot actess,
track assets or build internal applications
must use <u>Google Maps API Premier</u>, which
provides enhanced features, technical
support and a service-level agreement.

What's New

Styled Maps in the Static Map API.

After the appearance of the standard map
styles in the API, customizing the colors and
display of features and elements!

GIS Data Sources: Microsoft

GIS Data Sources: Canada

LiDAR Data

LiDAR Data Sources: NCALM

LiDAR Data Sources: USGS

Various states and local governments (via USGS portal)

LiDAR Data Sources: OpenTopography

LiDAR Data Sources: Bathimetry

ST. JOHN,

Web Map Service (WMS)

Mapping Toolbox: Features

Mapping Toolbox: Data

The MathWorks MATLAB&SIMULINE **Using Geospatial Analysis and Geodesy**

- · Distance and area calculations
- · 3D coordinate transformations
- · Spherical and ellipsoidal geometry
- Finding lines and polygons intersections
- Navigational calculations
- · Track and circle tools
- Map profiling

Figure: the Mathworks

ESRI ArcGIS

ESRI ArcGIS

ESRI ArcView: SHALSTAB

EPS 209 Course Evaluations

Fill out two forms:

- Instructor (Burkhard Militzer)
- GSI (Dino Bellugi)

Need a volunteer:

- Collect the forms
- Deliver them to McCone Hall
	- \checkmark Gretchen vonDuering (371)
	- \checkmark Margie Winn (398)
	- \checkmark Main office (307)

Lab Exercise in 15 minutes:

- Making a map
- Retrieving Web-based data
- Real-time weather
- Ortho-photos
- Digital Elevation Model

Computer Lab Assignment 7 Maps and Web Map Service (WMS) Data

This lab has three parts. In the first part, we learn to make a simple map. In the second part, we retrieve a satellite image and weather data from a WMS and we add it to the map, to simulate a cross country flight. In the third part we load a digital elevation model (DEM) and display it. Then we will retrieve aerial imagery and drape it on the DEM. As you learn new Matlab functions, look them up in the help for examples of their usage.

This spectacular "blue marble" image is the most detailed true-color image of the entire Earth to date. Using a collection of satellite-based observations, scientists and visualizers stitched together months of observations of the land surface, oceans, sea ice, and clouds into a seamless, true-color mosaic of every square kilometer (.386 square mile) of our planet. These images are freely available to educators, scientists, museums, and the public from NASA web site and WMS servers.

Part 1 – My first Matlab map

Here we make a very simple map of the United States to familiarize ourselves with some functionalities of the Mapping Toolbox.

(1) Let's start by making a blank map of the Continental US. Type the following commands in your editor and execute them:

```
fiqure;
usamp('conus');
```
(2) Now let's load a shapefile (a common file format for vector data) with the borders of the states. This file is called 'usastatehi.shp' and it comes packaged with the Mapping Tollbox: Type the following commands in your editor window and execute them:

```
states = shaperead('usastatehi.shp', 'UseGeoCoords', true);
```
 (3) The variable 'states' is an array of structures. Type states (1) into the command window. You will get a description of the data structure:

```
Geometry: 'Polygon'
BoundingBox: [2x2 double]
```
 Lon: [1x521 double] Lat: [1x521 double] Name: 'Alabama' LabelLat: 32.2827 LabelLon: -86.9206

Notice the variables present in this structure. In particular, note the two arrays 'Lon' and 'Lat', as there are what are used by Matlab to draw the shape of the state.

(3) Now let's load a shapefile (a common file format for vector data) with the location and names of major world cities. This file is called 'worldcities. shp'and it also comes packaged with the Mapping Tollbox: Type the following commands in your editor window and execute them:

```
cities = shaperead('worldcities.shp', 'UseGeoCoords', true);
```
Again, type cities(1) in the command window. Note that since the geometry is now a point, Lat and Lon now have only a single value.

(4) Now let's display the map: type the following commands in your editor window and execute them:

geoshow(states);

(5) Now let's add the names of the states to the map. Note that the structure in the shapefile also contained fields called 'LabelLat' and 'LabelLon'. We will use these coordinates to place our labels. Type the following commands in your editor window and execute them:

```
textm([states.LabelLat], [states.LabelLon], {states.Name},
'HorizontalAlignment', 'center', 'FontSize', 6);
```
(6) Now let's plot the cities to the map. Type the following commands in your editor window and execute them:

plotm([cities.Lat], [cities.Lon], '.r');

(7) Now let's add the names of the cities to the map. Note that the cities structure did not contain label coordinates. We will thus use the city coordinates but place our labels right above the city. Type the following commands in your editor window and execute them:

textm([cities.Lat], [cities.Lon], {cities.Name}, 'HorizontalAlignment', 'center', 'VerticalAlignment', 'bottom', 'FontSize', 5);

(8) Finally, add a title to the map:

title('My first map: the continental USA');

You made your first Matlab geographic map! It should look something like this:

Save your code as you will be adding to it in the next part.

Part 2 – Flying across the USA

In this part we will update the previous map with images and real-time information.

(1) Let's start again by making a blank map of the Continental US and reading the states boundaries and the major cities:

```
figure;
usamp('conus');states = shaperead('usastatehi.shp', 'UseGeoCoords', true);
cities = shaperead('worldcities.shp', 'UseGeoCoords', true);
```
(2) As we will be retrieving data for this map, we will need to define what the latitude and longitude limits are. We can do this by retrieving the map structure of the current blank map:

 $mstruct = qcm;$

The command 'gcm' means "get current map", and its output is assigned to the map structure mstruct. Type 'mstruct' in the command window to see its contents. In particular, note the fields 'maplatlimit' and 'maplonlimit'. These are the bounding box of your map. Assign them to your latlim and lonlim variables:

```
latlim = mstruct.maplatlimit;lonlim = mstruct.maplonlimit;
```
(3) Now we will retrieve the part of the "Blue Marble Earth" image (seen in the lecture) that fits into this map. The first step is finding the correct WMS server. We know the image is from NASA, so we will search for servers that have the word NASA in their URL:

nasa = wmsfind('NASA', 'SearchField', 'serverurl');

In the variable editor, explore the WMS layer array 'nasa', and note the diverse nature of its contents. Now we can refine the search to look for the words 'blue marble':

```
nasa = nasa.refine('blue marble');
```
The command 'refine' is an object function of the layer structure. Again, in the variable editor explore the WMS layer array 'nasa', and note the much reduced nature of its contents.

(4) Now we are ready to retrieve the image using the bounding box of our map:

```
[BM, R] = wmsread(nasa(1), 'ImageFormat', 'image/png', 'Latlim', latlim, 
'Lonlim', lonlim, 'CellSize', 0.1);
```
In the variable editor, look at the variables BM and R. The former is the RGB image, and the latter is a geo-referencing matrix used for drawing the image onto the map.

(5) Now we are ready to display the image and the state boundaries on our map:

```
figure(gcf);
geoshow(BM, R);
hold on
geoshow(states, 'FaceColor', 'none', 'EdgeColor', 'w');
geoshow(cities, 'Color', 'r', 'Marker', '.');
```
Note how BM and R are passed to 'geoshow'. Also note that to overlay the multiple layers one has to use "hold on" just like with regular Matlab figures.

(6) Let's pretend that we are about to fly between two of these cities. Let's trace our route by inputting the start and end points directly on the map:

```
figure(gcf);
disp('Input start and end points')
[lat lon] = inputm(2);
start = \lceil \text{lat}(1) \text{lon}(1) \rceil;
dest = [\text{lat}(2) \text{lon}(2)];
```
The function 'inputm' takes the number of desired points as a parameter. Click on the start and end city.

(7) The next step is to construct a great circle (the shortest distance on the surface of a sphere) between the two cities and to display it on the map. This is done with the navigation functions 'gcwaypts' and 'track', and the usual 'geoshow':

```
[lat_gc,lon_gc] = gcwaypts(start(1),start(2),dest(1),dest(2));
[lattrk_gc lontrk_gc] = track('gc', lat_gc, lon_gc, 'degrees');
figure(gcf);
geoshow(lattrk_gc, lontrk_gc, 'DisplayType', 'line', 'Color', 'r');
```
(8) One can measure the length of this path using the navigation function 'legs' which returns the bearing and the distance along the way:

 $[course_gc dist_gc] = legs(lat_gc, lon_gc, 'gc');$

disp(['Great circle path length: ' num2str(sum(dist_gc))]);

(9) We now have our course, but it would be nice to check the weather along the way. We can look up the radar data, much like we did for the blue marble image. Search for 'nexrad' among the WMS servers:

```
nexrad = wmsfind('nexrad', 'SearchField', 'serverurl');
```
In the variable editor, explore the WMS layer array 'nexrad, and note its contents. Now we can refine the search to look for the words 'current':

```
nextad = nextrad.refine('current');
```
(10) Retrieve the weather data and display it on the map:

```
[W, R2] = wmsread(nexrad(1), 'ImageFormat', 'image/png', 'Latlim', latlim,
'Lonlim', lonlim, 'ImageHeight', size(BM,1), 'ImageWidth', size(BM,2),<br>'BackgroundColor', [0 0 0]);
figure(qcf);qeoshow(W, R2);
```
Note how various parameters of 'wmsread' were used to get an image of the same size as the blue marble image and with a transparent background color. Note however, that when we show this image, we erase the blue marble image. We can fix this by adding the pixels showing precipitation to the blue marble image:

```
windex = any(W > 0, 3);
windex = cat(3, windex, windex, windex);WBM = BM;WBM(windex) = uint8(W(windex) . *255);figure(gcf);geoshow(WBM, R2);
```
Your map should now look something like this:

How is the weather along the route? If you were trying to plot a route that is less direct but avoids the bad weather, you'd probably want to look not just at the current radar image, but also at a sequence of images prior to the current one to see where the storm is going. Nexrad offers these data on the WMS server as well.

Part 3 – Retrieving elevation data and aerial imagery for the San **Francisco Bay Area**

Among the data repositories we saw in the lecture there was the San Francisco Bay Area Database $(BARD)$. Regional Open a browser and $\overline{Q}O$ this URL: to http://bard.wr.usgs.gov/getDEMSMap.html. Here you will see an index map with the USGS 7.5minute Digital Elevation Models (DEM). The tiles in this dataset have 10 meter grid spacing and units of (hopefully) meters.

Matlab has functions to read data directly from web sites such as this one, we will use them to download and display a DEM of your choice.

(1) Let's start by picking a DEM to explore. On the BARD web page hover with the cursor over the desired quadrangle to obtain its name. In the example listed below I use the 'sf_north' DEM, but feel free to change it to one of your choice:

```
demFilename = 'sf north.dem';demExt = \cdot, qz';
dataServerURL = 'http://bard.wr.usqs.qov/bard/elevation/';
```
(2) Now we can use the Matlab function 'urlwrite' to write the file from the URL to the current directory, and the function 'gunzip' to uncompress it:

```
URL = \lceildataServerURL demFilename demExt\lceil;
urlwrite(URL, [demFilename demExt]);
gunzip([demFilename demExt]);
```
(3) Use the mapping toolbox function 'usgs24kdem' to import the elevation data:

```
[Lat, Lon, Elv, Header] = uss24kdem(demFilename);
```
Note that this function returns matrices for latitude, longitude and elevation, as well as a header structure with the DEM info. Type 'Header' in the command window to explore its contents. Then get the bounding box of the DEM from the Lat and Lon matrices:

```
latlim = [min(Lat(:)) max(Lat(:))];
lonlim = [min(Lon(:)) max(Lon(:))]
```
(4) Display the elevation data using 'usamap' and 'geoshow':

```
figure;
usamap(latlim, lonlim);
geoshow(Lat, Lon, Elv, 'DisplayType','surface');
title(demFilename, 'Interpreter', 'none');
```
Note that Matlab used the standard color map (jet) to display the elevation. Generate a more topographic colormap by using the function 'demcmap':

demcmap(Elv);

Want a 3-D view? It's simple with the 'view' function:

view(3);

Perhaps we would like to exaggerate the vertical axis for a more 3-D effect. This can be done with the 'daspectm' function:

```
daspectm(<mark>'m'</mark>,1.5)
```
Your map should now look something like this:

```
122.5 W
```
sf north.dem

(5) Lets make and display a slope map, using Matlab's 'gradientm' function:

```
[aspect, slope, gradN, gradE] = gradient(Lat, Lon, Elv);figure;
usamap(latlim, lonlim);
geoshow(Lat, Lon, slope, 'DisplayType', 'texturemap');
colorbar
colorbar<br>title([demFilename ' Slope (degrees)'], 'Interpreter', 'none');
```
Does your map look like this?

(6) Now let's search for an aerial image to drape over our DEM. We can get one from WMS, much like we did for the radar data. Search for 'usgs' and 'california' among the WMS servers:

```
ortho = wmsfind('usgs*california', 'SearchField', 'serverurl', 'Latlim',
latlim, 'Lonlim', lonlim);
```
Note the use of the wildcard '*' in the search. In the variable editor, explore the WMS layer array 'ortho', and note its contents. We know we want the high-resolution (0.3m) color image, so we can refine the search to look for the words '0.3m' and 'color':

ortho = ortho.refine('0.3m*color', 'SearchField', 'LayerTitle');

In the variable editor one can see that some of the fields of the individual layers say '<Update using WMSUPDATE>'. One can update the information with the function 'wmsupdate':

 $ortho =$ wmsupdate(ortho(1));

(7) Now we are ready to retrieve the image. As you have already done it twice, offered here is an alternative method to create the request:

```
% create a WMS request structure
server = WebMapServer(ortho.ServerURL);
request = WMSMapRequest(ortho, server);
% modify map request for the desired limits, format and size
request. Latlim = latlim;
request.Lonlim = lonlim;request.ImageFormat = 'image/png';requestu. Image Height = size(Elv, 1);
request.ImageWidth = size(Elv,2);% Request the map
Map = server.getMap(request.RequestURL);
Ref = request. RasterRef;
```
(8) Let's display the DEM again, but with the elevation draped on it:

figure;

Spring 2011, EPS 209 "Matlab Applications in Earth Sciences", Instructors: Burkhard Militzer and Dino Bellugi

usamap(latlim, lonlim); geoshow(Lat, Lon, Elv, 'DisplayType', 'surface', 'CData', Map);
title([demFilename ' + ' ortho.LayerName], 'Interpreter', 'none');

Exaggerate and switch to 3-D view:

daspectm $('m', 1.5)$ $view(3)$

Isn't WMS groovy? Your map should look something like this:

sf_north.dem + SanFranciscoCA_0.3m_Color_Jun_2009_01

See you next week for final project presentations!

Support Vector Machine Classification

2nd International Summer School on Water Research

Landslide modeling and Early Warning Systems *8 July 2013*

Dino Bellugi *Massachusetts Institute of Technology*

Support Vector Machine Classification

SVM-based Classification

- Classification
- Object representation
- Training and validation
- Linearly separable data
- Linearly non-separable data
- Non-linearly separable data
- The SVM formulation
- The Kernel trick

Rock Classification

- Rock image database
- Rock characteristics
- Creating a rock descriptor
- SVM Software
- SVM Cookbook
- Training an SVM
- A test
- Results

Landslide Identification

• Deep seated landslides

Landslide Prediction

- Landslide database
- Landslide characteristics
- Creating the descriptor
- Training the SVM
- A preliminary test

Discussion

- Storm classification
- A real application (Luigi!)

Some slides adapted from: Dino Bellugi - EPS 209:

"*Matlab Applications in Earth Science*"

Michael Jordan - CS 294:

"*Practical Machine Learning*" University of California, Berkeley

Classification

- In classification problems, each entity in some domain can be placed in one of a discrete set of categories: yes/no, friend/foe, good/bad/indifferent, etc.
- Given a training set of labeled entities, develop a rule for assigning labels to entities in a test set
- For example:
	- Observe whether a given medication affects various patients positively or negatively over several years (the training set).
	- Given this data, extract a rule allowing us to predict whether or not any new patient will respond positively or negatively to the medication.
- Many variations on this theme:
	- binary classification
	- multi-category classification
	- non-exclusive categories

Example: face detection

Example: object recognition

Example: object recognition

Try to find: blimp, clutter, grasshopper, picnic-table, refrigerator, watermelon

Example: object recognition

Object Representation

- Each object to be classified is represented as a pair (*x, y*):
	- x is a description of the object (see examples of data types in the following slides)
	- *y* is a label (assumed binary for now: 1 or -1)
- Success or failure of a machine learning classifier often depends on choosing good descriptions of objects
	- the choice of description can also be viewed as a learning problem
	- but good human intuitions are often needed here
- Vectorial data:
	- physical attributes
	- textual attributes
	- context
	- history

feature vector (x)

Example: Spam Filter

- Input: email
- Output: spam/ham
- Setup:
	- Get a large collection of example emails, each labeled "spam" or "ham"
	- Note: someone has to hand label all this data
	- Want to learn to predict labels of new, future emails
- Features: The attributes used to make the ham / spam decision
	- Words: FREE!
	- Text Patterns: \$dd, CAPS
	- Non-text: SenderInContacts
	- …

First, I must solicit your confidence in this transaction, this is by virture of its nature as being utterly confidencial and top secret. …

TO BE REMOVED FROM FUTURE MAILINGS, SIMPLY REPLY TO THIS MESSAGE AND PUT "REMOVE" IN THE SUBJECT.

99 MILLION EMAIL ADDRESSES FOR ONLY \$99

Ok, Iknow this is blatantly OT but I'm beginning to go insane. Had an old Dell Dimension XPS sitting in the corner and decided to put it to use, I know it was working pre being stuck in the corner, but when I plugged it in, hit the power nothing happened.

Example: Digit Recognition

Training and Validation

Some State of the Art Classifiers

- **Support vector machines (SVMs)**
- Decision trees
- Random forests
- Kernelized logistic regression
- Kernelized discriminant analysis
- Kernelized perceptron
- Bayesian classifiers
- Boosting and other ensemble methods

Some Resources

- Google 'Berkeley practical machine learning' for more information
- Trevor Hastie's "The elements of statistical learning: data mining, inference, and prediction." Springer. 2001
- Nello Cristianini's web page: http://www.support-vector.net/

Which Hyper-plane to Use?

Setting up the Optimization Problem

The Optimization Problem

The maximum margin can be characterized as a solution to an optimization problem:

 max γ s.t. $y \cdot (\vec{w} \cdot \vec{x} + b) \ge \gamma$, $\forall (x, y)$ in training set $\|\vec{w}\|^2 = 1$

Linear Hard-Margin SVM Formulation

• Simple manipulation yields an equivalent problem: find *w,b* that solves

$$
\min \frac{1}{2} ||w||^2
$$

s.t. $y_i (w \cdot x_i + b) \ge 1, \forall x_i$

- Problem is convex, so there is a unique global minimum value (when feasible).
- There is also a unique minimizer, i.e. *w* and *b* value that provides the minimum.
- Quadratic Programming
	- very efficient computationally with procedures that take advantage of the special structure

Linear Non-Separable Case

Allow some instances to fall within the margin, but penalize them.

Introduce slack variables ξ

Formulating the Optimization Problem

$$
\max \ \gamma - C \sum \xi_i
$$

s.t. $y_i \cdot (\vec{w} \cdot \vec{x}_i + b) \ge \gamma - \xi_i$, $\forall i$
 $\xi_i \ge 0$, $\forall i$
 $\|\vec{w}\|^2 = 1$

Objective function penalizes for misclassified instances and those within the margin

C trades-off margin width and misclassifications

Linear Soft-Margin SVM's

• Equivalent problem:

$$
\min \frac{1}{2} ||w||^2 + C \sum_i \xi_i
$$

 $y_i(w \cdot x_i + b) \geq 1 - \xi_i, \ \forall x_i$ $\xi_i \geq 0$

- Algorithm tries to push ξ*ⁱ* to zero while maximizing margin
- As *C*→0, we get the hard-margin solution
- Notice: algorithm does not minimize the *number* of misclassifications (NP-complete problem) but the sum of distances from the margin hyperplanes
- Other formulations can use ξ*ⁱ ²*instead

Robustness of Hard vs. Soft Margin SVM's

Soft Margin SVM Hard Margin SVM

Linear Classifiers in High-Dimensional Spaces

Mapping Data to High-Dimensional Spaces

Find function $\Phi(x)$ to map to a different space, then SVM formulation becomes:

$$
\min \frac{1}{2} ||w||^2 + C \sum_{i} \xi_i \qquad \qquad \xi_i \ge 0
$$

s.t. $y_i(w \cdot \Phi(x) + b) \geq 1 - \xi_i, \forall x_i$ $\xi_i \geq$

- Data appear as Φ(x), weights *w* are now weights in the new space
- Explicit mapping expensive if $\Phi(x)$ is very high dimensional
- Solving the problem without explicitly mapping the data is desirable

The Kernel Trick

- $\Phi(x_i) \cdot \Phi(x_j)$: means map data into new space, then take the inner product of the new vectors
- We can instead simply find a function such that: $K\langle x_{i}\cdot x_{j}\rangle = \Phi(x_{i})\cdot \nabla$ $\Phi(\mathsf{x}_j)$, i.e., the image of the inner product of the data is the inner product of the images of the data
- Then, we do not need to explicitly map the data into the highdimensional space to solve the optimization problem

Kernels

• Some common kernels

- Linear kernel: $k(x,z) = x^{T}z$ **equivalent to linear algorithm**
- Polynomial kernel: $k(x, z) = (1 + x^Tz)^d$ **polynomial decision rules**
- RBF kernel: **k(x,z) = exp(-||x-z||2/2**σ**) highly nonlinear decisions**

Histograms of Oriented Gradients for Human Detection

Navneet Dalal and Bill Triggs INRIA Rhône-Alps, 655 avenue de l'Europe, Montbonnot 38334, France {Navneet.Dalal,Bill.Triggs}@inrialpes.fr, http://lear.inrialpes.fr

Figure 1. An overview of our feature extraction and object detection chain. The detector window is tiled with a grid of overlapping blocks in which Histogram of Oriented Gradient feature vectors are extracted. The combined vectors are fed to a linear SVM for object/non-object classification. The detection window is scanned across the image at all positions and scales, and conventional non-maximum suppression is run on the output pyramid to detect object instances, but this paper concentrates on the feature extraction process.

Figure 6. Our HOG detectors cue mainly on silhouette contours (especially the head, shoulders and feet). The most active blocks are centred on the image background just outside the contour. (a) The average gradient image over the training examples. (b) Each "pixel" shows the maximum positive SVM weight in the block centred on the pixel. (c) Likewise for the negative SVM weights. (d) A test image. (e) It's computed R-HOG descriptor. (f,g) The R-HOG descriptor weighted by respectively the positive and the negative SVM weights.

Representing shape with a spatial pyramid kernel

Anna Bosch
University of Girona
Computer Vision Group
17003 Girona, Spain aboschr@eia.udg.es Andrew Zisserman
University of Oxford
Robotics Research Group
OX1 3PJ Oxford, UK
az@robots.ox.ac.uk

Xavier Munoz University of Girona
Computer Vision Group
17003 Girona, Spain xmunoz@eia.udg.es

- Implements HOG on a quad-tree
- Canny edges, Sobel gradients
- No smoothing
- Gradients transferred to edges
- Binned for orientation
- Weighted by their strength
- PHOG descriptor: Concatenation of HOG descriptors for each level of pyramid (BFS)
- Matlab code available from the Robotics Research Group (Visual Geometry), University of Oxford: www.robots.ox.ac.uk/~vgg

Figure 1: Shape spatial pyramid representation. Top row: an image and grids for levels $l = 0$ to $l = 2$; Below: histogram representations corresponding to each level. The final PHOG vector is a weighted concatenation of vectors (histograms) for all levels. Remaining rows: images from the same and from different categories, together with their histogram representations.

CIVR 2007

Rock Classification**Rock-forming Type of rock and Example** source material process **IGNEOUS Crystallization Melting of rocks in** (solidification of hot, deep crust and magma or lava) upper mantle **Granite SEDIMENTARY Deposition, Weathering and** burial, and erosion of rocks lithification exposed at surface **Sandstone METAMORPHIC** Rocks under high **Recrystallization** temperatures and of new minerals pressures in deep in solid state crust and upper mantle **Gneiss**

Figure 3.24
Understanding Earth, Sixth Edition
© 2010 W. H. Freeman and Company

Rock Classification: Igneous

i Opener
n**ding Earth, Sixth Edition**
H. Freeman and Company

Rock Classification: Sedimentary

Figure 5.10
Understanding Earth, Sixth Edition
© 2010 W. H. Freeman and Company

Rock Classification: Metamorphic

Rock Database

Virtual Geology Museum

sponsored by Cochise College

Hall of Rocks -introduction by Opal's Pals

Attractions: Igneous Rock Photos Sedimentary Rock Photos Metamorphic Rock Photos Rocks Used in Building Quick Tours of Rock Types

Roger Weller, curator (wellerr@cochise.edu)

Igneous Rocks

Sedimentary Rocks

Training Set: Igneous Rocks (85 samples)

Training Set: Metamorphic Rocks (56 samples)

Training Set: Sedimentary Rocks (70 samples)

Rock Descriptor

Igneous

Color Spaces

% convert image to R, G, B, HSV and to Gray G = double(rgb2gray(I))./255; HSV = rgb2hsv(I); H = HSV(:,:,1); R = I(:,:,1); Gr = I(:,:,2); B = I(:,:,3);

Multiple Scales (Spatial Pyramid)

function [distrG bSeps]= … makeHistograms(G, nBins, imScale, noZero, doCat, opts) % makes normalized histograms % with nBins bins at scale imScale % if doCat is false: % returns a matrix nBins by nBlocks % if doCat is true: % returns a vector of length nBins x nBlocks % if noZero is true: % only nonzero elements are considered

Rock Color

Grain Contours: Oriented Edges

% creates a matrix of oriented edges: % each canny edge pixel contains the angle of the gradient direction

Rock Texture: Local Standard Deviation

Rock Texture: Local Entropy

A lot more histograms than what is shown: at 3 scales there are 25 histograms per image

Rock Texture: Local Range

Co-Occurrence Matrix

Gray Level Co-Occurrence Matrix (GLCM)

- GLCM functions characterize texture
- Calculate how often pairs of pixels with specific values and in a specified spatial relationship occur in an image
- Function of angle and distance
- Various properties can be extracted
- In Matlab: graycomatrix() and graycoprops()

offsets = $[0 1; 0 2; 0 3; 0 4; \dots]$

Co-Occurrence Matrix

circuitBoard = rgb2gray(imread('board.tif')); imshow(circuitBoard);

% create horizontal offsets $offsets0 = [zeros(40, 1) (1:40)']$;

% get GLCM and stats glcms = graycomatrix(circuitBoard,'Offset',offsets0); stats = graycoprops(glcms,'Contrast Correlation');

% plot correlation figure, plot([stats.Correlation]); title('Texture Correlation as a function of offset'); xlabel('Horizontal Offset'); ylabel('Correlation')**;**

Global Values

% compute normalized entropy at all levels EG = []; EH = []; sEG = []; sEH = []; for n = 1:nLevs EG = [EG; getNEntrs(G, n, opts)]; EH = [EH; getNEntrs(H, n, opts)]; sEG = [sEG; getNEntrs(sG, n, opts)]; sEH = [sEH; getNEntrs(sH, n, opts)]; end % compute standard deviation at all levels DG = []; DH = []; sDG = []; sDH = []; for n = 1:nLevs DG = [DG; getStddevs(G, n, opts)]; DH = [DH; getStddevs(H, n, opts)]; sDG = [sDG; getStddevs(sG, n, opts)]; sDH = [sDH; getStddevs(sH, n, opts)]; end % compute variance at all levels VG = []; VH = []; sVG = []; sVH = []; for n = 1:nLevs VG = [VG; getVariances(G, n, opts)]; VH = [VH; getVariances(H, n, opts)]; sVG = [sVG; getVariances(sG, n, opts)]; sVH = [sVH; getVariances(sH, n, opts)]; end % compute mean at all levels AG = []; AH = []; sAG = []; sAH = []; for n = 1:nLevs AG = [AG; getMeans(G, n, options)]; AH = [AH; getMeans(H, n, options)]; sAG = [sAG; getMeans(sG, n, options)]; sAH = [sAH; getMeans(sH, n, options)]; end % compute median at all levels MG = []; MH = []; sMG = []; sMH = []; for n = 1:nLevs MG = [MG; getMedians(G, n, options)]; MH = [MH; getMedians(H, n, options)]; sMG = [sMG; getMedians(sG, n, options)]; sMH = [sMH; getMedians(sH, n, options)]; end

A lot more values than what is shown: at 3 scales there are 25 values per field

The Descriptor (22,052 Dimensions)

% concatenate descriptor descriptor = [...

```
GDis; HDis; sGDis; sHDis; ... % intensity, hue histogram
RDis; GrDis; BDis; sRDis; sGrDis; sBDis; ... % RGB histogram
rGDis; rHDis; srGDis; srHDis; ... % local range histogram
eGDis; eHDis; seGDis; seHDis; ... % local entropy histogram
dGDis; dHDis; sdGDis; sdHDis; ... % local std histogram
statsG.Correlation'; sstatsG.Correlation'; ... % intensity correlation
statsH.Correlation'; sstatsH.Correlation'; ... % hue correlation
statsG.Contrast'; sstatsG.Contrast'; ... % intensity contrast
statsH.Contrast'; sstatsH.Contrast'; ... % hue contrast
statsG.Energy'; sstatsG.Energy'; ... % intensity energy
statsH.Energy'; sstatsH.Energy'; ... % hue energy
statsG.Homogeneity'; sstatsG.Homogeneity'; ... % intensity homogeneity
statsH.Homogeneity'; sstatsH.Homogeneity'; ... % hue homogeneity
EG; EH; sEG; sEH; ... % global entropy
VG; VH; sVG; sVH; ... % global variance
DG; DH; sDG; sDH; ... % global std
AG; AH; sAG; sAH; ... % global mean
MG; MH; sMG; sMH; ... % global median
];
```
SVM Software

LIBSVM: a Library for Support Vector Machines

Chih-Chung Chang and Chih-Jen Lin *Department of Computer Science*

National Taiwan University, Taipei 106, Taiwan

http://www.csie.ntu.edu.tw/~cjlin *(Version 3.0 released: September 13, 2010)*

Abstract

LIBSVM is a library for support vector machines (SVM). Its goal is to help users to easily use SVM as a tool. In this document, we present all its implementation details. For the use of LIBSVM, the README file included in the package and the LIBSVM FAQ provide the information.

Different SVM formulations Efficient multi-class classification Cross validation for model selection Probability estimates Various kernels (including precomputed kernel matrix) Weighted SVM for unbalanced data Both C++ and Java sources GUI demonstrating SVM classification and regression Python, R, MATLAB, Perl, Ruby, Weka, Common LISP, CLISP, Haskell, and LabVIEW, interfaces. C# .NET code and CUDA extension is available. It's also included in some data mining environments: RapidMiner and PCP. Automatic model selection which can generate contour of cross validation accuracy.

SVM Cookbook

A Practical Guide to Support Vector Classication

Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin *Department of Computer Science* National Taiwan University, Taipei 106, Taiwan http://www.csie.ntu.edu.tw/~cjlin *(Initial version: 2003 Last updated: April 15, 2010)*

Abstract

The support vector machine (SVM) is a popular classification technique. However, beginners who are not familiar with SVM often get unsatisfactory results since they miss some easy but significant steps. In this guide, we propose a simple procedure which usually gives reasonable results.

We propose that beginners try the following procedure first:

- Transform data to the format of an SVM package
- Conduct simple scaling on the data
- Consider the RBF kernel K(x; y) = $e^{-\gamma ||x-y||^2}$
- Use grid search and cross-validation to find the best parameters C and γ
- Use the best parameters C and to train the whole training set
- Test

SVM Type and Cross Validation

Parameter Search: Coarse

Parameter Search: Fine

Training the SVM

Training:

- Use the best discovered parameters
- Train on the entire training data (no cross-validation)

Check:

- Use the trained model on the training data
- Ideally you should get 100% accuracy

```
% use best parameters
svm_params = [' -c ', num2str(bestc), ' -g ', num2str(bestg) ' '];
% train
svm_model = svmtrain(trainLabels, trainDescriptors, [svm_type svm_params]);
% test
[labels, accuracy, value] = ... 
   svmpredict(trainLabels, trainDescriptors, svm_model, svm_opts);
```
Classifying with the SVM

Validating:

- Use the trained model on the separate testing data with labels (same as on previous slide but on data that was not part of the training)
- Decide if your accuracy is good enough

```
% test on separate labeled data
[labels, accuracy, value] = ... 
   svmpredict(testLabels, testDescriptors, svm_model, svm_opts);
```

```
Predicting:
```
• Use the trained model on new data with unknown labels

```
% generate random labels (2-class in this example)
randLabels = double(round(rand(numInstances, 1)));
% predict on new data
[labels, accuracy, value] = ... 
   svmpredict(randLabels, neDescriptors, svm_model, svm_opts);
```
iRock: Results

How well did it do?

Note: the selection of the training and testing images was entirely random (and no rocks were harmed in the process)

Man vs. Machine: the Turing Test

Turing, A.M. (1950). Mind, 59, 433-460. **COMPUTING MACHINERY AND INTELLIGENCE**

By A. M. Turing

I propose to consider the question, "Can machines think?"
This should begin with definitions of the meaning of the terms "machine" and "think." The definitions might be

framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous, if the
meaning of the words "machine" and "think" are to be found by examining how they are commonly used it is
diffic unambiguous words.

The new form of the problem can be described in terms of a game which we call the 'imitation game." It is
played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The
interrogato

C: Will X please tell me the length of his or her hair?

Now suppose X is actually A, then A must answer. It is A's object in the game to try and cause C to make the wrong identification. His answer might therefore be:

"My hair is shingled, and the longest strands are about nine inches long."

In order that tones of voice may not help the interrogator the answers should be written, or better still, typewritten. The ideal arrangement is to have a tele-printer communicating between the two rooms.
Alternatively the question and answers can be repeated by an intermediary. The object of the game for the
third player (B) i

We now ask the question. "What will happen when a machine takes the part of A in this game?" Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman? These questions replace our original, "Can machines think?"

Alan Mathison Turing, 1912-1954

Man vs. Machine: Deep Blue

It was the second victory of the match for the computer -- there were three draws -making the final score $3\frac{1}{2}$ to $2\frac{1}{2}$, the first time any chess champion has been beaten by $\,$ a machine in a traditional match. Mr. Kasparov, 34, retains his title, which he has held since 1985, but the loss was nonetheless unprecedented in his career; he has never before lost a multigame match against an individual opponent.

Garry Kimovich Kasparov World Chess Champion 1985-2000

Man vs. Machine: Watson

Facing certain defeat at the bands of a room-size LB.M. computer on Wednesday evening, Ken Jennings, famous for
winning 74 games in a row on the TV quiz show, acknowledged the obvious. "I, for one, welcome our new
computer overlords," he wrote on his video screen, borrowing a line from a "Simpsons" episode.

From now on, if the answer is "the computer champion on "Jeopardy!," the question will be, "What is Watson?

For I.B.M., the showdown was not merely a well-publicized stunt and a \$1 million prize, but proof that the company has taken a big step toward a world in which intelligent machines will understand and respond to humans, and perhaps inevitably, replace some of them.

Watson, specifically, is a "question answering machine" of a type that artificial intelligence researchers have struggled with for decades $-$ a computer akin to the one on "Star Trek" that can understand questions posed in natural language and answer them.

Man vs. Machine: ISSWR Students & iRock

Man vs. Machine: ISSWR Students & iRock

- Take one picture card and three colored voting cards
- Study the picture card while we get ready
- A random sequence of 26 images will be shown
- Vote quickly by raising one of the colored cards
- One volunteer to call the vote
- Another volunteer to tally the counts on the board **Ready?**

ISSWR vs. iRock: Image 1

Truth: Limestone – Sedimentary iRock: Sedimentary

ISSWR vs. iRock: Image 2

Truth: Schist – Metamorphic iRock: Metamorphic

ISSWR vs. iRock: Image 3

Truth: Shale – Sedimentary iRock: Sedimentary

ISSWR vs. iRock: Image 4

Truth: Diorite – Igneous iRock: Igneous

ISSWR vs. iRock: Image 5

Truth: Andesite – Igneous iRock: Igneous

ISSWR vs. iRock: Image 6

Truth: Conglomerate – Sedimentary iRock: Sedimentary

ISSWR vs. iRock: Image 7

Truth: Gypsum – Sedimentary iRock: Sedimentary

ISSWR vs. iRock: Image 8

Truth: Marble – Metamorphic iRock: Metamorphic

ISSWR vs. iRock: Image 9

Truth: Shale – Sedimentary iRock: Sedimentary

ISSWR vs. iRock: Image 10

Truth: Granodiorite – Igneous iRock: Igneous

ISSWR vs. iRock: Image 11

Truth: Anhydrite – Sedimentary iRock: Sedimentary

ISSWR vs. iRock: Image 12

Truth: Granite – Igneous iRock: Igneous

ISSWR vs. iRock: Image 13

Truth: Marble – Metamorphic iRock: Metamorphic

ISSWR vs. iRock: Image 14

Truth: Rhyolite – Igneous iRock: Metamorphic

ISSWR vs. iRock: Image 15

Truth: Skarn – Metamorphic iRock: Metamorphic

ISSWR vs. iRock: Image 16

Truth: Limestone – Sedimentary iRock: Sedimentary

ISSWR vs. iRock: Image 17

Truth: Schist – Metamorphic iRock: Metamorphic

ISSWR vs. iRock: Image 18

Truth: Syenite – Igneous iRock: Igneous

ISSWR vs. iRock: Image 19

Truth: Sandstone – Sedimentary iRock: Sedimentary

ISSWR vs. iRock: Image 20

Truth: Schist – Metamorphic iRock: Metamorphic

ISSWR vs. iRock: Image 21

Truth: Conglomerate – Sedimentary iRock: Sedimentary

ISSWR vs. iRock: Image 22

Truth: Diabase – Igneous iRock: Igneous

ISSWR vs. iRock: Image 21

Truth: Limestone – Sedimentary iRock: Sedimentary

ISSWR vs. iRock: Image 24

Truth: Siltstone – Sedimentary iRock: Sedimentary

ISSWR vs. iRock: Image 25

Truth: Schist – Metamorphic iRock: Metamorphic

ISSWR vs. iRock: Image 26

Truth: Volcanic Sandstone – Igneous iRock: Igneous

Man vs. Machine: ISSWR Students & iRock

… out of 26 correct 24 out of 26 correct

iRock: Recap

6andesite2659.jpg

6syenite-homblende111a.jpg

6trip-whitney9.jpg

6diorite120a.jpg

6volcanic-sandstone7-39a.jpg

6rxgranite-polished4.jpg

Truth: Igneous iRock: Igneous (87.5%), Sedimentary (12.5%), Metamorphic (0%)

iRock: Recap

6conglomerate1842a.jpg

6limestone-oolitic-

rock72c.jpg

6conglomerate1843a.jpg

6sandstone-flagstone-
rock61a.jpg

6gypsum-rock75b.jpg

6shale-arenaceousrock53a.jpg

6limestone-chalkrock67a.jpg

6shale-carbonaceousrock54a.jpg

6limestone-oolitic-
rock72a.jpg

6travertine155a.jpg

Truth: Sedimentary iRock: Sedimentary (91%), Igneous (0%), Metamorphic (9%)

iRock: Recap

6schist-quartz-sericite-
rock90b.jpg

6schist-tourmaline-micarock89b.jpg

6mrx-marble-polishedswirl1.jpg

garnet179c.jpg

6schist-talc-rock88b.jpg

Truth: Metamorphic iRock: Metamorphic (100%), Igneous (0%), Sedimentary (0%)

6marble1860.jpg

6skarn-wollastonite-

iRock: Another Test (after re-training)

6diorite120a.jpg

6diorite1877a.jpg

6granite-hornblenderock014b.jpg

6irx-diorite-polished1.jpg

6pyroxenite-harzburgite-
rock043b.jpg

6trip-whitney9.jpg

Truth: Igneous iRock: Igneous (87.5%), Metamorphic (12.5%), Sedimentary (0%)

iRock: Another Test (after re-training)

Truth: Sedimentary iRock: Sedimentary (80%), Igneous (0%), Metamorphic (20%)

iRock: Another Test (after re-training)

6gneiss-biotite-rock79b.jpg

Sgneiss1850a.jpg

6mrx-gneiss-flat1.jpg

6mrx-marble-polished breccia5.jpg

6marble1900.jpg

6marble-tobermorite180b.jpg

Not so lucky this time!

Truth: Metamorphic iRock: Metamorphic (42.9%), Igneous (57.1%), Sedimentary (0%)

iRock: Discussion

What happened and why?

How much training? How much testing?

Can we take better advantage of prior knowledge? How could we apply this technique to landslide prediction? How could we apply this technique to landslide identification?

How many features? Which features?

Landslide (deep) identification

Signature:

- Rougher texture - Edges around scarp - Differently dissected - Differently sloping

USGS 10m data

Signature:

- Smoother texture
- Less defined edges
- Flatter slopes
- More uniform slope direction

Can we learn the signature independently of the type of data?

Lib-SVM (thanks Subhransu!)

- Matlab implementation:
	- LibSVM
	- Training, test, and
	- classify routines - Linear, Radial Basis,
	- or Sigmoid Kernels
- Learning:
	- 6 training patches (red – landslide, green – non landslide)

Lib-SVM

(thanks Subhransu!)

- Matlab implementation:
	- LibSVM
	- Training, test, and
	- classify routines
	- Linear, Radial Basis, or Sigmoid Kernels

- Learning:

- 6 training patches (red – landslide, green – non landslide)

-Testing:

- 10 new patches (orange – landslide, yellow – non landslide)

Lib-SVM Results

Lib-SVM Results

- Much to my surprise: Nine out of ten correct! (correct – green, incorrect - red)
- False positive: Not sure who was right …

Lib-SVM Results

Landslide (shallow) prediction

Shalstab: a compact simple model

[Montgomery and Dietrich, 1994]

Shalstab: Performance (over-prediction)

Soil depth

- **Soil production:**

$$
-\frac{\partial z_b}{\partial t} = \varepsilon e^{-\alpha h}
$$

- **Soil transport:**

$$
\tilde{q} = \frac{K\nabla z}{1 - \left(\left|\nabla z\right| / S_c\right)^2}
$$

5m contours

Soil depth (m) 0.0 or N.A. $0.0 - 0.1$ $0.1 - 0.2$ $0.2 - 0.4$ $0,4 - 0.6$

 $0.6 - 0.8$ $0.8 - 1$

 $1.0 - 1.5$ $1, 5 - 2$ >2

- **Regionally calibrated**
- **No landsliding in this realization!**

Observation: landslides in thick soils

Root Strength

[Benda & Dunne, 1997; Schmidt et al., 1999; Montgomery et al., 2009]

The descriptor

True positives and true negatives?

True positives and true negatives?

Training and testing

Training and testing

Parameter search (radial basis function)

Typical values

5m contours channels site boundary mapped slides predicted slides

Shalstab $log(q/T)$ unstable <-3.4 $-3.4 - -3.1$ $-3.1 - 2.8$ $-2.8 - -2.5$ $-2.5 - -2.2$ $-2.2 - 1.9$ -1.9 stable

Test area: all data pixels in red polygon (not seen in training phase)

(+) all pixels inside mapped slides (not in test area), and (-) a **4-pixel** buffer around them

Smoothing: None

Background: Shalstab draped on shaded relief

Pixel Statistics: Accuracy = 97.0 %, Precision = 34.8%, Recall = 12.9%, F-Score: 18.9%, True Positive Rate = 12.9%, False Positive Rate = 0.7%

F-Score: 12.3%, True Positive Rate = 25.7%, False Positive Rate = 8.2%

F-Score: 15.8%, True Positive Rate = 26.1%, False Positive Rate = 13.7%

Results are encouraging:

- Implicitly figured out a Shalstab-like rule
- Reduced overprediction
- Couple physical and empirical models?
- Need landslide databases with long term observations!
- Can we apply to the temporal domain?

Application to storms: Seattle landslides

Application to storms: Seattle landslides

Summary

- Data-driven approaches are easy to implement given good training data
- They can be used to identify geomorphological features in a landscape
- Such methods also have good predictive potential
- Coupling mechanistic and empirical slope stability models can help reduce over-prediction
- Similar approach can be used to improve the prediction of landslidetriggering storms

We need large, detailed, accurate, and long-term landslide datasets!

TEACHING EVALUATIONS

The following pages contain student teaching evaluations two courses at the University of California, Berkeley:

- Instructor, Spring 2011, EPS209 "Matlab Applications in Earth Science". New graduate course offering a practical toolbox for analyzing Earth science data, and to explore selected problems in earth and environmental sciences, with particular focus on image processing and machine learning techniques. Responsible for curriculum, lectures, and labs development, grades, and office hours. Co-developed and co-taught with Prof. Burkhard Militzer.
- Graduate Student Instructor, Fall 2009, EPS50 "The Planet Earth". Instructors: Prof. Michael Manga and Prof. Doug Dreger. Undergraduate introductory course on geology and geophysics. Gave lectures, guided labs and field trips, advised students, graded assignments, and held office hours.

 \sim

GRADUATE STUDENT INSTRUCTOR EVALUATION

Considering both the limitations and the possibilities of the subject matter and the course, how would you rate the overall teaching effectiveness of this graduate student instructor?

GRADUATE STUDENT INSTRUCTOR EVALUATION

Considering both the limitations and the possibilities of the subject matter and the course, how would you rate the overall teaching effectiveness of this graduate student instructor?

What comments or suggestions do you have regarding the teaching effectiveness of this GSI? (Please feel free to use the other side of this page if you need to).

Agam, I just wish the labs of the were. a) less complex 4 shorter b) Offer the class for optronal 3units and have a lab every week to get through

* I'm maybe a different case b/c lhave

no programming SIE!/15.

GRADUATE STUDENT INSTRUCTOR EVALUATION

Considering both the limitations and the possibilities of the subject matter and the course, how would you rate the overall teaching effectiveness of this graduate student instructor?

What comments or suggestions do you have regarding the teaching effectiveness of this GSI? (Please feel free to use the other side of this page if you need to).

It's really cool, just sometimes the homework assignments are not very clear so I need to read a couple times to understand. I really like the class.

GRADUATE STUDENT INSTRUCTOR EVALUATION

Considering both the limitations and the possibilities of the subject matter and the course, how would you rate the overall teaching effectiveness of this graduate student instructor?

GRADUATE STUDENT INSTRUCTOR EVALUATION

Considering both the limitations and the possibilities of the subject matter and the course, how would you rate the overall teaching effectiveness of this graduate student instructor?

thanks for the hard mork Dine!

GRADUATE STUDENT INSTRUCTOR EVALUATION

Considering both the limitations and the possibilities of the subject matter and the course, how would you rate the overall teaching effectiveness of this graduate student instructor?

Very good class

GRADUATE STUDENT INSTRUCTOR EVALUATION

Considering both the limitations and the possibilities of the subject matter and the course, how would you rate the overall teaching effectiveness of this graduate student instructor?

GRADUATE STUDENT INSTRUCTOR EVALUATION

GSI: Dino Bellug:

Course/Section: EPS 207 Semester/Year: Sl

Please circle the number that indicates the degree to which these statements describe your GSI.

Considering both the limitations and the possibilities of the subject matter and the course, how would you rate the overall teaching effectiveness of this graduate student instructor?

and the company of the state.

 $\mathcal{L}(\mathcal{A})$ and $\mathcal{L}(\mathcal{A})$ are $\mathcal{L}(\mathcal{A})$. The simple state $\mathcal{L}(\mathcal{A})$

Fantastre class! I really appresent all of
the hard work that you put in.

GRADUATE STUDENT INSTRUCTOR EVALUATION

Considering both the limitations and the possibilities of the subject matter and the course, how would you rate the overall teaching effectiveness of this graduate student instructor?

What comments or suggestions do you have regarding the teaching effectiveness of this GSI? (Please feel free to use the other side of this page if you need to).

Is a voly Tanowrop Eauctor. HIS VELTURES AND
WERE voly EFFECTIVE AND WILL SERVE AS סמוּע LABS THOROUGH Is very APPOACHABLE AND EXTROMERY KNOWLEDGABLE DINO ABout TAE **WERTH** would betwitting THEA CLASS FROM DINU ACAIN. \mathcal{L}

GRADUATE STUDENT INSTRUCTOR EVALUATION

 \mathbf{r}

Considering both the limitations and the possibilities of the subject matter and the course, how would you rate the overall teaching effectiveness of this graduate student instructor?

Evaluate and Return the homework!
We got & Feedback during the whole class!

GRADUATE STUDENT INSTRUCTOR EVALUATION

Semester/Year: 57 209 Course/Section: EDS 209 GSI: Divo Bellugi Please circle the number that indicates the degree to which these statements describe your GSI. In general, your GSI: not at all very $n.a./$ don't know descriptive descriptive $\overline{5}$ $\overline{2}$ 3 4 $\left(\ \right)$ 1. appears to have a good 1 knowledge of the subject \int_0^1 $\overline{4}$ $\left(\ \right)$ 2. is well prepared $\mathbf{1}$ $\overline{2}$ 3 $\widehat{5}$ $\mathbf{1}$ $\overline{2}$ $\left(\ \right)$ 3. uses class time effectively 3 $\overline{4}$ $\widehat{5}$ $\overline{2}$ 3 $\overline{4}$ $()$ 4. explains new material clearly $\mathbf{1}$ 5. has clear objectives for each $\mathbf{1}$ $\overline{2}$ 3 $\overline{\mathbf{4}}$ $()$ class session $\mathbf{1}$ $\overline{2}$ 3 $\overline{4}$ $()$ 6. is concerned that students $5²$ learn the material $\widehat{5}$ $()$ 7. raises challenging questions or $\mathbf{1}$ $\overline{2}$ 3 $\overline{\mathbf{4}}$ reviews of the material covered 8. knows if the class is $\mathbf{1}$ $\overline{2}$ 3 $\boldsymbol{4}$ $5¹$ $\left(\ \right)$ understanding him/her 5 9. presents clear summaries or $\mathbf{1}$ $2¹$ 3. $\overline{\mathbf{4}}$ $\left(\ \right)$ reviews of the material covered 5. $\overline{2}$ $\left(\ \right)$ 10. makes you feel comfortable about 1 3 4 asking questions or expressing ideas 11. is thoughtful and precise when $\mathbf{1}$ $\overline{2}$ 3 4 $\left(\ \right)$ answering questions $()$ 12. helps clarify points not $\mathbf{1}$ $\overline{2}$ 3 $\overline{\mathbf{4}}$ understood in lecture

Considering both the limitations and the possibilities of the subject matter and the course, how would you rate the overall teaching effectiveness of this graduate student instructor?

What comments or suggestions do you have regarding the teaching effectiveness of this GSI? (Please feel free to use the other side of this page if you need to).

Great job on lectures Quality of HW and Lab is outstanding.
Really cares about teaching
Excited about subject

GRADUATE STUDENT INSTRUCTOR EVALUATION

GSI: Dino Bellugi Course/Section: EPS 209 Semester/Year: Spring 201 Please circle the number that indicates the degree to which these statements describe your GSI. In general, your GSI: not at all very $n.a./$ descriptive descriptive don't know (5) $\mathbf{1}$ $\overline{2}$ $\mathbf{3}$ 4 $\left(\ \right)$ 1. appears to have a good knowledge of the subject $\binom{5}{ }$ 2. is well prepared $\mathbf{1}$ $\overline{2}$ 3 4 $\left(\ \right)$ G) $\mathbf{1}$ $\overline{2}$ 3. uses class time effectively 3 4 $\left(\ \right)$ \circ 4. explains new material clearly $\mathbf{1}$ $\overline{2}$ 3 4 $\left(\ \right)$ 5. has clear objectives for each $\mathbf{1}$ $\overline{2}$ 3 4 ලා $()$ class session (5) $\overline{2}$ $\overline{\mathbf{3}}$ $()$ 6. is concerned that students $\mathbf{1}$ $\overline{\mathbf{4}}$ learn the material 5 7. raises challenging questions or $\mathbf{1}$ $\overline{2}$ 3 $\overline{4}$ $()$ reviews of the material covered 8. knows if the class is $\overline{2}$ 3 $\left(4\right)$ 5 $\mathbf{1}$ $()$ understanding him/her $\left(4\right)$ 9. presents clear summaries or $\mathbf{1}$ $\overline{2}$ 3 5 $()$ reviews of the material covered ග 10. makes you feel comfortable about 1 $\overline{2}$ 3 $\overline{\mathbf{4}}$ $\left(\ \right)$ asking questions or expressing ideas (5) 11. is thoughtful and precise when $\mathbf{1}$ $\overline{2}$ 3 $\overline{4}$ $()$ answering questions (වි 12. helps clarify points not $\mathbf{1}$ $\overline{2}$ 3 $\overline{\mathbf{4}}$ $()$ understood in lecture

Considering both the limitations and the possibilities of the subject matter and the course, how would you rate the overall teaching effectiveness of this graduate student instructor?

 $\ddot{}$

GRADUATE STUDENT INSTRUCTOR EVALUATION

 $-$ Continued on reverse $-$

 $\ddot{}$

Considering both the limitations and the possibilities of the subject matter and the course, how would you rate the overall teaching effectiveness of this graduate student instructor?

Best GSI ever.

GRADUATE STUDENT INSTRUCTOR EVALUATION

GSI: Dino Bellugi

 \sim

 $\overline{11}$

Course/Section: EPS 209 Semester/Year: Spr_{up} //

Please circle the number that indicates the degree to which these statements describe your GSI.

Considering both the limitations and the possibilities of the subject matter and the course, how would you rate the overall teaching effectiveness of this graduate student instructor?

Divid has been a great GSI and clearly put
a lot of effort into nualing tuis dass both on his nar

 $\overline{1}$

GRADUATE STUDENT INSTRUCTOR EVALUATION

Considering both the limitations and the possibilities of the subject matter and the course, how would you rate the overall teaching effectiveness of this graduate student instructor?

What comments or suggestions do you have regarding the teaching effectiveness of this GSI? (Please feel free to use the other side of this page if you need to).

I know it is hard to achieven in "2" int class, but it will be much niker if you could explain more about theory behind image segmentation. In addition, sightlim metiles for both lab and homeunte will be necessary to develop more efficant coding skill.

GRADUATE STUDENT INSTRUCTOR EVALUATION

GSI: Dino

Course/Section: EPS 5D

Semester/Year: Fall 09

Please circle the number that indicates the degree to which these statements describe your GSI.

GRADUATE STUDENT INSTRUCTOR EVALUATION, Continued

ś.

Considering both the limitations and the possibilities of the subject matter and the course, how would you rate the overall teaching effectiveness of this graduate student instructor?

GRADUATE STUDENT INSTRUCTOR EVALUATION

GSI: Dino Bellueji Course/Section: 50

Semester/Year: Fall 2009

Please circle the number that indicates the degree to which these statements describe your GSI.

GRADUATE STUDENT INSTRUCTOR EVALUATION, Continued

Considering both the limitations and the possibilities of the subject matter and the course, how would you rate the overall teaching effectiveness of this graduate student instructor?

GRADUATE STUDENT INSTRUCTOR EVALUATION

Semester/Year: Fall o 9 $linO$ Course/Section: GSI: Please circle the number that indicates the degree to which these statements describe your GSI. $n.a.$ very not at all In general, your GSI: don't know descriptive descriptive 5 $()$ 3 4 $\mathbf{1}$ $\overline{2}$ 1. appears to have a good knowledge of the subject $\left(\right)$ $5⁻$ 3 $\overline{4}$ $\overline{2}$ $\mathbf{1}$ is well prepared $2.$ $5⁻$ $()$ 3 $\overline{4}$ $\overline{2}$ uses class time effectively $\mathbf{1}$ 3. 5 $\left(\right)$ 4^{-} $\overline{2}$ 3 $\mathbf{1}$ 4. explains new material clearly $()$ 5) 3 $\overline{4}$ $\overline{2}$ $\mathbf{1}$ 5. has clear objectives for each class session $\widehat{4}$ $()$ 5 3 $\overline{2}$ 6. is concerned that students $\mathbf{1}$ learn the material $\left(\right)$ $\overline{5}$ $\overline{4}$ $\overline{2}$ 3 7. raises challenging questions or $\mathbf{1}$ reviews of the material covered $5²$ $()$ 3 $\overline{4}$ $\overline{2}$ $\mathbf{1}$ 8. knows if the class is understanding him/her 5 $\left(\right)$ 3 4 $\overline{2}$ 9. presents clear summaries or $\,1$ reviews of the material covered $()$ $\overline{5}$ 10. makes you feel comfortable about 1 3 $\overline{4}$ $\overline{2}$ asking questions or expressing ideas $()$ 5 $\overline{4}$ $\overline{2}$ 3 11. is thoughtful and precise when 1 answering questions $()$ 5 $\overline{3}$ $\overline{4}$ $\overline{2}$ 12. helps clarify points not 1 understood in lecture

GRADUATE STUDENT INSTRUCTOR EVALUATION, Continued

Considering both the limitations and the possibilities of the subject matter and the course, how would you rate the overall teaching effectiveness of this graduate student instructor?

What comments or suggestions do you have regarding the teaching effectiveness of this GSI? (Please feel free to use the other side of this page if you need to).

Very good, terkes time to explain.

GRADUATE STUDENT INSTRUCTOR EVALUATION

GSI: DINO BOUNGOT

Course/Section: 67560

Semester/Year: FAW ZOOG

Please circle the number that indicates the degree to which these statements describe your GSI.

GRADUATE STUDENT INSTRUCTOR EVALUATION, Continued

Considering both the limitations and the possibilities of the subject matter and the course, how would you rate the overall teaching effectiveness of this graduate student instructor?

What comments or suggestions do you have regarding the teaching effectiveness of this GSI? (Please feel free to use the other side of this page if you need to).

- DINO IS AWOSCHO! VOTH KNOW, MULTISUNDOTESINDING I am the the warristenes when no one conservation

GRADUATE STUDENT INSTRUCTOR EVALUATION, Continued

Considering both the limitations and the possibilities of the subject matter and the course, how would you rate the overall teaching effectiveness of this graduate student instructor?

GRADUATE STUDENT INSTRUCTOR EVALUATION

 $GSI: 2000$ plugi

Course/Section: $\overline{\text{SO}}$

Please circle the number that indicates the degree to which these statements describe your GSI.

GRADUATE STUDENT INSTRUCTOR EVALUATION, Continued

Considering both the limitations and the possibilities of the subject matter and the course, how would you rate the overall teaching effectiveness of this graduate student instructor?

GRADUATE STUDENT INSTRUCTOR EVALUATION, Continued

Considering both the limitations and the possibilities of the subject matter and the course, how would you rate the overall teaching effectiveness of this graduate student instructor?

GRADUATE STUDENT INSTRUCTOR EVALUATION, Continued

Considering both the limitations and the possibilities of the subject matter and the course, how would you rate the overall teaching effectiveness of this graduate student instructor?

GRADUATE STUDENT INSTRUCTOR EVALUATION, Continued

Considering both the limitations and the possibilities of the subject matter and the course, how would you rate the overall teaching effectiveness of this graduate student instructor?

What comments or suggestions do you have regarding the teaching effectiveness of this GSI? (Please feel free to use the other side of this page if you need to).

no comments,

GRADUATE STUDENT INSTRUCTOR EVALUATION

GSI: DINO

Course/Section: EPS SO Semester/Year: FallO9

Please circle the number that indicates the degree to which these statements describe your GSI.

GRADUATE STUDENT INSTRUCTOR EVALUATION, Continued

Considering both the limitations and the possibilities of the subject matter and the course, how would you rate the overall teaching effectiveness of this graduate student instructor?

GRADUATE STUDENT INSTRUCTOR EVALUATION

 $()$

 $5¹$

 \bigoplus

 $3¹$

 $\overline{2}$

 $\mathbf{1}$

12. helps clarify points not understood in lecture

answering questions

GRADUATE STUDENT INSTRUCTOR EVALUATION, Continued

Considering both the limitations and the possibilities of the subject matter and the course, how would you rate the overall teaching effectiveness of this graduate student instructor?

What comments or suggestions do you have regarding the teaching effectiveness of this GSI? (Please feel free to use the other side of this page if you need to).

I wish he could provide better explanations in reviews before lato. He goes really fast through the review

GRADUATE STUDENT INSTRUCTOR EVALUATION, Continued

Considering both the limitations and the possibilities of the subject matter and the course, how would you rate the overall teaching effectiveness of this graduate student instructor?

GRADUATE STUDENT INSTRUCTOR EVALUATION, Continued

Considering both the limitations and the possibilities of the subject matter and the course, how would you rate the overall teaching effectiveness of this graduate student instructor?

What comments or suggestions do you have regarding the teaching effectiveness of this GSI? (Please feel free to use the other side of this page if you need to).

I think Dino is sometimes confused by Not sure of
the carese Material, especially the rocks and Mineral

GRADUATE STUDENT INSTRUCTOR EVALUATION

GSI: Dino Bellugi Course/Section: ERS SO < Semester/Year: 56/109

Please circle the number that indicates the degree to which these statements describe your GSI.

GRADUATE STUDENT INSTRUCTOR EVALUATION, Continued

Considering both the limitations and the possibilities of the subject matter and the course, how would you rate the overall teaching effectiveness of this graduate student instructor?

