Image analysis of malign melanoma: Waveles and svd

Dan Dolonius

University of Gothenburg

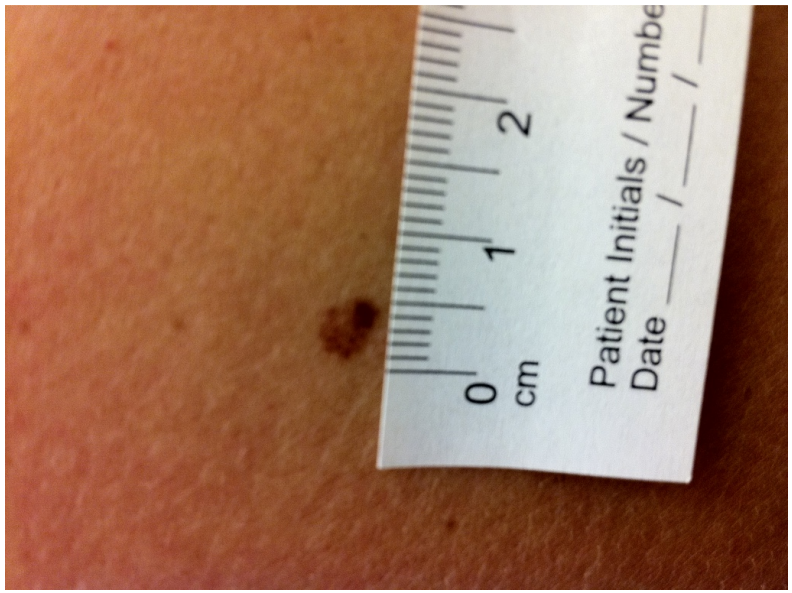
gusdolod@student.gu.se

April 28, 2015

Overview

Melanoma

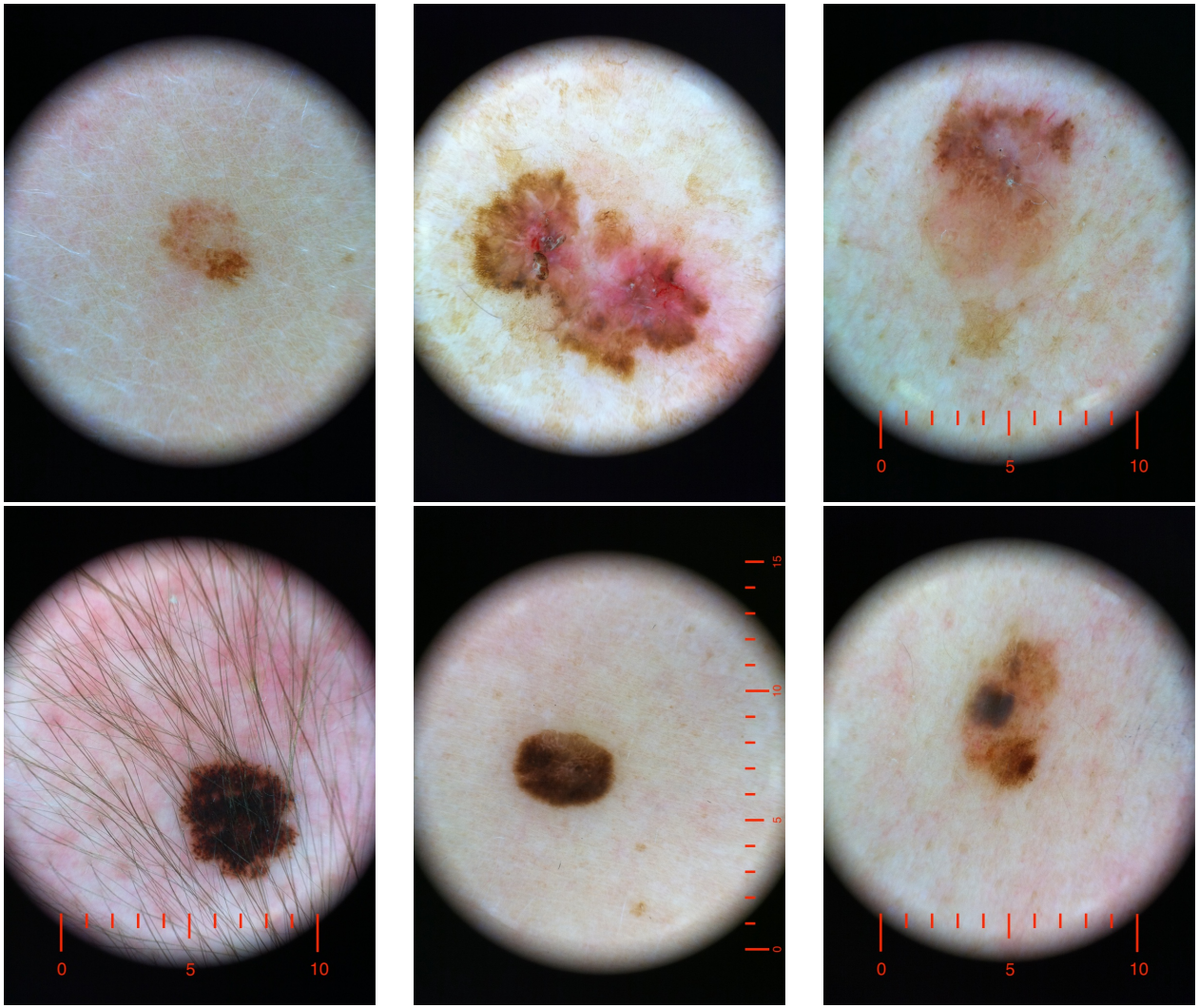
- A type of skin cancer.
- Correlation to sun bathing (UV-rays).
- Important to spot early.
- Not easy to identify.



Types of melanoma

- Benign.
  - Not dangerous.
  - A.k.a. Mole.
- Malign.
  - Lethal.
  - Grows.
  - Irregular.
  - Colorful.
  - Looks quite different from benign.
- Dysplastic.
  - As benign not dangerous.
  - Looks similar to malign.
  - Makes classification problematic.

Types of melanoma



Benign Malign Dysplastic

ABCDE

- **ABCDE**-Rule
  - **A**ssymetry.
  - **B**order.
  - **C**olor.
  - **D**iameter.
  - **E**volution.

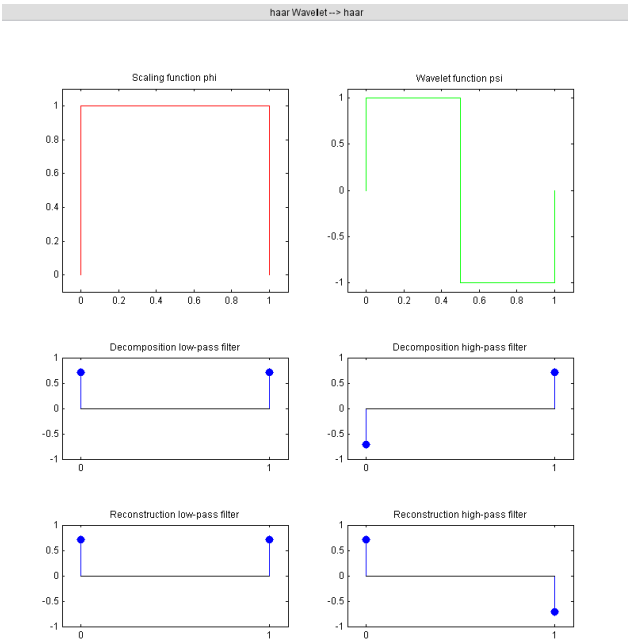
Classification

- Dysplastic and malign share similar features.
- Need lots of experience to be able to classify.
- Also use tools to aid.
  - Dermatoscope.
  - The other thing.
- Our approach: Computer aided image analysis.

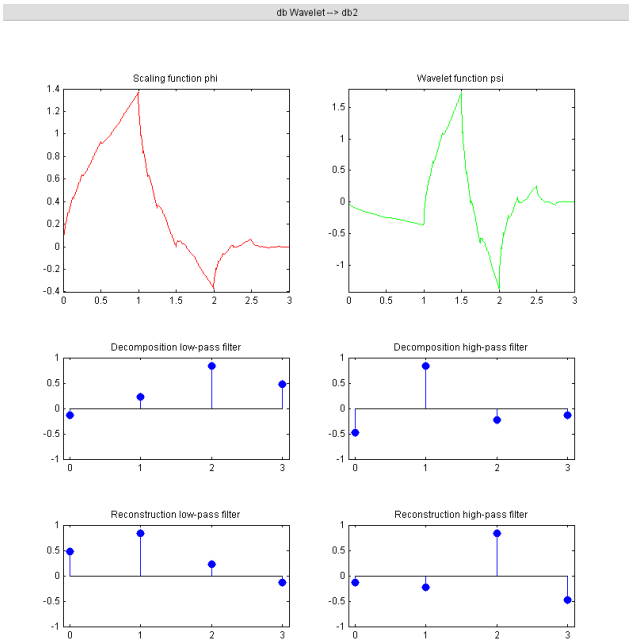
Wavlets

- Similar to fourier transform.
- Difference between consecutive resolutions.
  - Using different up/down-sampling filters.
  - Basically what we call wavelets.
  - Can be viewed as localized fourier transforms.
- Scaling function and wavelet.
  - Scaling: Low pass filter. (Approximation)
  - Wavelet: High pass filter (Detail)

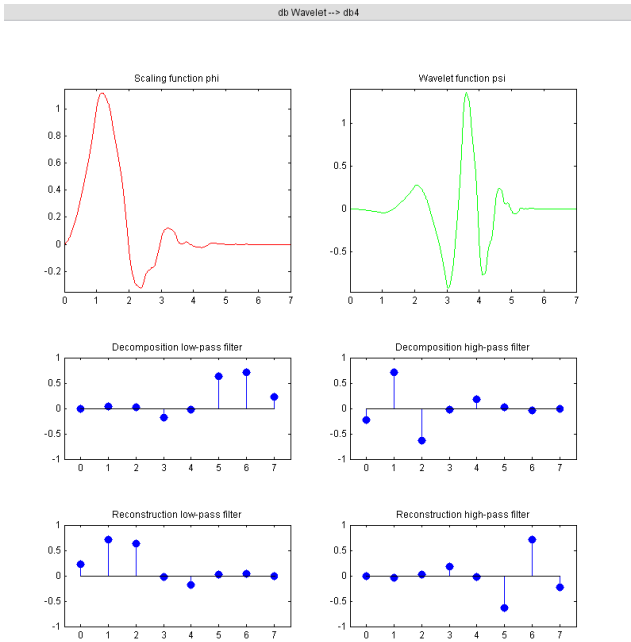
Some examples



Haar / Db1



Db2



Db4

Deccomposition for 2D arrays (images)

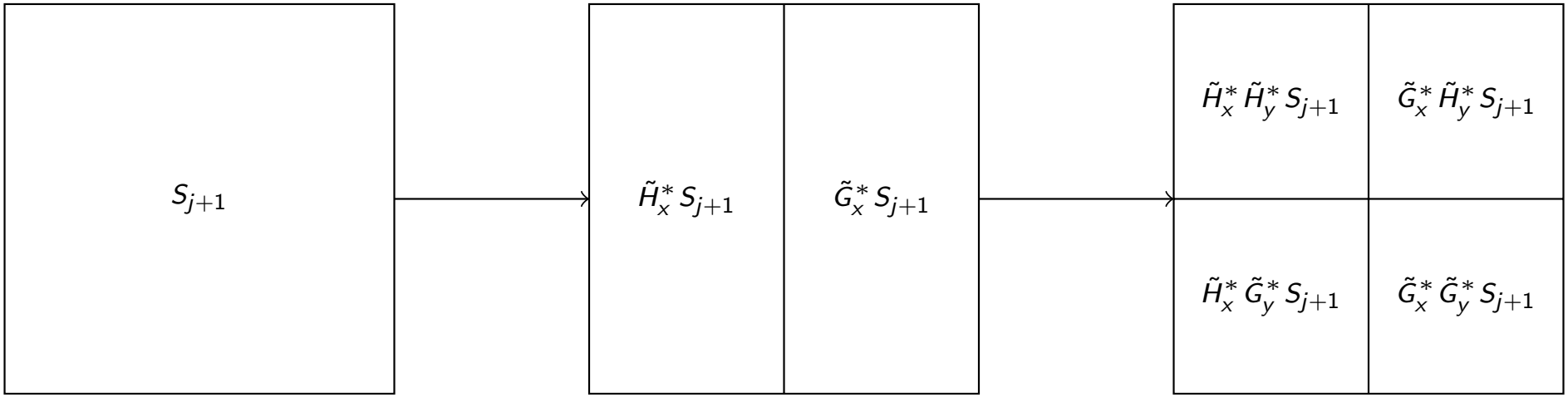


Figure: 2D Wavelet decomposition.

Deccomposition for 2D arrays (images)

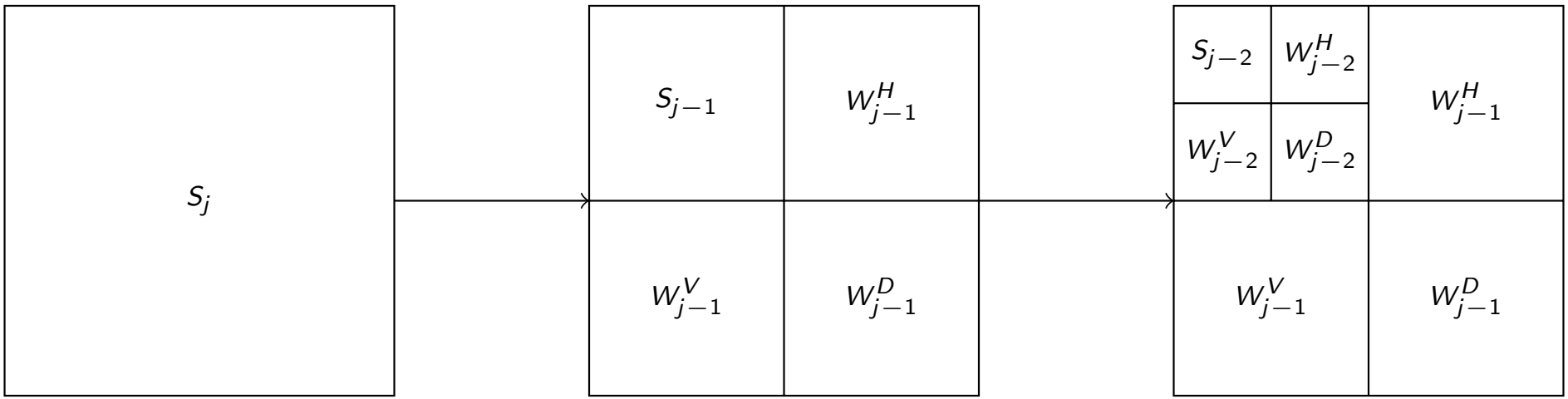
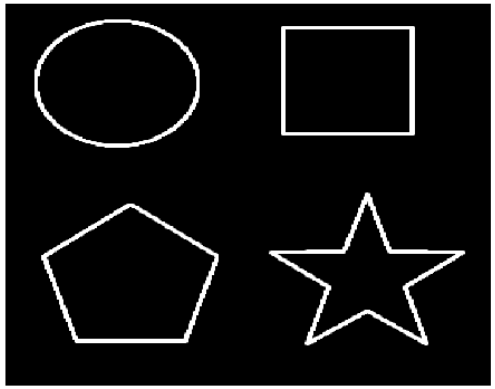


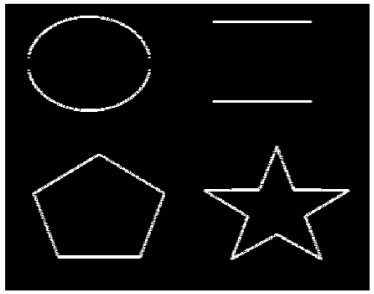
Figure: Two steps of decomposition matrix.

Edge detection

Original

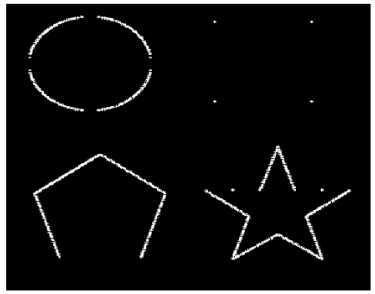


Reconstruct using only details



Horizontal details

Vertical details



Diagonal details

SVD

- Remember eigenvalues and vectors.
  - $Au = \lambda u$ [1]
    - u Eigenvector.
    - λ Eigenvalue.
- Eigen Decomposition.
  - $A = U\Lambda U^T$
    - A Symmetric.
    - Λ Diagonal. Eigenvalues.
    - U Orthogonal. Eigenvectors.
- Singular Value Decomposition
  - $A = U\Sigma V^T$
    - A not necessarily symmetric.
    - Σ Quasidiagonal. Singularvalues. Similar to eigenvalues.
    - U Orthgonal.
    - V Orthgonal.

[1] Typo in paper!

Advantages of SVD

- Don't need to calculate **all** the values.
- Need only calculate the n largest.
- *Economy* or *compact-SVD*
- Different methods for desired importance.
  - Speed.
  - Accuracy.
  - Orthogonality.
  - Number of values.

Principal Component Analysis

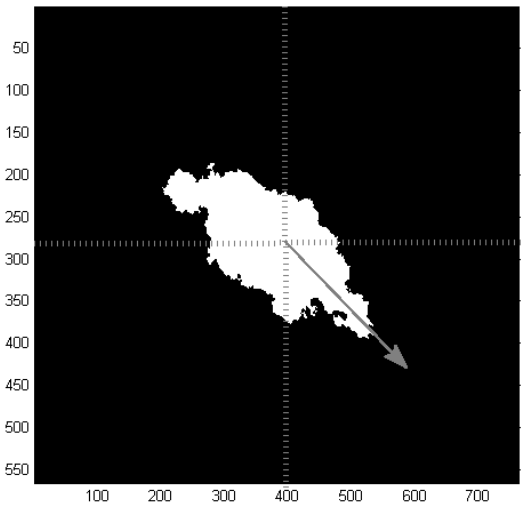
- Project data along orthogonal axes of highest variance.
  - In Decreasing order.
- Good for making the dataset more compact.
  - Use combination of variables with high variance as parameters instead.
  - Can discard the directions that does not vary a lot.
- Can be used to aid cluster identification.
  - Clusters may cause high variance along a line.
  - Don't need to find a line since it will end up in along an axis after projection.
  - Basically we do fit lines when doing the PCA.
- Other uses as well as will be shown later.

PCA + SVD = True

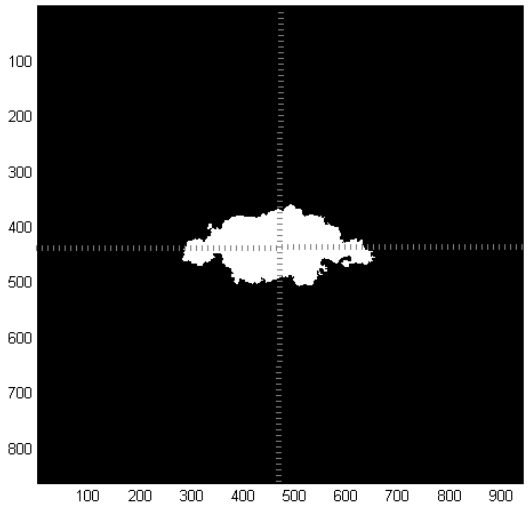
- PCA by brute force is painfully slow.
- Can use Eigen decomposition.
  - Given covariance matrix $C = XX^T$ [2].
  - X is data with zero mean.
  - $C = TDT^T$
  - Can be shown that the principal components are in Y when $Y = TX$.
  - Still have to compute XX^T.
- Even better, use SVD!
  - Can be shown that:
  - $X = U\Sigma V^T$
  - $Y = U^T X$

[2] abusing notation

Rotation



Before

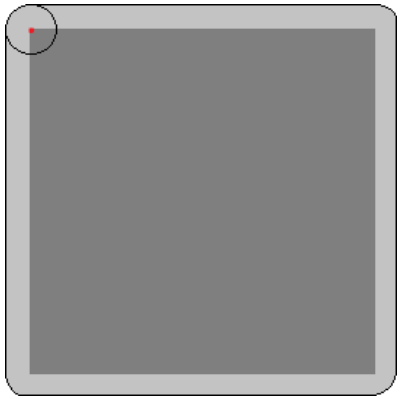


Afer

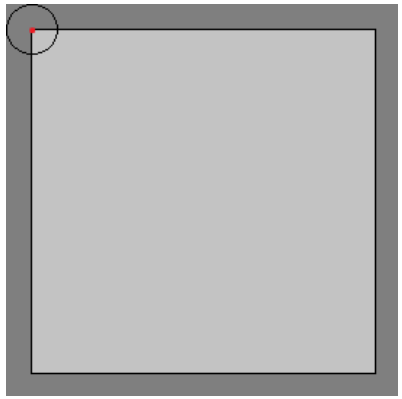
Preprocessing

- Will no go in to much detail.
- Noise removal.
  - Used median filter.
  - Lots of other options. Could have used wavelets.
- Masking out the lesion.
  - Converted to LAB-color space.
  - Used L channel for thresholding (Lightness). Lesions darker than skin.
  - Other options as well. E.g. $0.2126R + 0.7152G + 0.0722B$, common in computer graphics.
- Remove hairs.
- Clear border.
- Crop and rotate image.
  - Optimize the lesion to fill image.
  - Can also use this to compare fill rate of each quadrant (Irregularity).

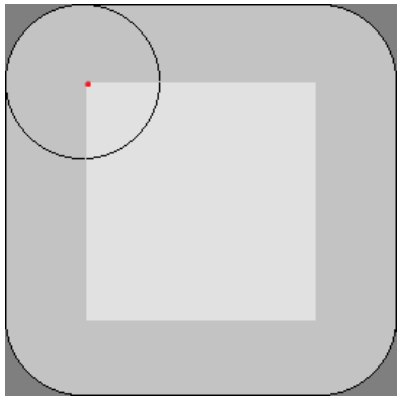
Morphological operators



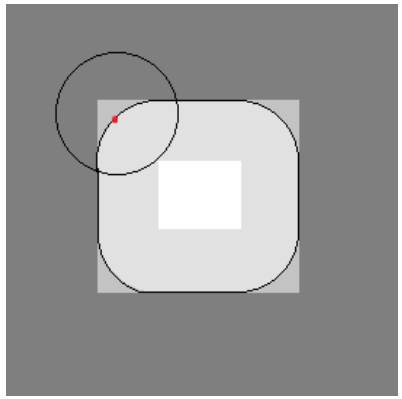
Dilation



Erosion

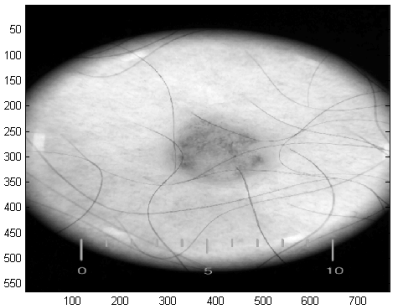


Opening

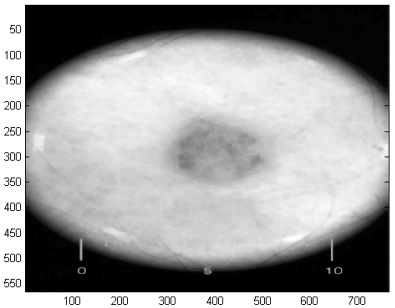


Closing

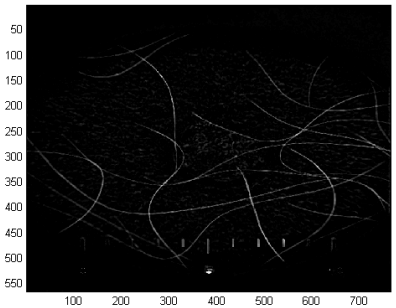
Remove hairs



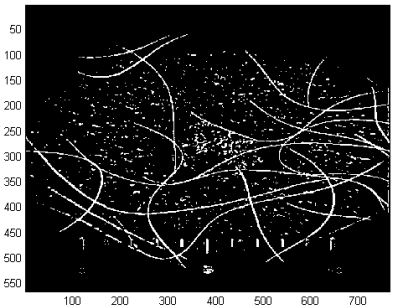
Original



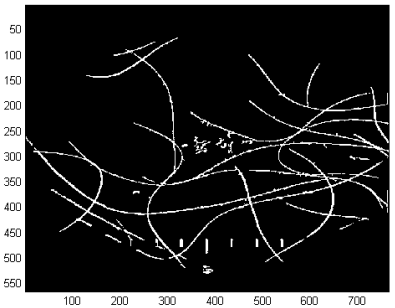
Opening



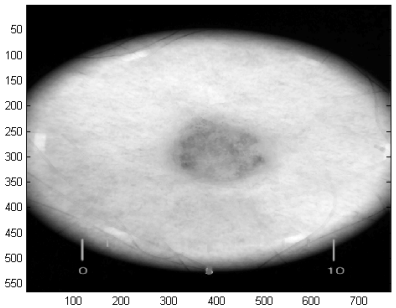
Difference



Threshold

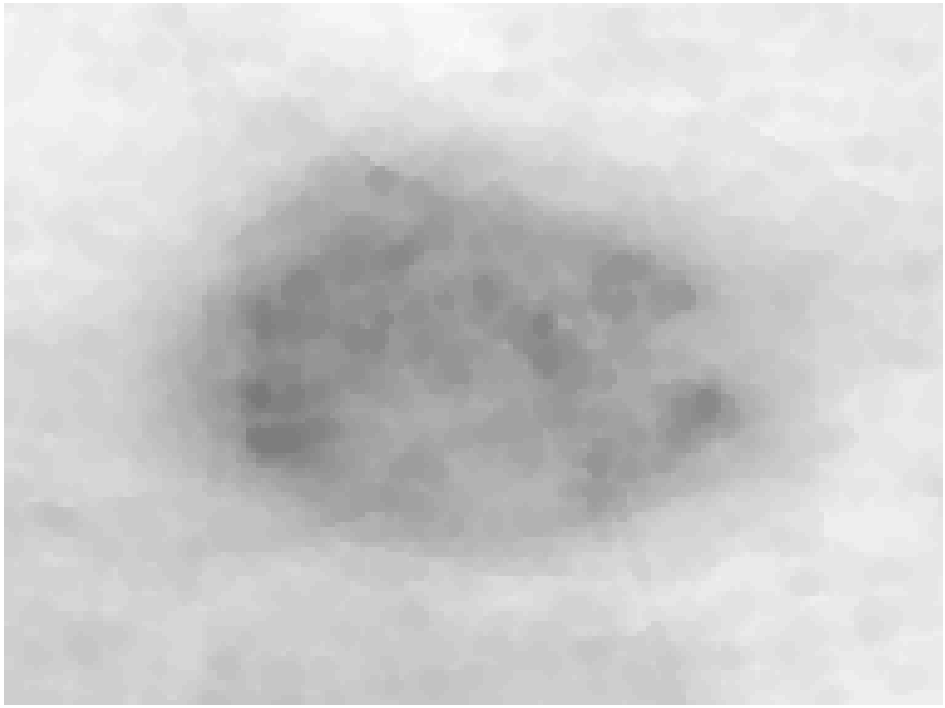


Morphological

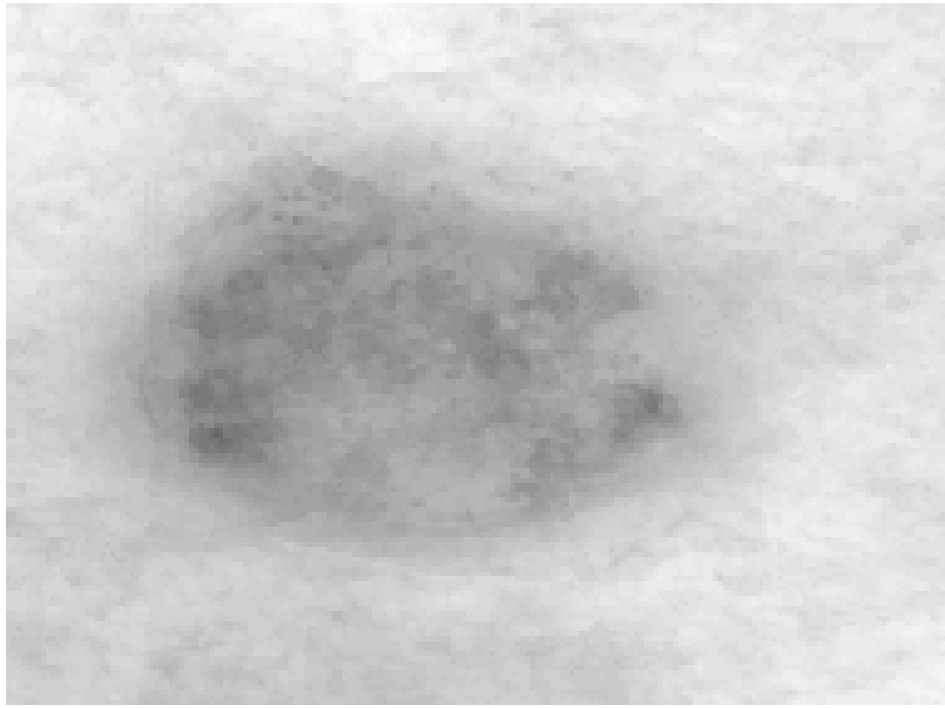


Final

Closeup

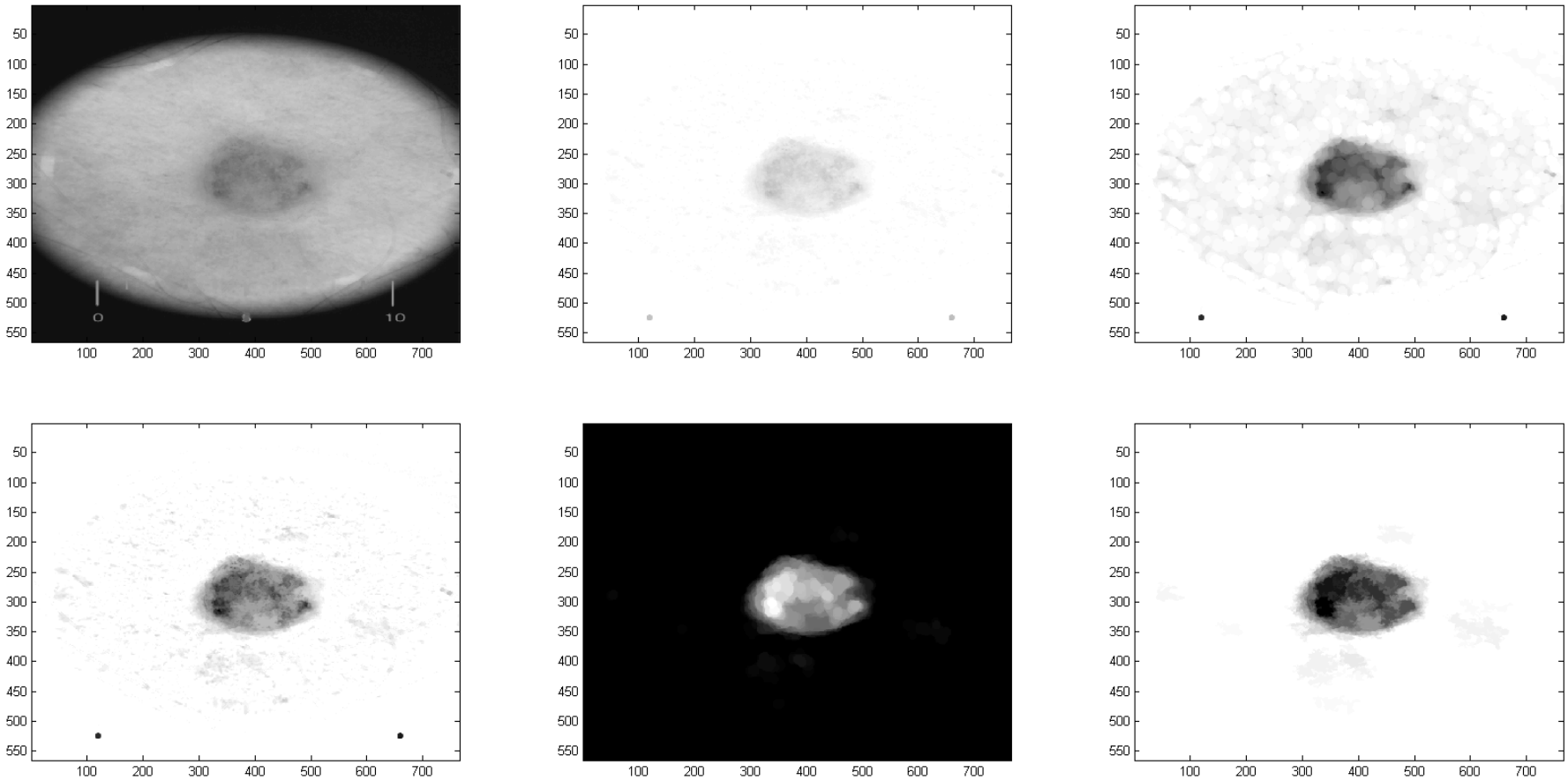


Just opening



Replcaing only masked areas

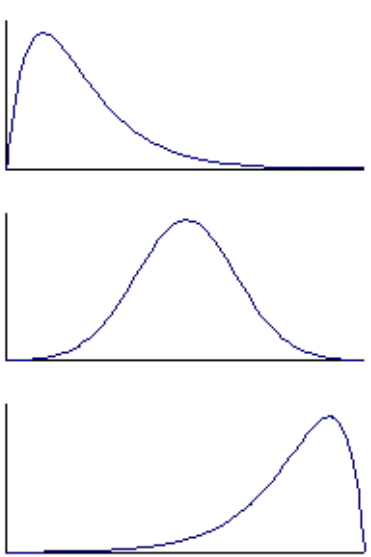
Segmentation



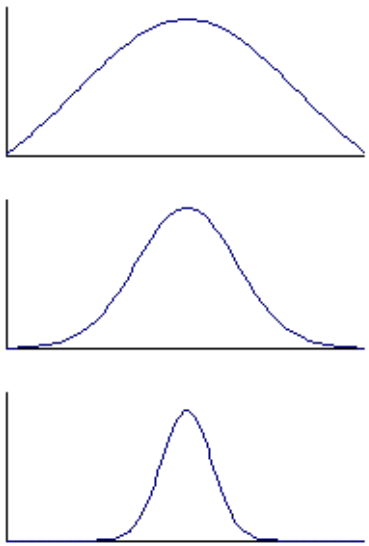
Segmentation to aid e.g. thresholding.

Skewness and kurtosis

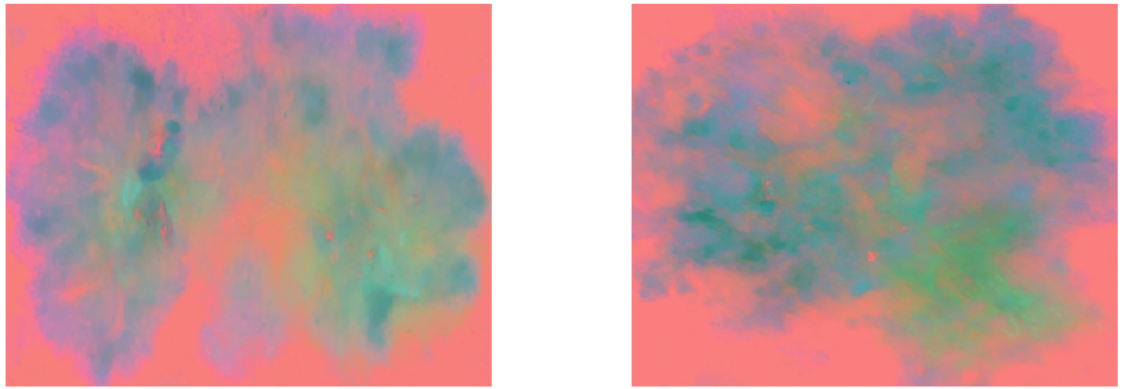
- Healthy tissue can be expected to have a normal distribution.
- That is, no exceptional skewness or kurtosis.
- Check this for wavelet coefficients for each row and col.



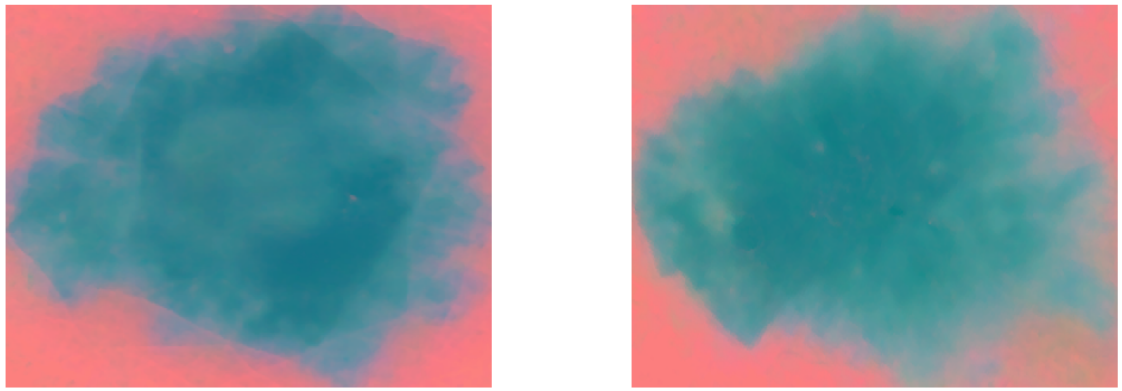
Skewness



Kurtosis



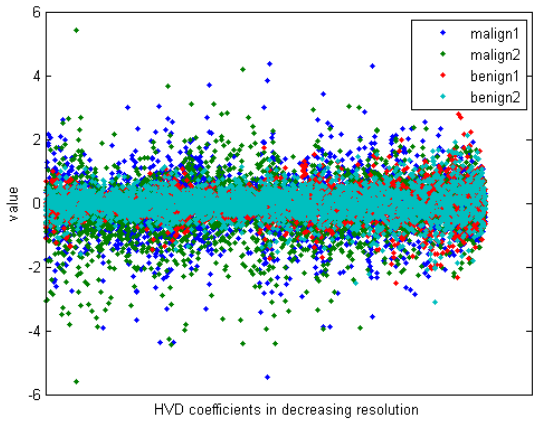
Malign



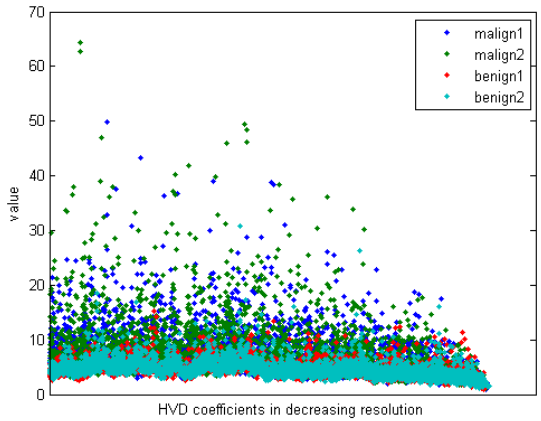
Benign

Figure: Test data in LAB-color space.

Skewness and kurtosis of wavelet coefficients



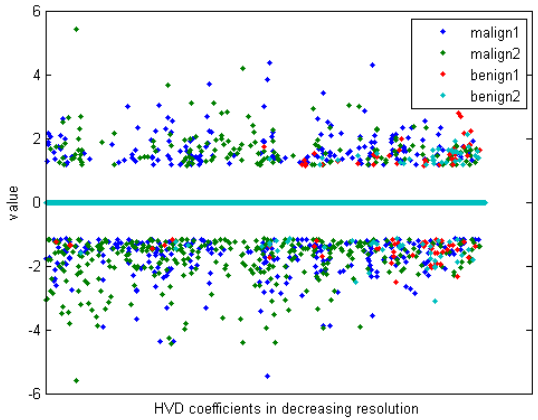
Skewness



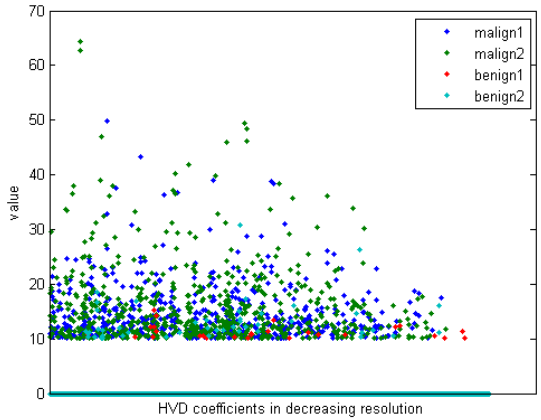
Kurtosis

DB4	Malign1	Malign2	Benign1	Benign2
Skewness absolute mean	0.5014	0.7190	0.2972	0.3189
Skewness variance	0.5242	1.0891	0.1912	0.2171
Kurtosis mean	6.9720	8.5350	5.1325	4.9342
Kurtosis variance	16.5221	43.6926	4.8227	4.6182

Table: Mean and variance of skewness and kurtosis from spectrum



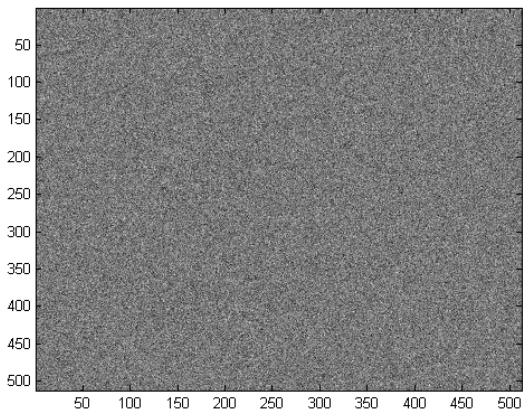
Skewness



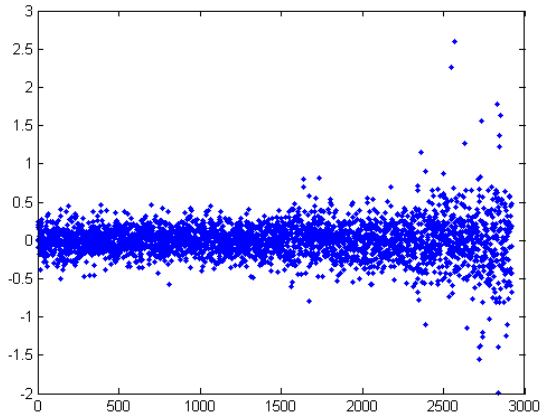
Kurtosis

DB4	Malign1	Malign2	Benign1	Benign2
Skewness absolute mean	1.8160	2.0639	1.5549	1.5725
Skewness variance	0.3633	0.8117	0.0648	0.0911
# of values above threshold	428	493	41	32
Kurtosis mean	16.7860	19.4424	14.8717	13.0884
Kurtosis variance	38.3466	78.9864	12.4312	1.4258
# of values above threshold	492	576	53	20

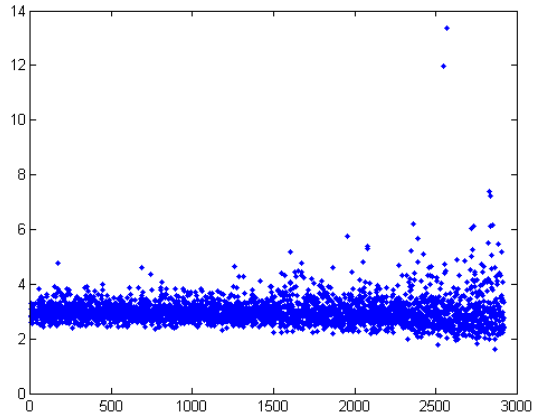
Table: Mean, variance and number of values above threshold for skewness and kurtosis from spectrum.



Normal noise



Skewness



Kurtosis

	Normal
Skewness absolute mean	0.1990
Skewness variance	0.1023
Kurtosis mean	2.9854
Kurtosis variance	0.3149

Table: Skewness and kurtosis values for the control image.

Other parameters

Irregularity $Ir := \frac{N^2}{4\pi A}$ Solidity $So := \frac{A}{C_A}$ Convexity $Co := \frac{C_N}{N}$

- N: Perimeter (Number of pixels along border).
- A: Area (Total number of pixels).
- C_A: Area of convex hull.
- C_N: Area of perimeter for convex hull.

PCA for neighbours

- Usually when using SVD with images we treat the whole image as a matrix.
- Try instead using a (3×3) neighborhood for each pixel instead.
- We get 27 bases (each pixel has 3 values, rgb).
- By re-projecting the image on each of these bases we can identify different details.
- Unfortunately didn't have time to investigate this much.
- Ended up using simple mean, variance, skewness and kurtosis for each image.

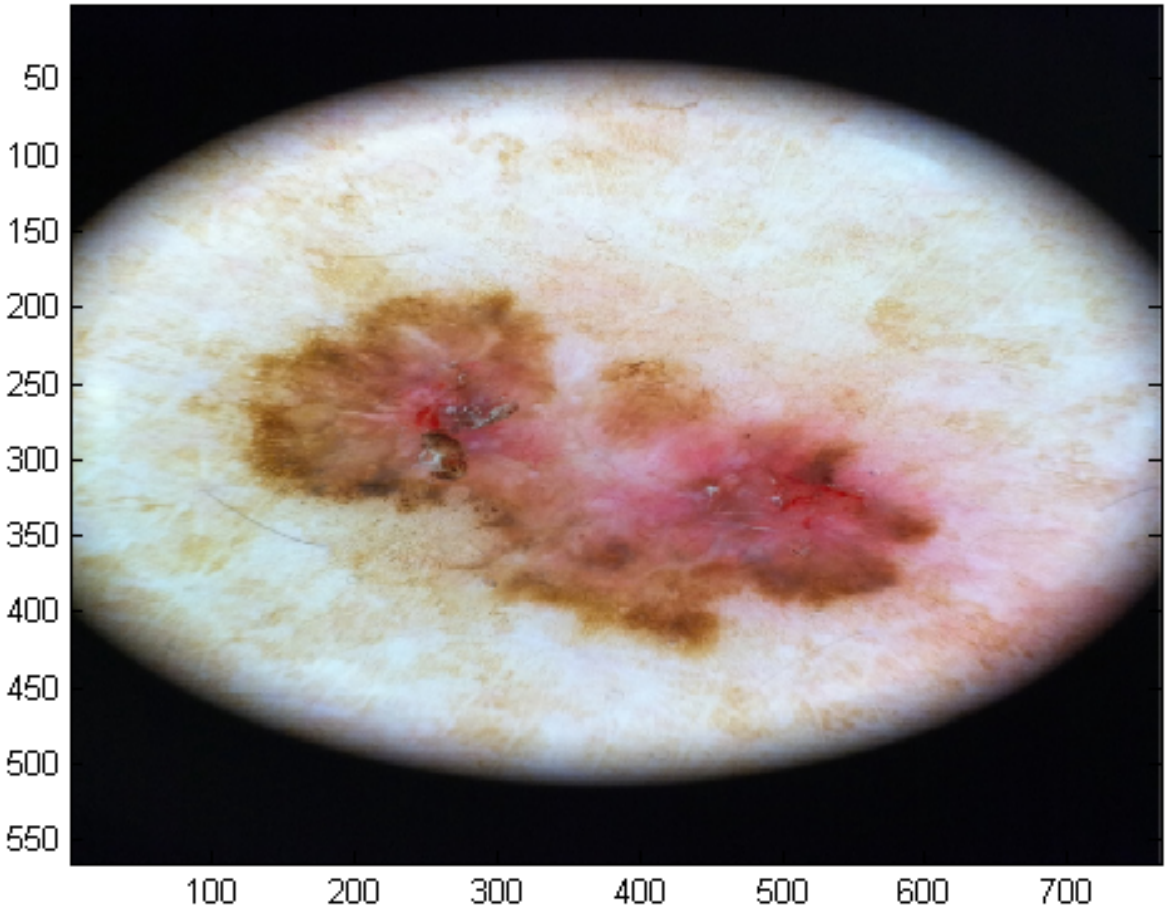


Figure: Test image.

Examples

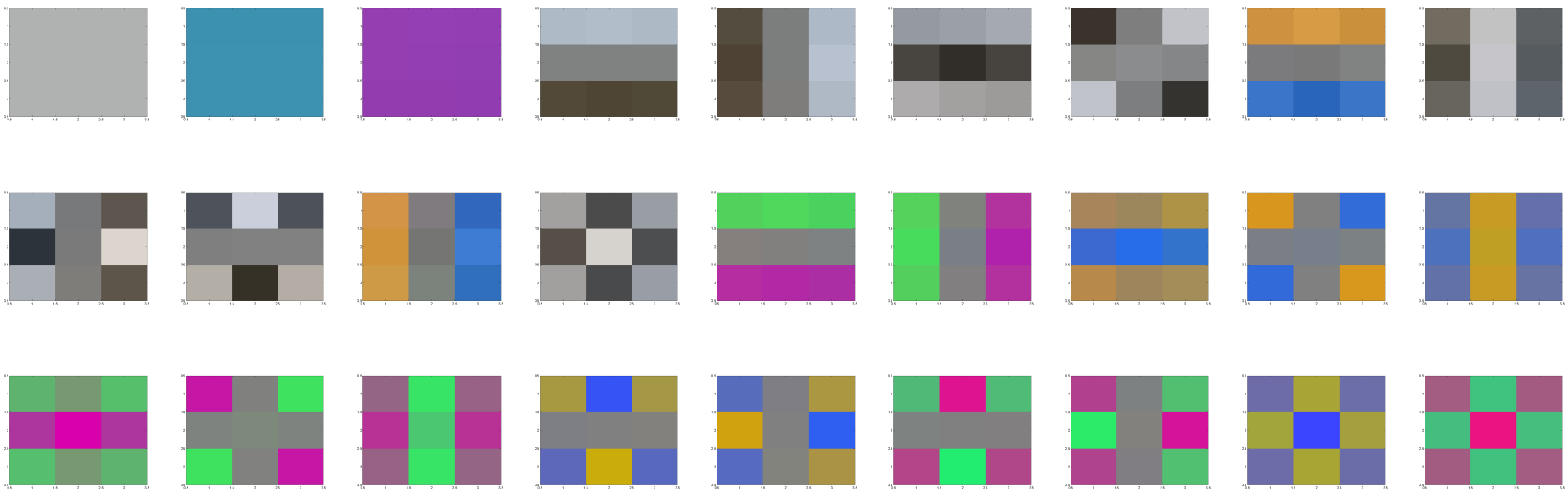


Figure: The different bases.

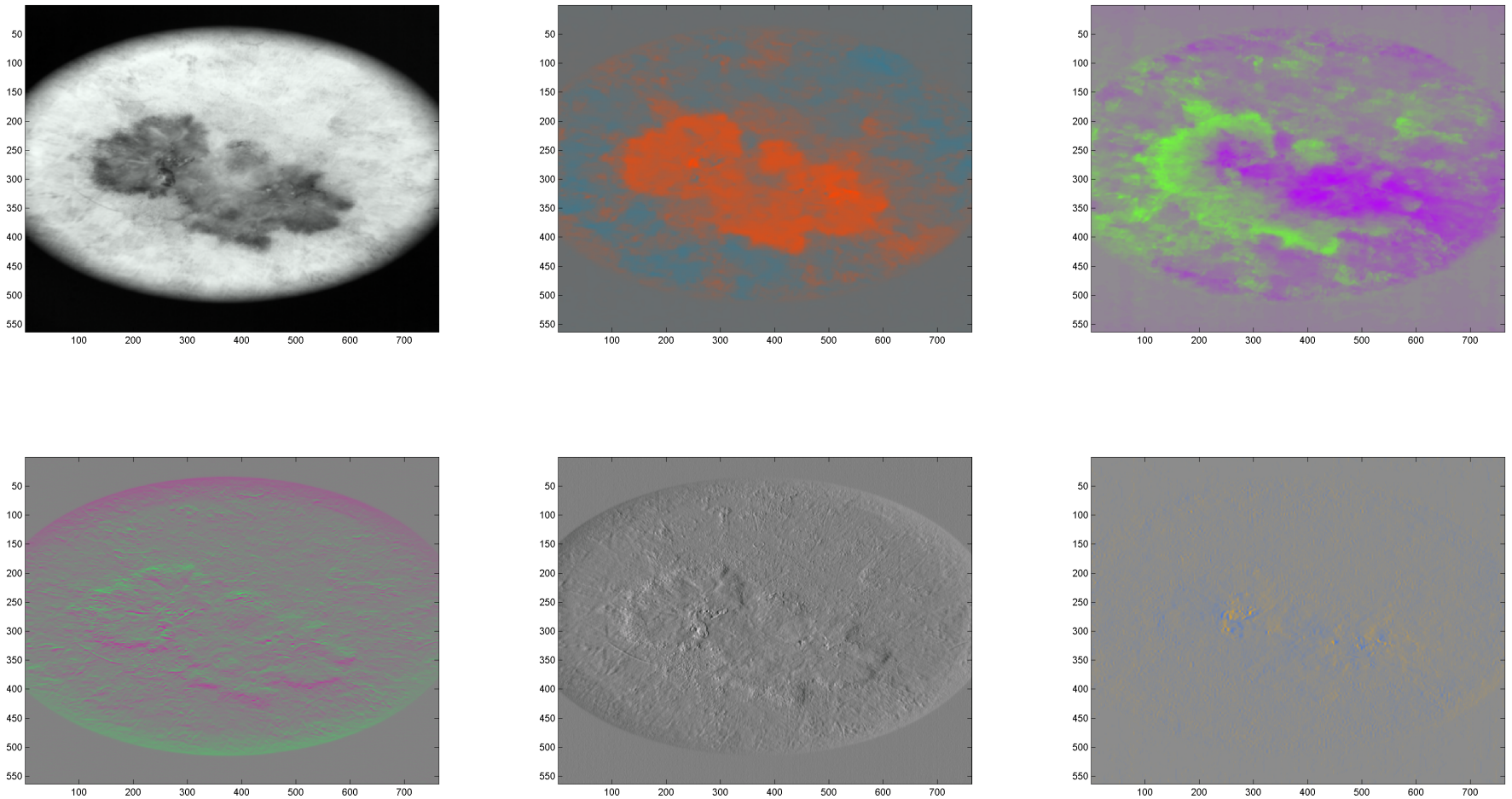


Figure: Reprojected images.

Results

- Used linear regression.
- Forward and backward selection.
  - When fitting, alternate between adding and removing parameters.

1. *Use all the data to figure out what parameters to include.*
2. *Take e.g. 70% as training data and 30% as test.*
3. *Fit the training data to the model acquired in step one using simple regression.*
4. *Evaluate the test data and save results.*
5. *Go to step 2. Stop after n steps or until convergence of e.g. the mean squared error of the model or the misclassification rate.*

- Sensitivity
  - Percentage of correctly identified malign lesions
- Specificity
  - Percentage of correctly identified benign lesions
- Can tweak classification threshold to favor one over the other.

Final remarks and results

- By tweaking the input data (Procedure steps, transforming parameters) managed to fairly easily get aroun 80% in specificity and sensitivity.
- Some best results up to 93%.
- Don't trust model too much!
- Many parameters per data-points (10:100).
- Not homogeneous data. Had to manipulate.
  - Malign images had higher resolution.
  - Had to downscale.
  - Usually some polynomial is used to avoid aliasing.
  - Could possibly affect wavelets (Some have close relations to polynomials, "vanishing moments").
- However consistent results show that there is information which is worth investigating.

I have spoken!