Steganalysis with a Computational Immune System

By

Jacob Jackson, Gregg Gunsch, Roger Claypoole, Gary Lamont

Presented At

The Digital Forensic Research Conference

DFRWS 2002 USA Syracuse, NY (Aug 6th - 9th)

DFRWS is dedicated to the sharing of knowledge and ideas about digital forensics research. Ever since it organized the first open workshop devoted to digital forensics in 2001, DFRWS continues to bring academics and practitioners together in an informal environment. As a non-profit, volunteer organization, DFRWS sponsors technical working groups, annual conferences and challenges to help drive the direction of research and development.

http:/dfrws.org

Air Force Institute of Technology

Educating the World's Best Air Force

U.S. AIR FORCE

Blind Steganography Detection Using a Computational Immune System Approach: A Proposal

Capt Jacob T. Jackson
Gregg H. Gunsch, Ph.D
Roger L. Claypoole, Jr., Ph.D
Gary B. Lamont, Ph.D

Integrity - Service - Excellence

Overview

- Research goal
- Wavelet analysis background
- Computational Immune Systems (CIS) background and methodology
  - Genetic algorithms (GAs)
- Research concerns

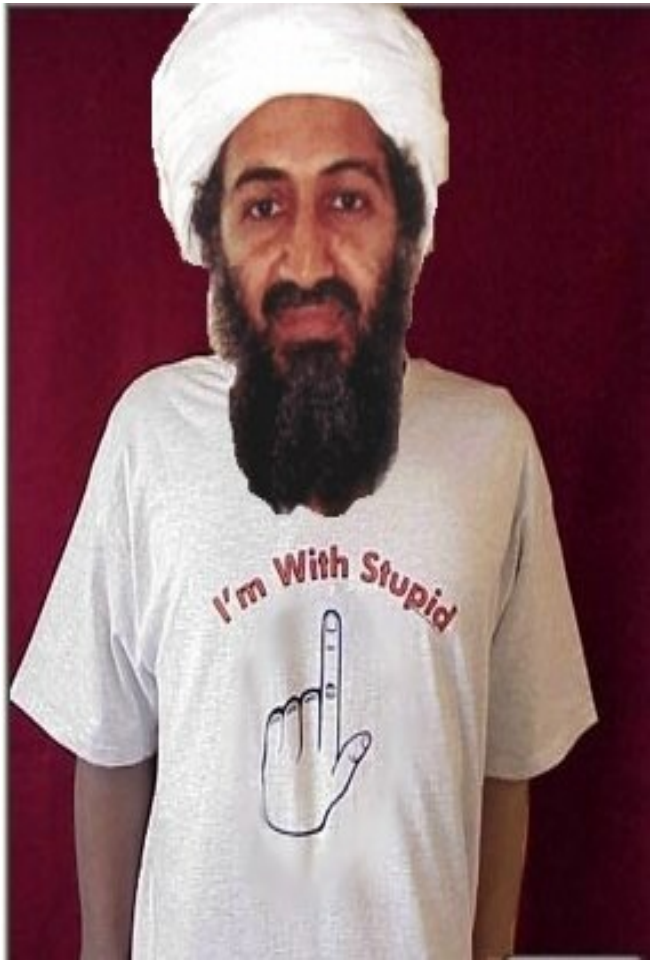
Motivation

"Lately, al-Qaeda operatives have been sending hundreds of encrypted messages that have been hidden in files on digital photographs on the auction site eBay.com….The volume of the messages has nearly doubled in the past month, indicating to some U.S. intelligence officials that al-Qaeda is planning another attack."

- *USA Today*, 10 July 2002

"Authorities also are investigating information from detainees that suggests al Qaeda members -- and possibly even bin Laden -- are hiding messages inside photographic files on pornographic Web sites."

- *CNN,* 23 July 2002



Research Goal

Develop CIS classifiers, which will be evolved using a GA, that distinguish between clean and stego images by using statistics gathered from a wavelet decomposition.

- Out of scope
  - Development of a full CIS
  - Embedded file size or stego tool prediction
  - Embedded file extraction

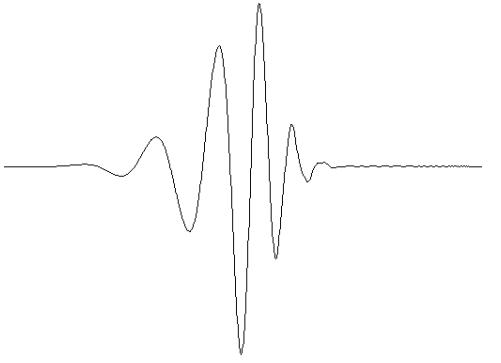
Farid's Research

- Gathered statistics from wavelet analysis of clean and stego images
- Fisher linear discriminant (FLD) analysis
- Tested Jpeg-Jsteg, EzStego, and OutGuess
- Results
  - Jpeg-Jsteg detection rate 97.8% (1.8% false +)
  - EzStego detection rate 86.6% (13.2% false +)
  - OutGuess detection rate 77.7% (23.8% false +)
- Novel images, but known stego tool

Ref: Fari

Wavelet Analysis

- Scale - compress or extend a mother wavelet
  - Small scale (compress) captures high frequency
  - Large scale (extend) captures low frequency
- Shift along signal
- Wavelet coefficient measures similarity between signal and scaled, shifted wavelet - filter
- Continuous Wavelet Transform (CWT)



Mother Wavelet

Small Scale

Large Scale

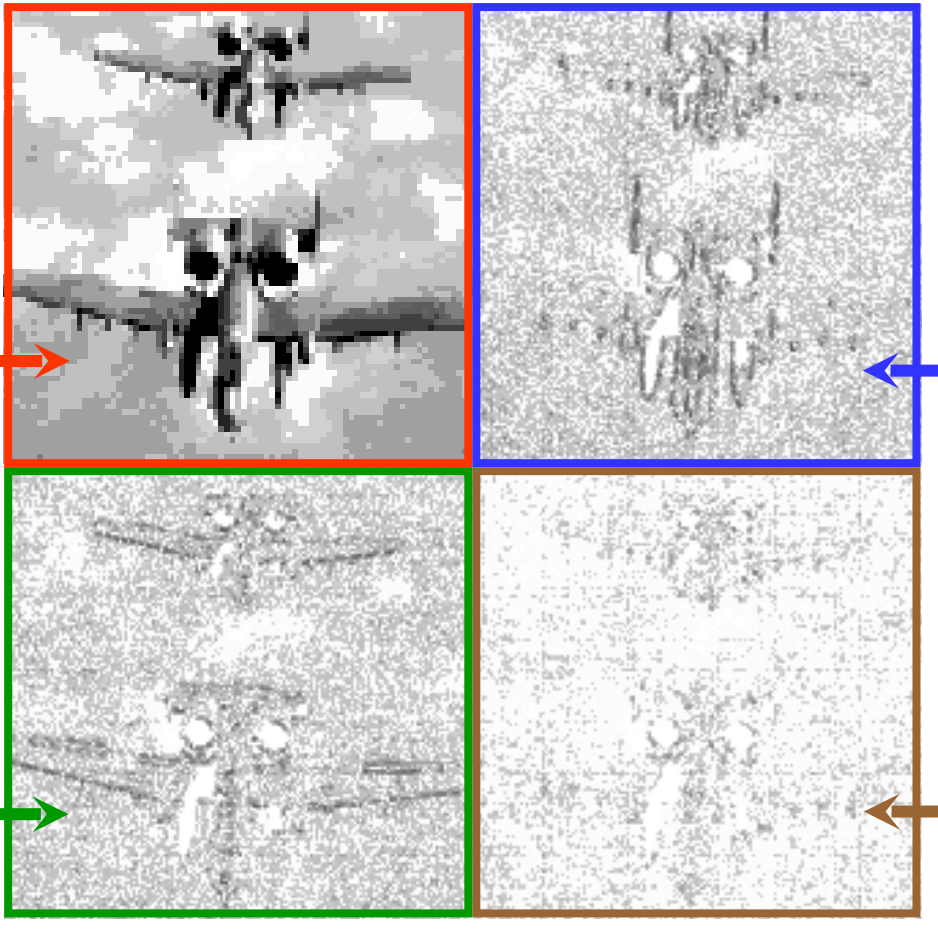
Ref: Hubb, Riou

Wavelet Analysis

- Discrete Wavelet Transform (DWT)
  - Wavelet function ⌈
    - Implemented with unique high pass filter
    - Wavelet coefficients capture signal details
  - Scaling function ⌈
    - Implemented with unique low pass filter
    - Scaling coefficients capture signal approximation
  - Shifting and scaling by factors of two (dyadic) results in efficient and easy to compute decomposition
  - For images apply specific combinations of ⌈ and ⌈ along the rows and then along the columns

Ref: Hubb, Riou

Wavelet Analysis



LL subband (approximation)

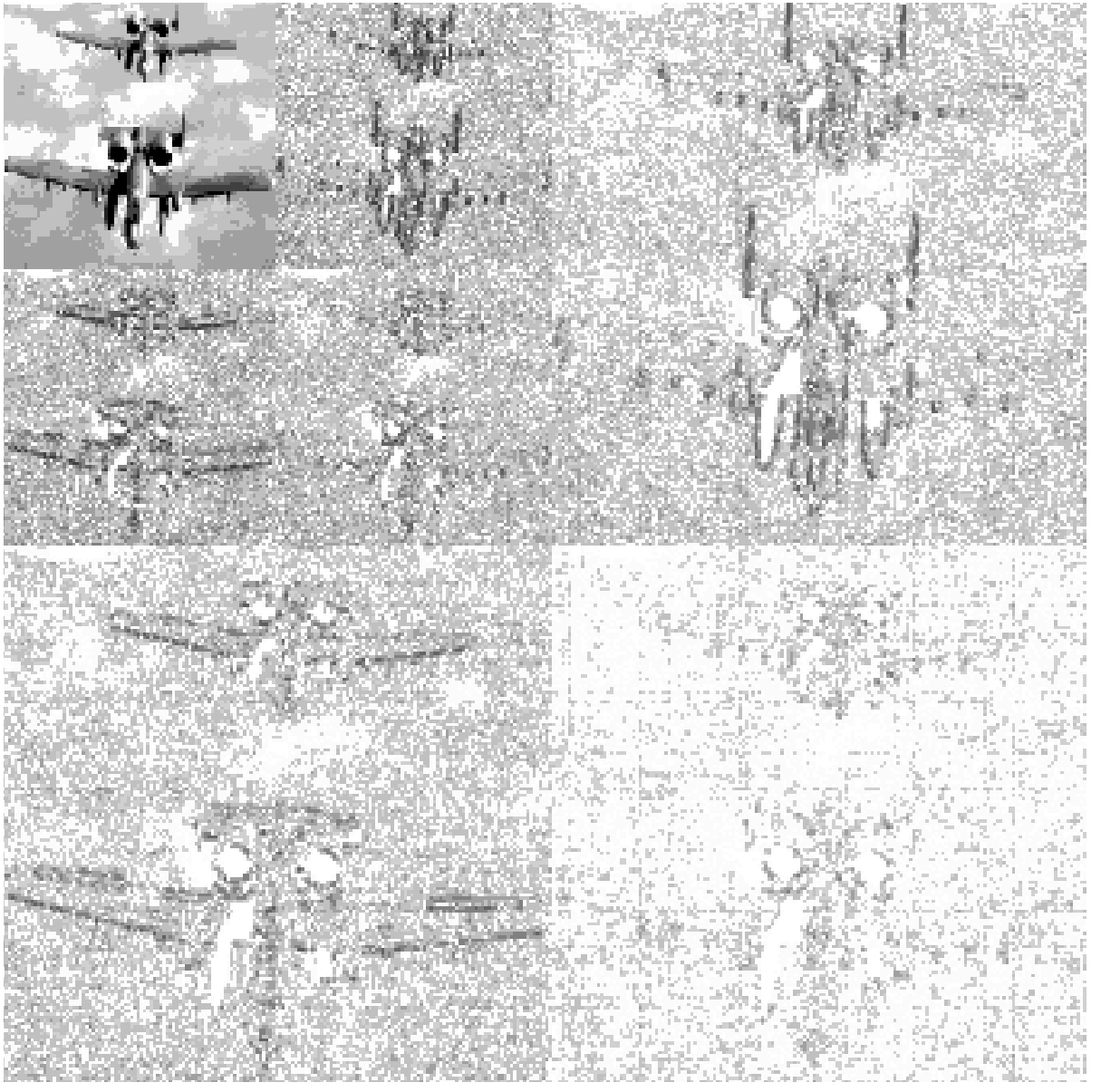
HL subband (vertical edges)

LH subband (horizontal edges)

HH subband (diagonal edges)

Ref: Mend

Wavelet Analysis



Wavelet Statistics

- Mean, variance, skewness, and kurtosis of wavelet coefficients at LH, HL, HH subbands for each scale
- Same statistics on the error in wavelet coefficient predictor
  - Use coefficients from nearby subbands and scales
  - Linear regression to predict coefficient
  - Can predict because coefficients have clustering and persistence characteristics
- 72 statistics

	1	2	...	72
Image 1	1011	1100	...	0010
Image 2	0010	1010	...	1000
⋮				

Ref: Fari

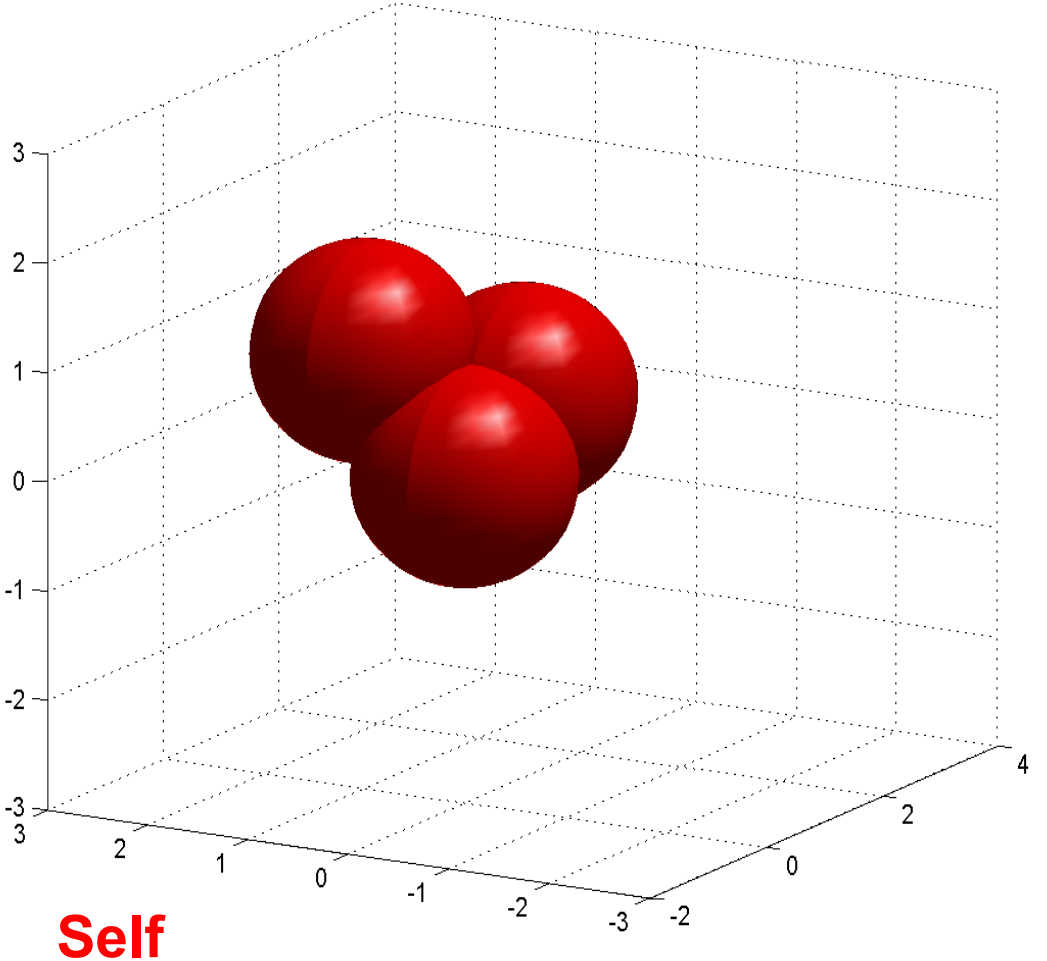
Computational Immune System

- Model of biological immune system
- Attempts to distinguish between ***self*** and ***nonself***
  - Self - allowable activity
  - Nonself - prohibited activity
- Definitions of self and nonself drift over time
- Ways of distinguishing between self and nonself
  - Pattern recognition - FLD
  - Neural networks
  - Classifier (also called antibody or detector)

Ref: Will

Self and Nonself

- Self - hypervolume represented by clean image wavelet statistics
- Nonself - everything else



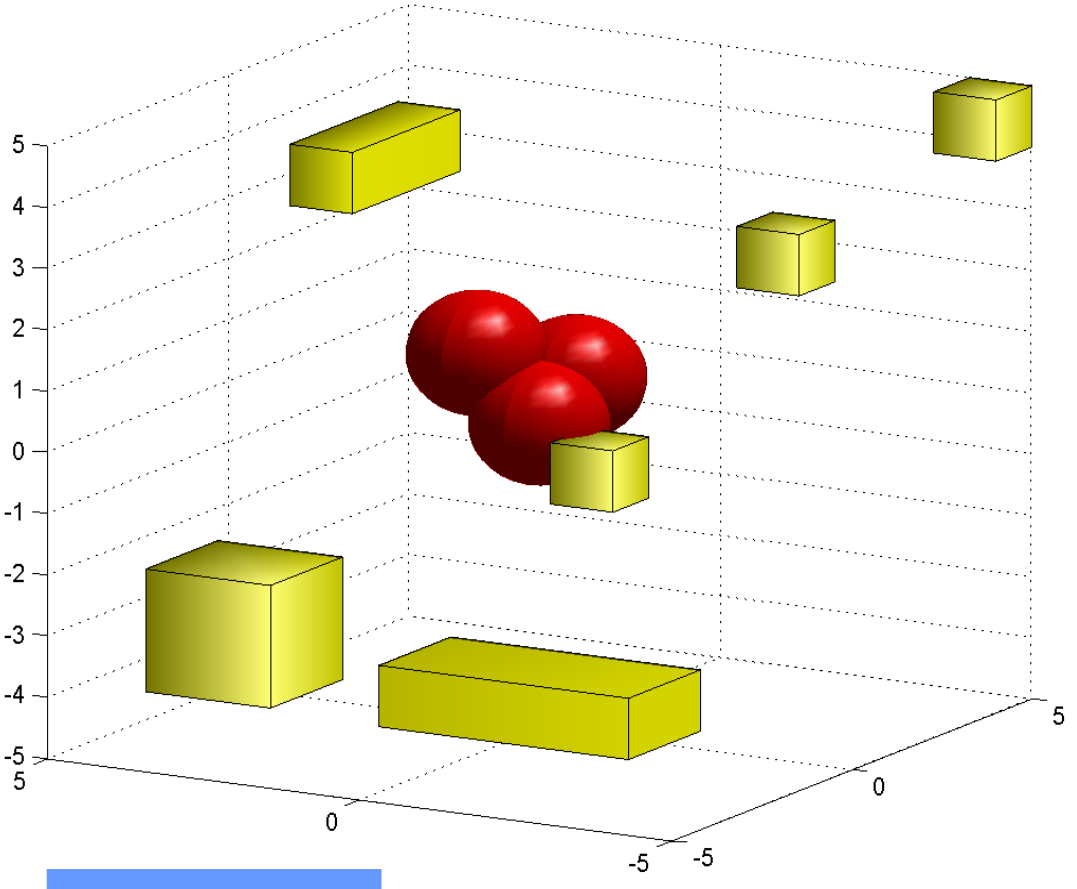
Self

Nonself - everything else

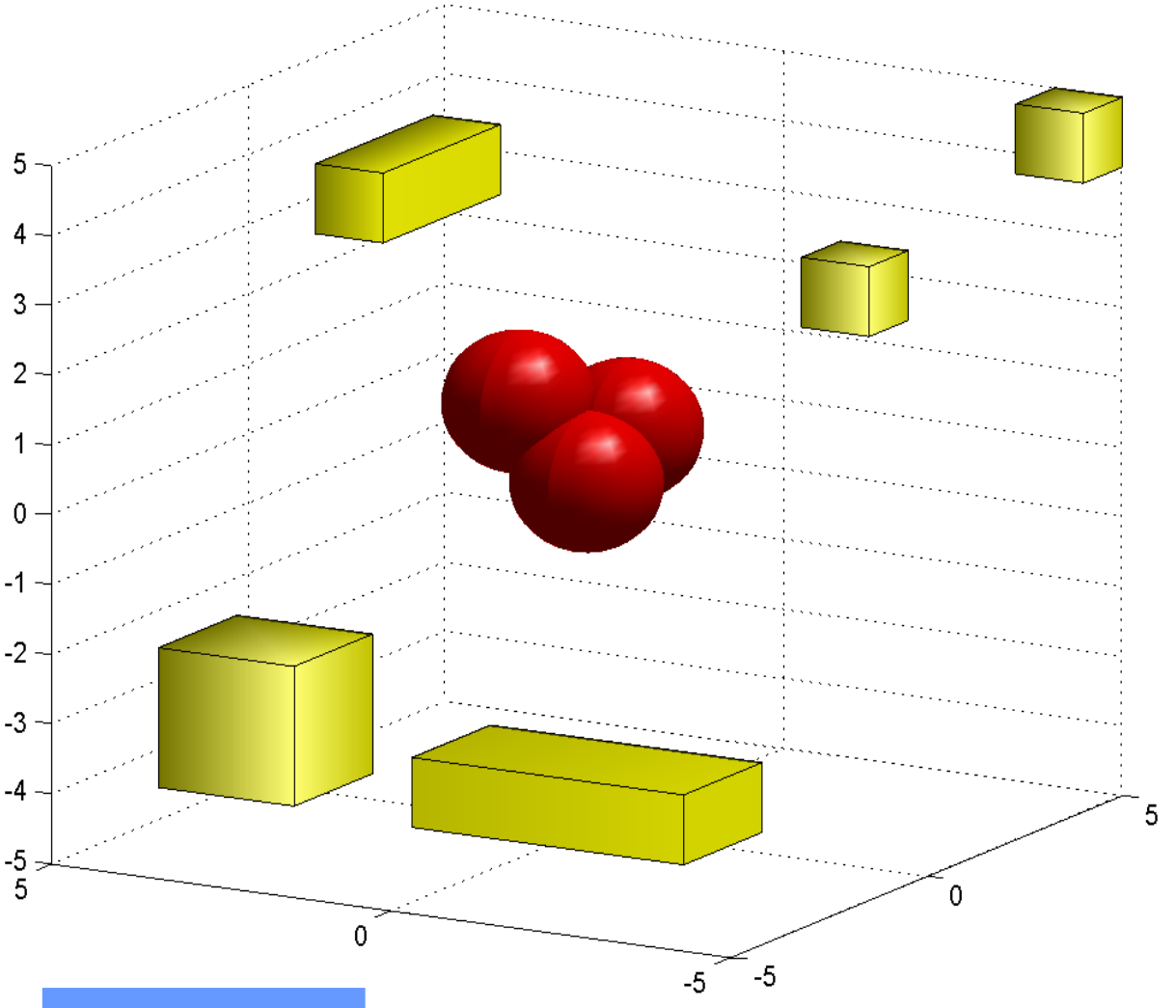
Classifiers

- Randomly generated
- Location,range, and mask
- Might impinge on self

	1	2	...	72
Location	1011	1100	...	0010
Range	1111	1010	...	1000
Mask	0	1	...	1



After Negative Selection

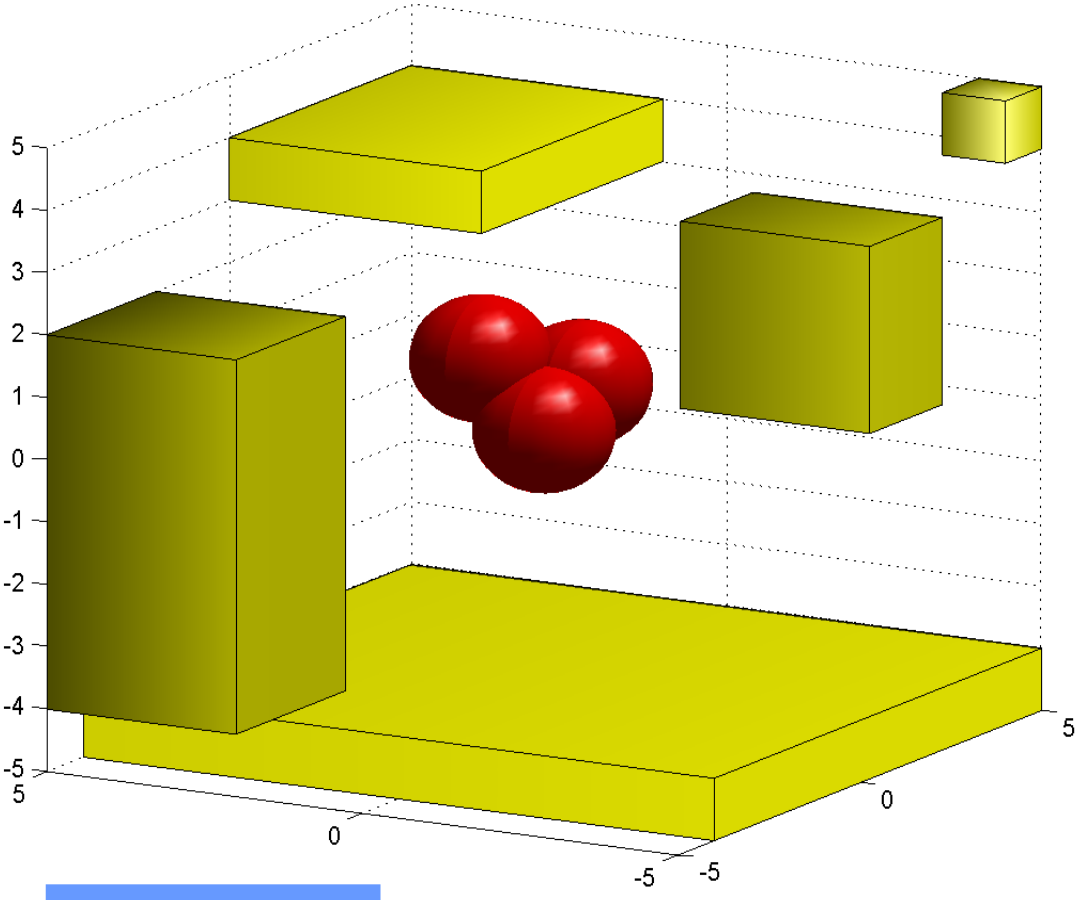


Self

Classifiers

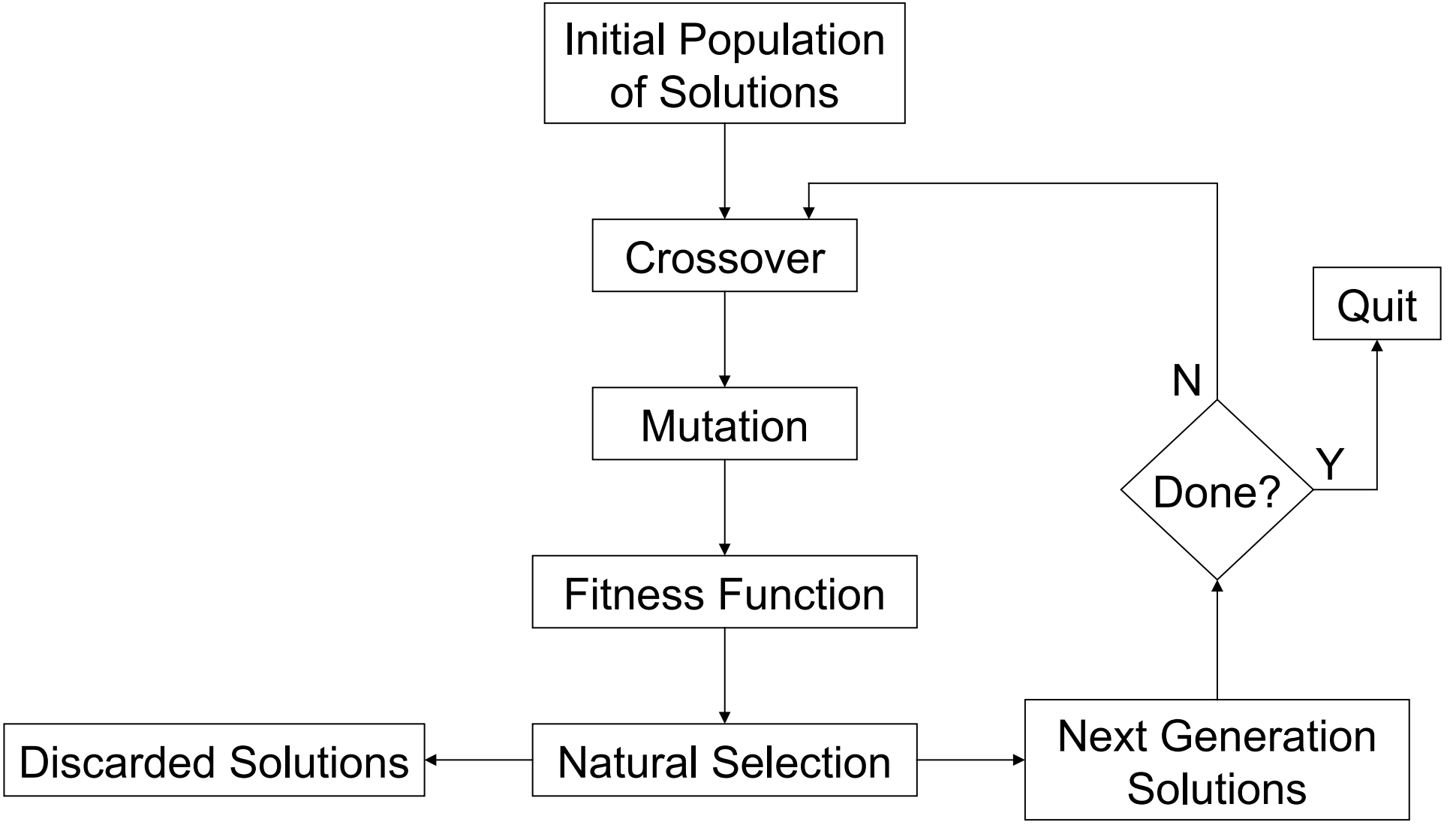
Affinity Maturation

- Goal is to make classifiers as large as possible without impinging on self
- Done using a GA
  - Multi-directional search for best solution(s)
  - Crossover - exchanges information between solutions
  - Mutation - slow search of solution space
  - Fitness function - reward growth and penalize impinging on self
  - Natural selection - keep the best classifiers



Ref: BeasA, BeasB

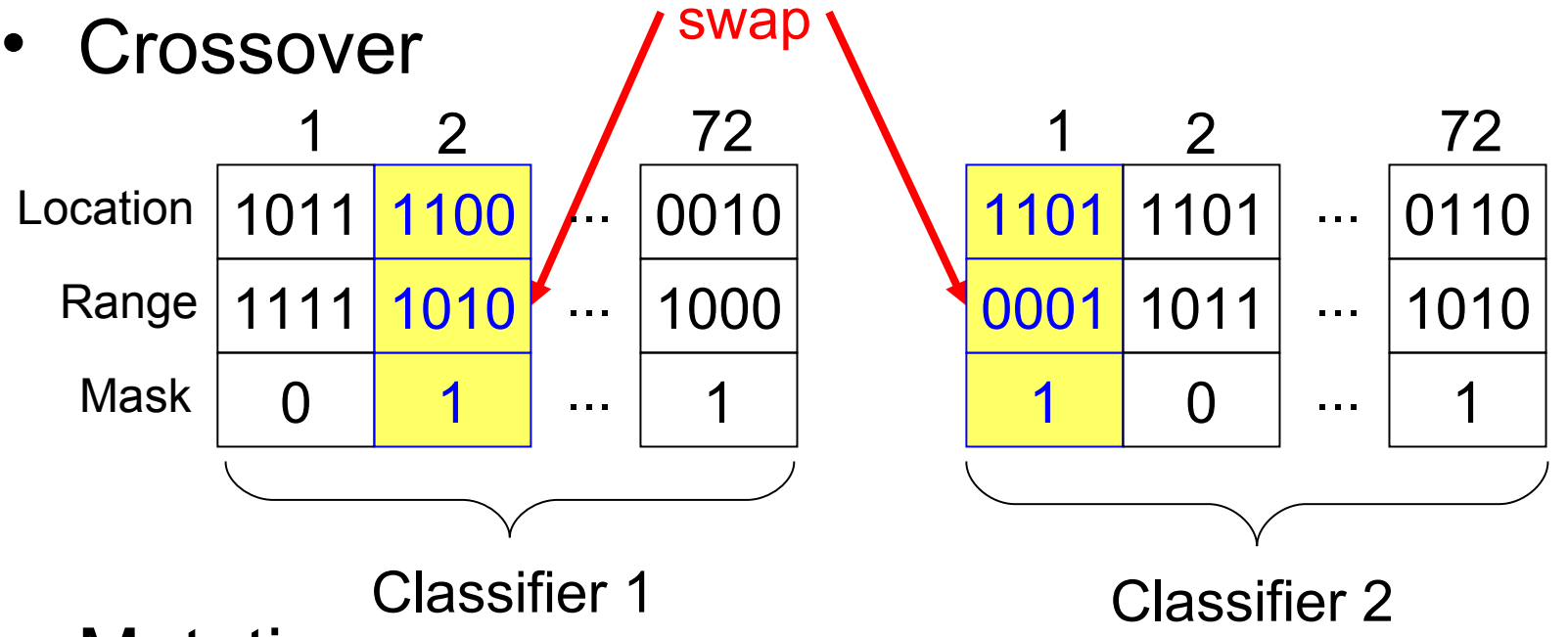
GA



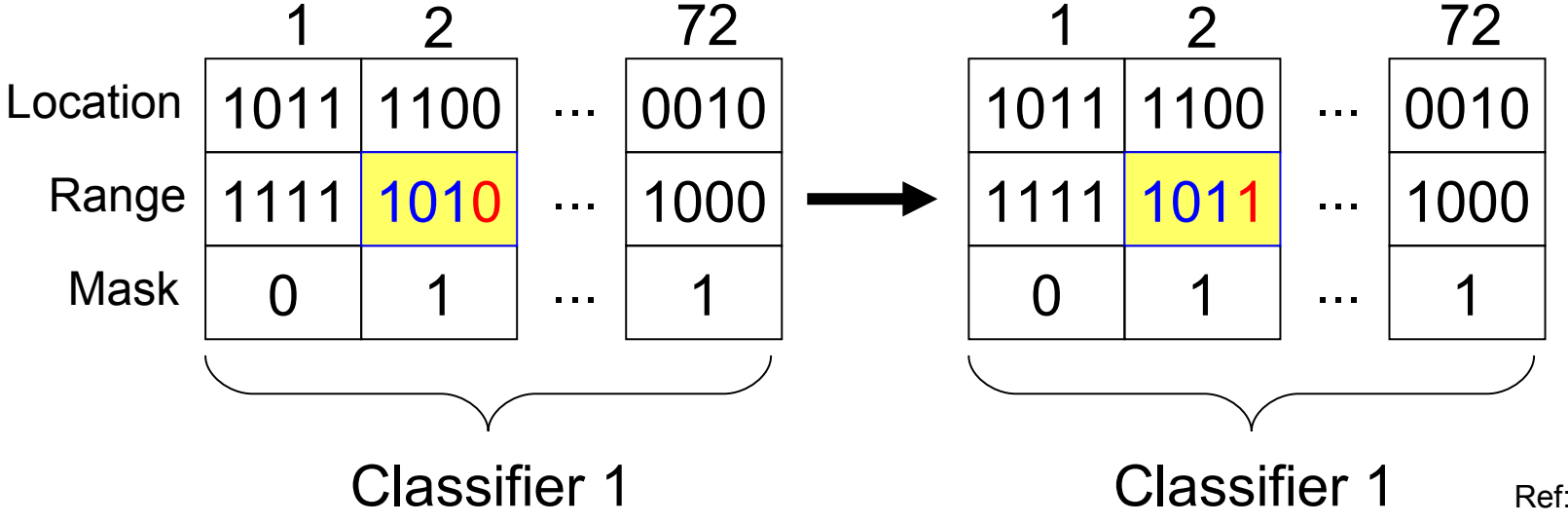
Ref: BeasA, BeasB

GA

- Crossover



- Mutation



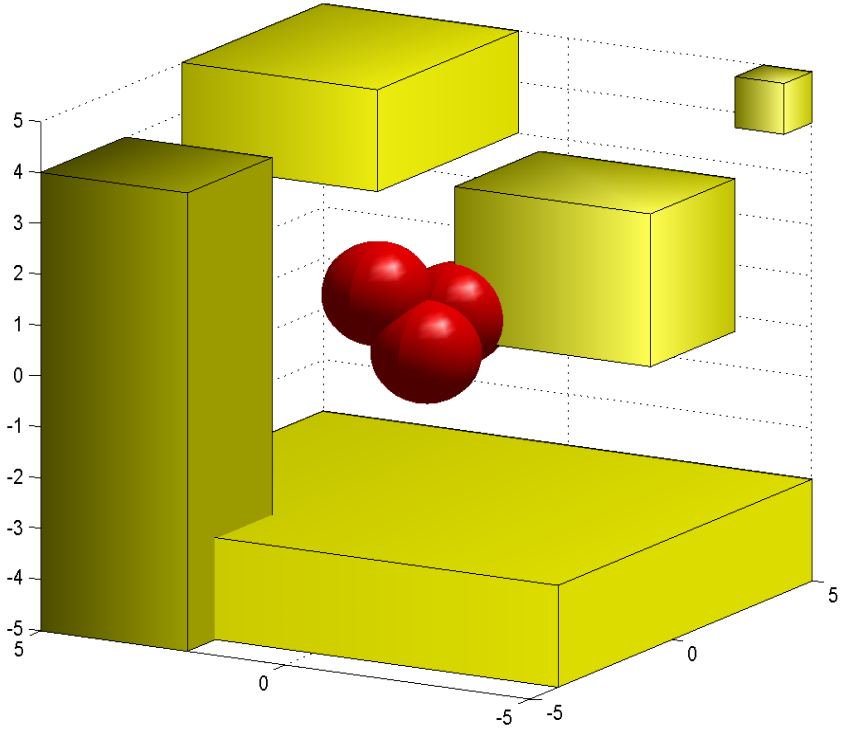
Ref: BeasA, BeasB

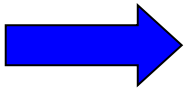
GA

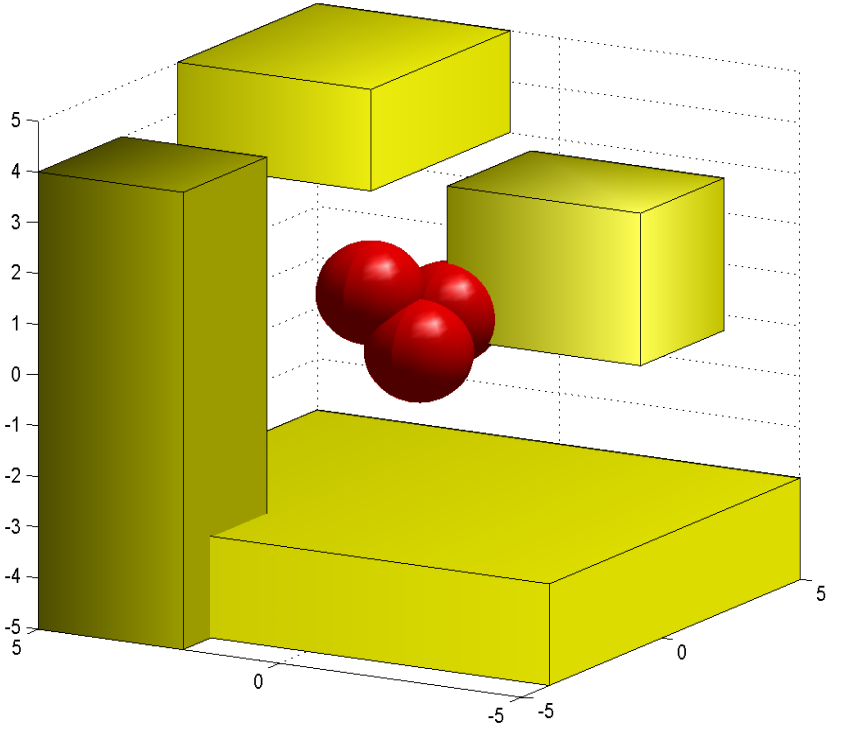
- Fitness function
  - Assign a fitness score - classifier with largest volume without impinging on self gets greatest score
  - Multiobjective approach
- Natural selection - binary tournament selection with replacement
  - Randomly select two classifiers to participate in tournament
  - Compare fitness scores – best goes on to next generation
  - Place both classifiers back in tournament pool
  - Maintains diversity in generations

Ref: BeasA, BeasB

Natural Selection

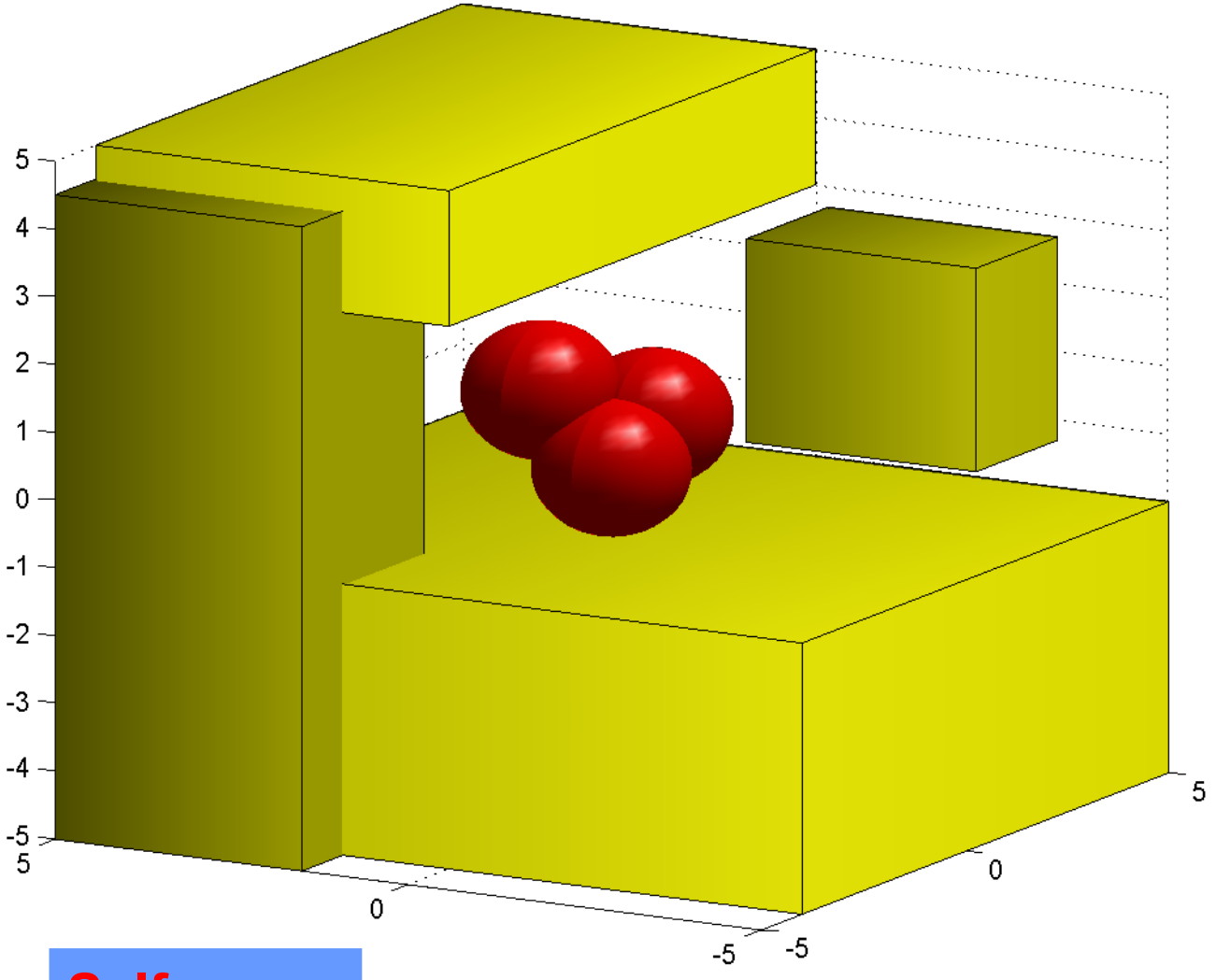






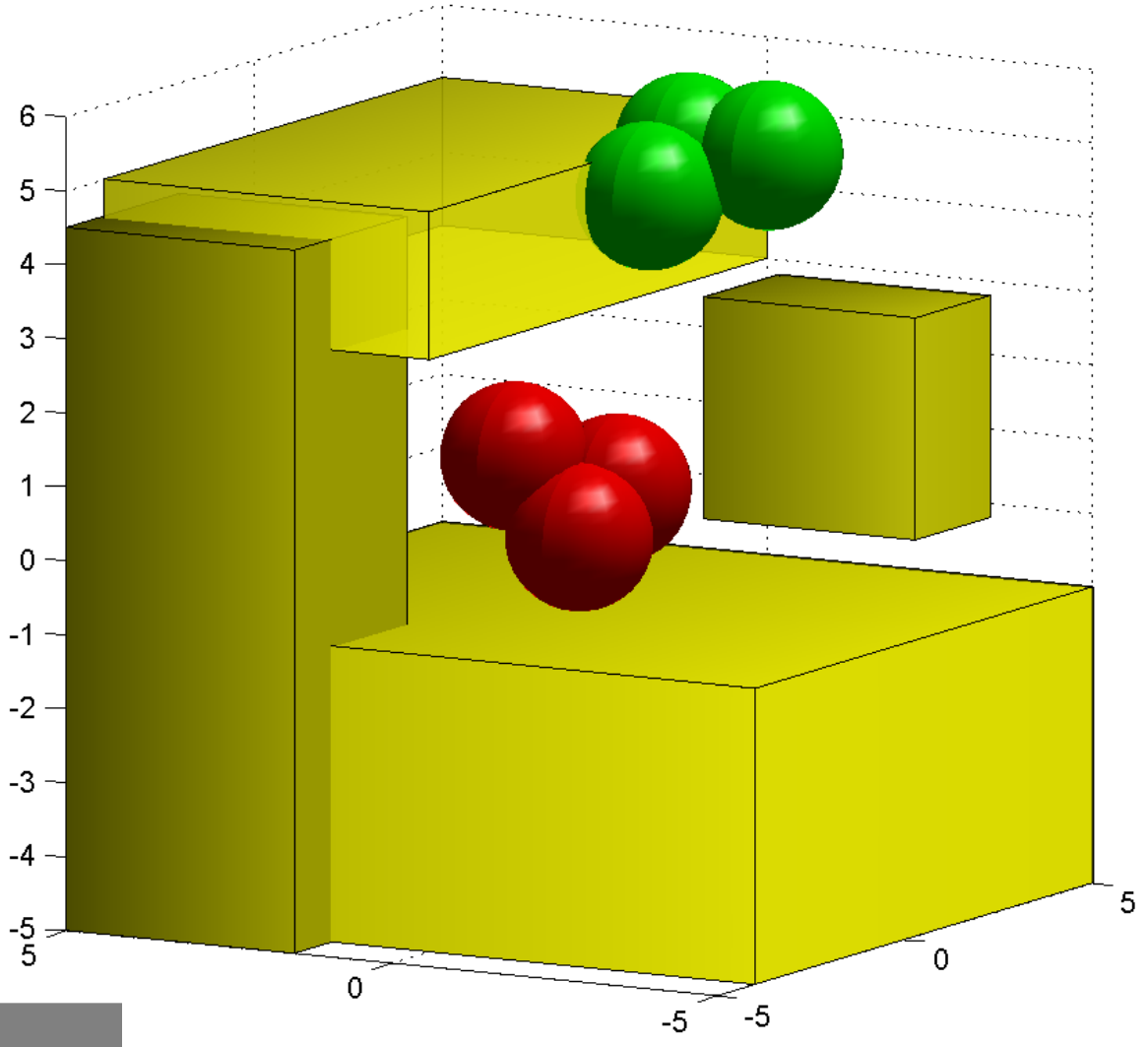
Self
Classifiers

Next Generation Result



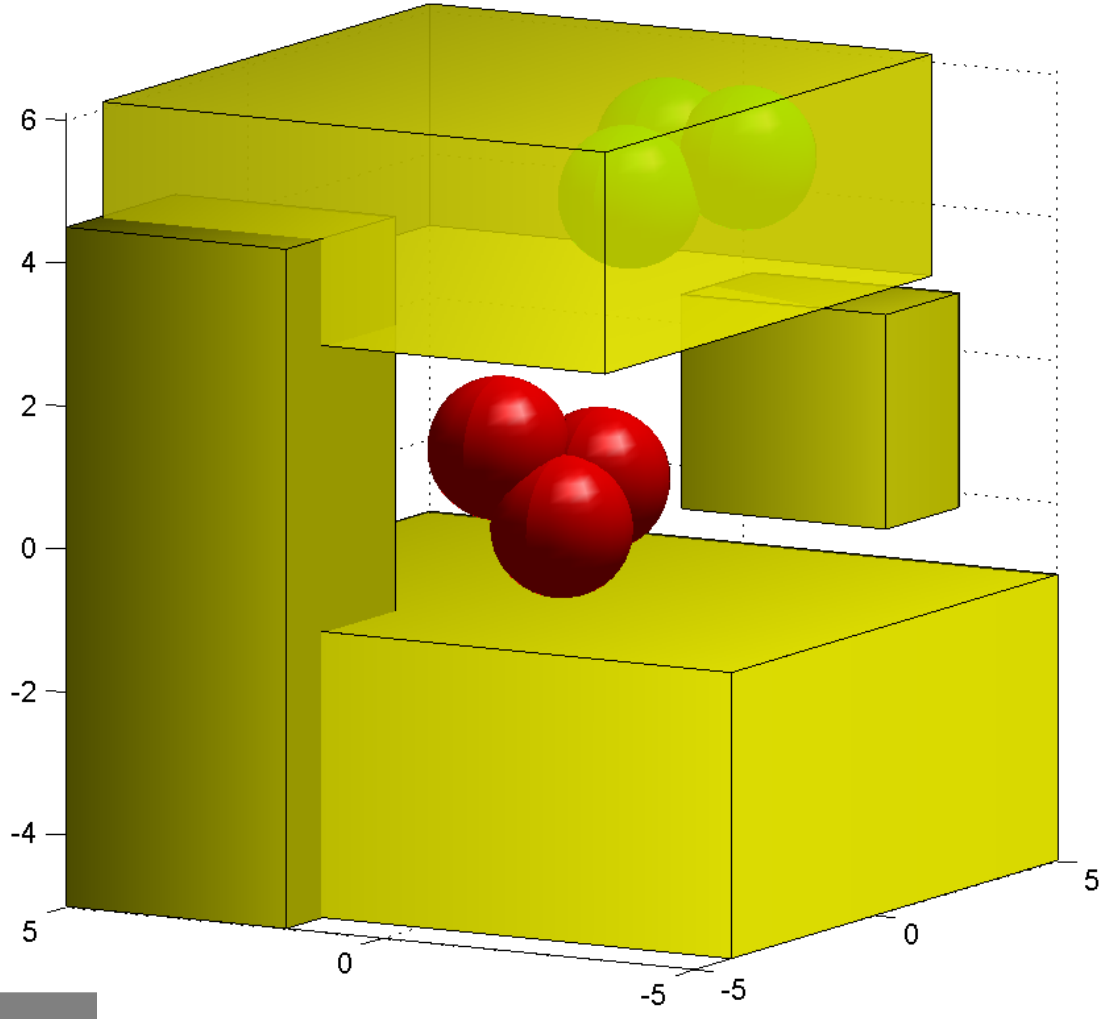
Self
Classifiers

Known Nonself



Self
Classifiers
Known nonself

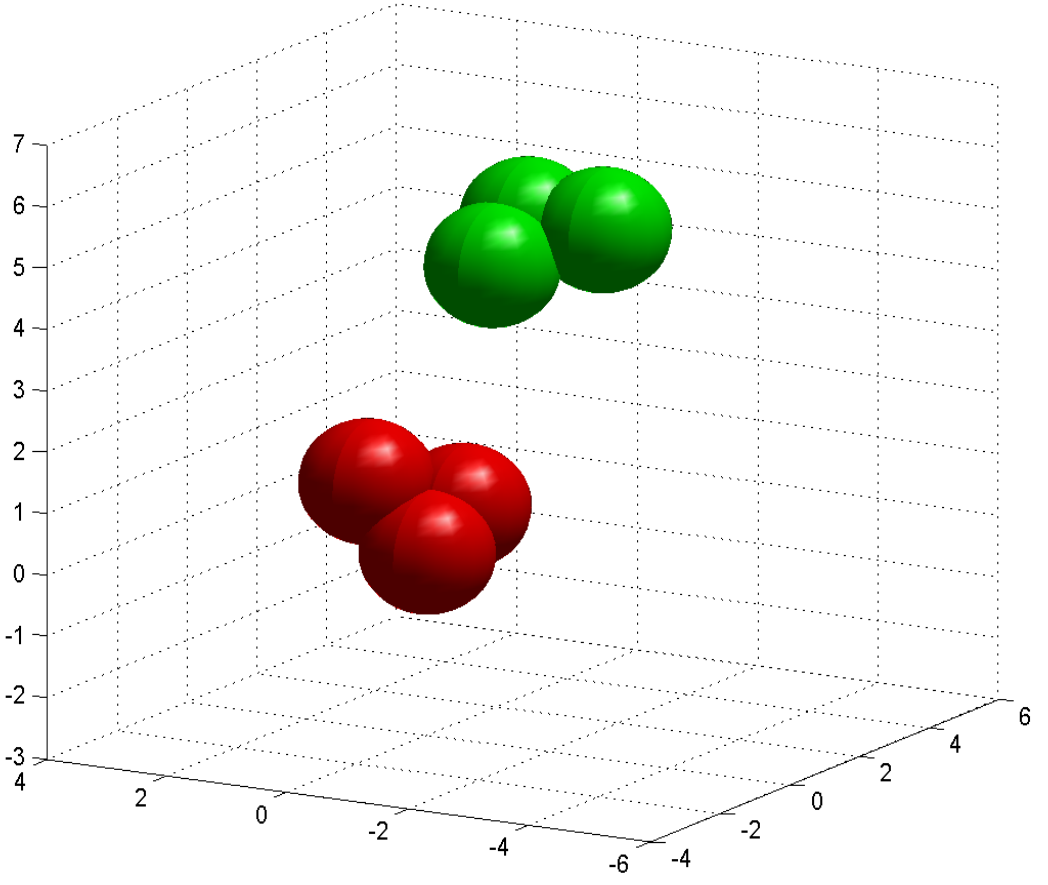
Finished?



Self
Classifiers
Known nonself

Research Concerns

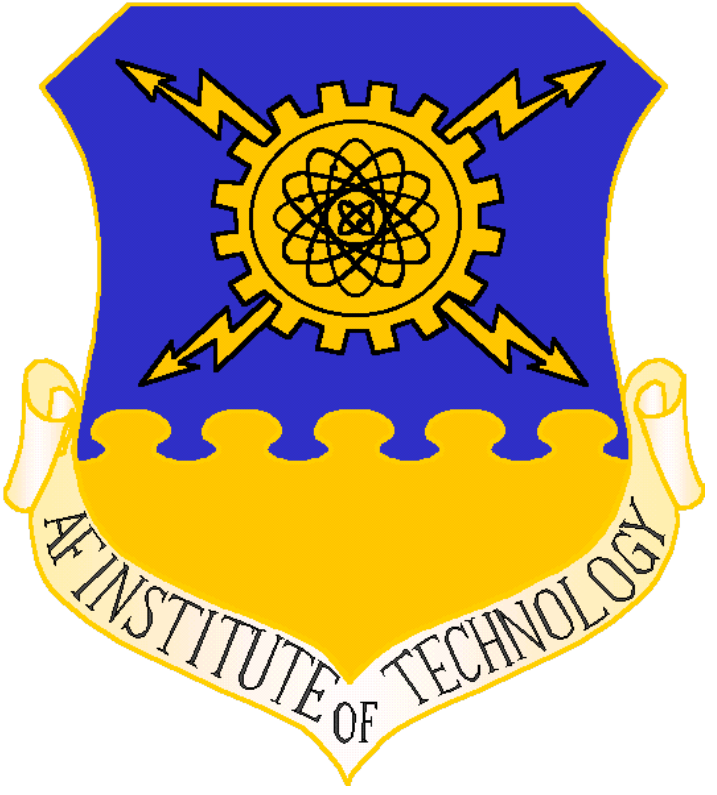
- Self and known nonself hypervolumes not disjoint
- Picking the best statistics and coefficient predictors
- Computation time associated with GAs



Overview

- Research goal
- Wavelet analysis background
- Computational Immune Systems (CIS) background and methodology
  - Genetic algorithms (GAs)
- Research concerns

Questions



AFIT

Backup Charts

References

[BeasA] Beasley, David and others. “An Overview of Genetic Algorithms: Part 1, Fundamentals,” *University Computing*, 15(2): 58-69 (1993).

[BeasB] Beasley, David and others. “An Overview of Genetic Algorithms: Part 2, Research Topics,” *University Computing*, 15(4): 170-181 (1993).

[Fari] Farid, Hany. *Detecting Steganographic Messages in Digital Images*. Technical Report TR2001-412, Hanover, NH: Dartmouth College, 2001.

[Frid] Fridrich, Jessica and Miroslav Goljan. “Practical Steganalysis of Digital Images – State of the Art,” *Proc. SPIE Photonics West 2002: Electronic Imaging, Security and Watermarking Contents IV,* 4675:1-13 (January 2002).

[Hubb] Hubbard, Barbara Burke. *The World According to Wavelets.* Wellesley, MA: A K Peters, 1996.

[John] Johnson, Neil F. and others. *Information Hiding: Steganography and Watermarking – Attacks and Countermeasures*. Boston: Kluwer Academic Publishers, 2001.

[Katz] Katzenbeisser, Stefan and Fabien A. P. Petitcolas, editors. *Information Hiding Techniques for Steganography and Digital Watermarking*. Boston: Artech House, 2000.

[Mend] Mendenhall, Capt. Michael J. *Wavelet-Based Audio Embedding and Audio/Video Compression.* MS thesis, AFIT/GE/ENG/01M-18, Graduate School of Engineering, Air Force Institute of Technology (AETC), Wright-Patterson AFB OH, March 2001.

[Riou] Rioul, Oliver and Martin Vetterli. “Wavelets and Signal Processing,” *IEEE SP Magazine,* 14-38 (October 1991).

[West] Westfield, Andreas and Andreas Pfitzmann. “Attacks on Steganographic Systems - Breaking the Steganographic Utilities EzStego, Jsteg, Steganos, and S-Tools - and Some Lessons Learned,” *Lecture Notes in Computer Science*, 1768: 61-75 (2000).

[Will] Williams, Paul D. and others. “CDIS: Towards a Computer Immune System for Detecting Network Intrusions,” *Lecture Notes in Computer Science,* 2212:117-133 (2001).

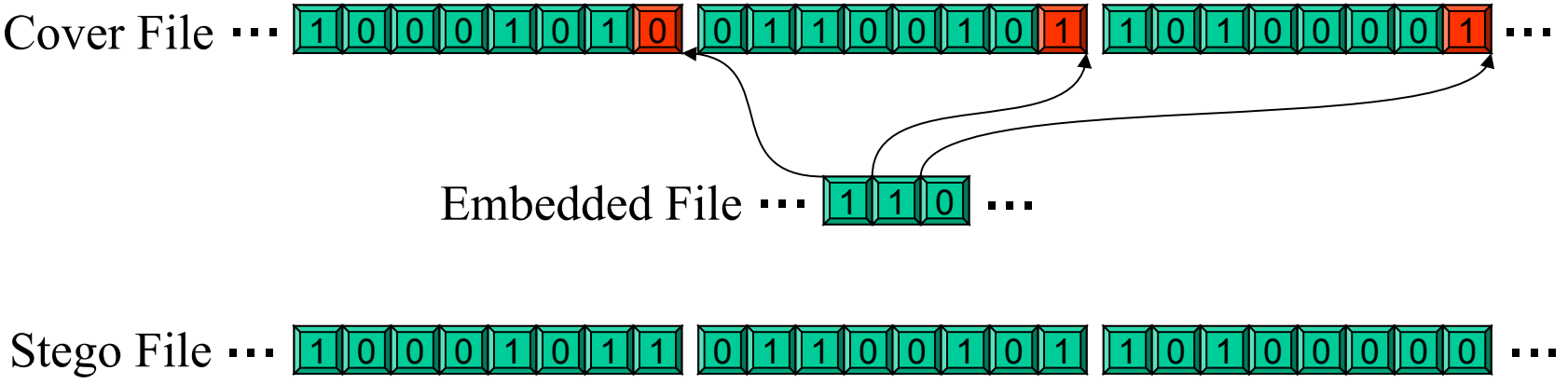
Steganography and Steganalysis

- Steganography
  - Goal – hide an embedded file within a cover file such that embedded file's existence is concealed
  - Result is called stego file
  - Substitution (least significant bit), transform, spread spectrum, cover generation, etc
- Steganalysis
  - Goals – detection, disabling, extraction, confusion of steganography
  - Visible detection, filtering, statistics, etc

Ref: Katz, West, John, Frid, Fari

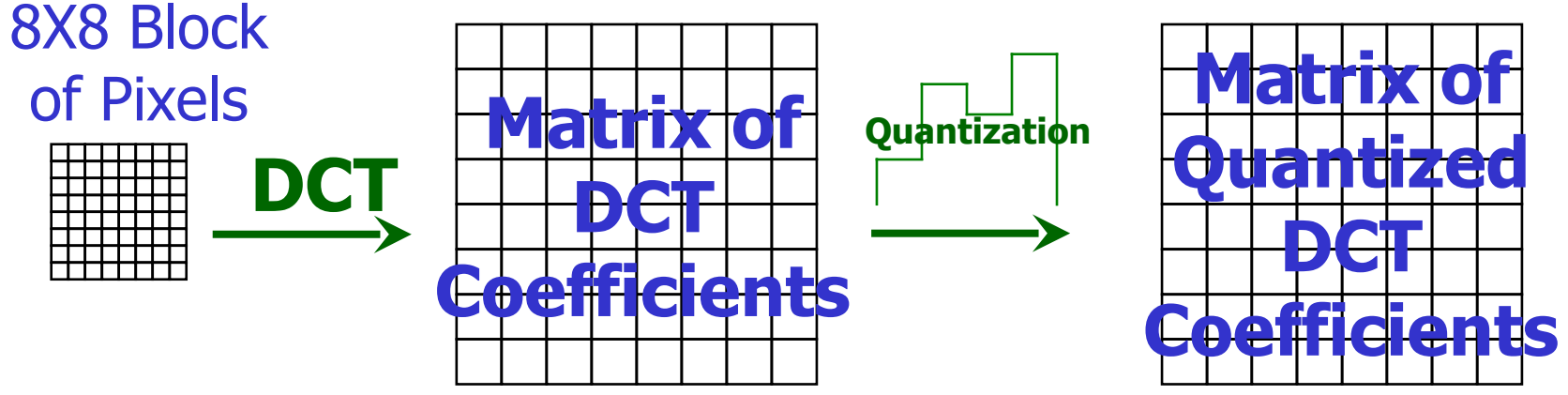
Steganography

- Least significant bit (LSB) substitution
  - Easy to understand and implement
  - Used in many available stego tools



Steganography

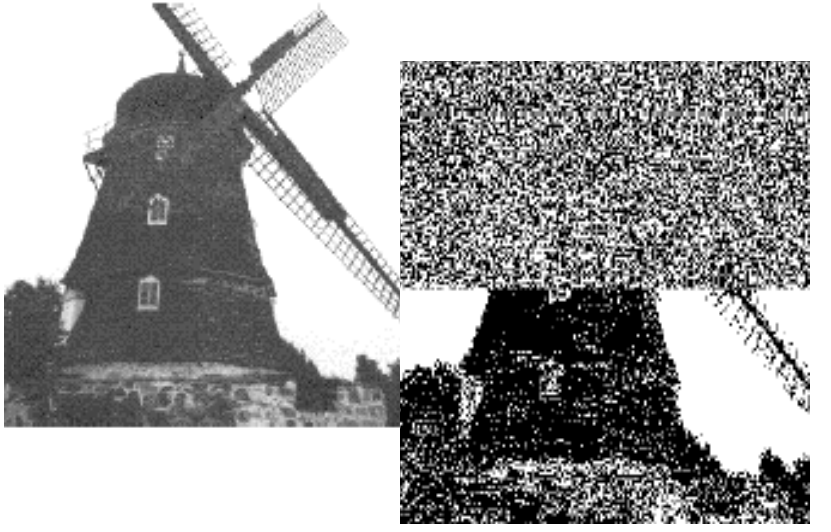
- Hiding in Discrete Cosine Transform (DCT)
  - Embed in difference between DCT coefficients
  - Embed in quantization rounding decision



Steganalysis

- Visible detection
  - Color shifts
  - Filtering – Westfield and Pfitzmann
- Simple statistics
  - Close color pairs
  - Raw quick pairs – Fridrich
  - OutGuess stego tool provides statistical correction
- Complex statistics
  - RS Steganalysis – Fridrich
  - Wavelet-based steganalysis – Farid

Stego



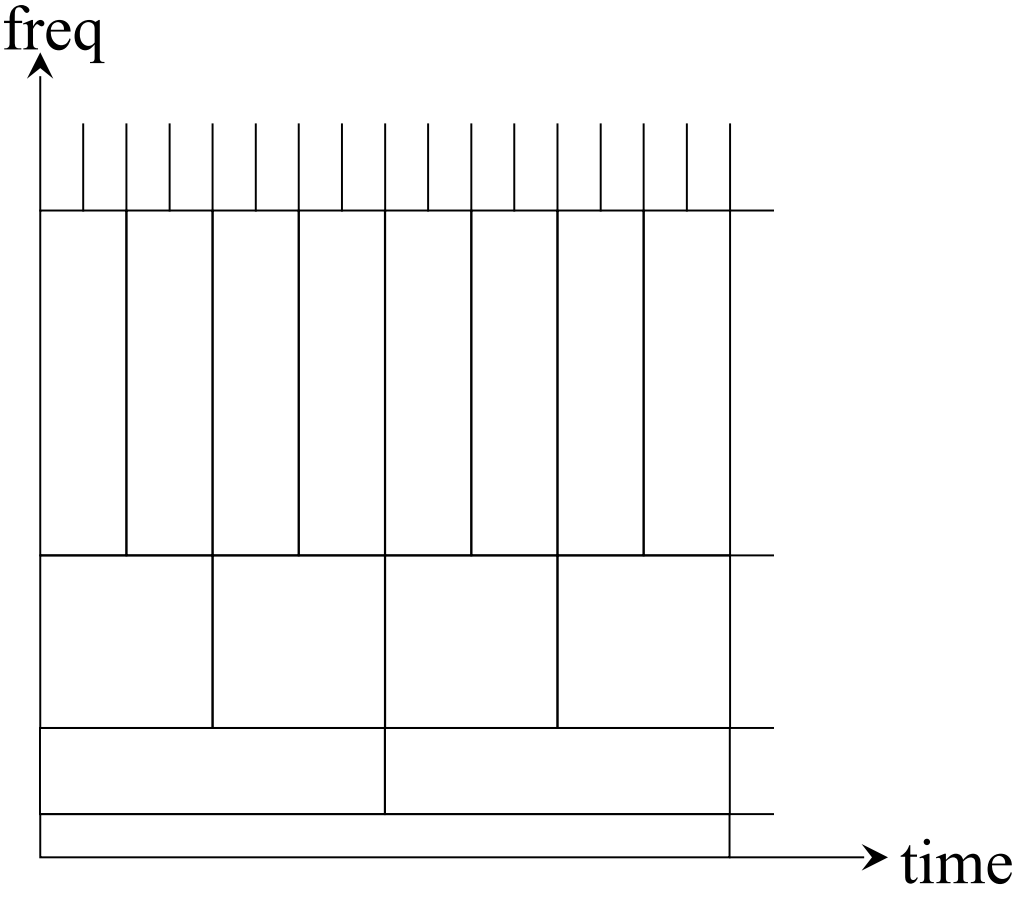
Filtered Stego

Image Formats

- 8-bit .bmp, .jpg, color .gif, and grayscale .gif
- Allow for testing of substitution and transform stego techniques
- Using EzStego, Jpeg-Jsteg, and OutGuess
  - User friendly tools
  - Good functionality
  - Range of detection ease
- Conversion to grayscale for wavelet analysis

Wavelet Analysis

- Fourier Transform
  - Good for stationary signals
  - Doesn't capture transient events very well
- Short-Time Fourier Transform offers good frequency or time resolution, but not both
- Wavelet analysis
  - Long time window for low frequencies
  - Short time window for high frequencies



Farid's Research

Image #	Jpeg-Jsteg		EzStego		OutGuess	
	Clean	Stego	Clean	Stego	Clean	Stego
1		X		X	O	
2	X		O			X
3	O		X			X
4		X	X		X	
5		O		O		X
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
499	X			O	X	
500		X	X			O

X = Training Set
O = Testing Set

Not Enough Statistics