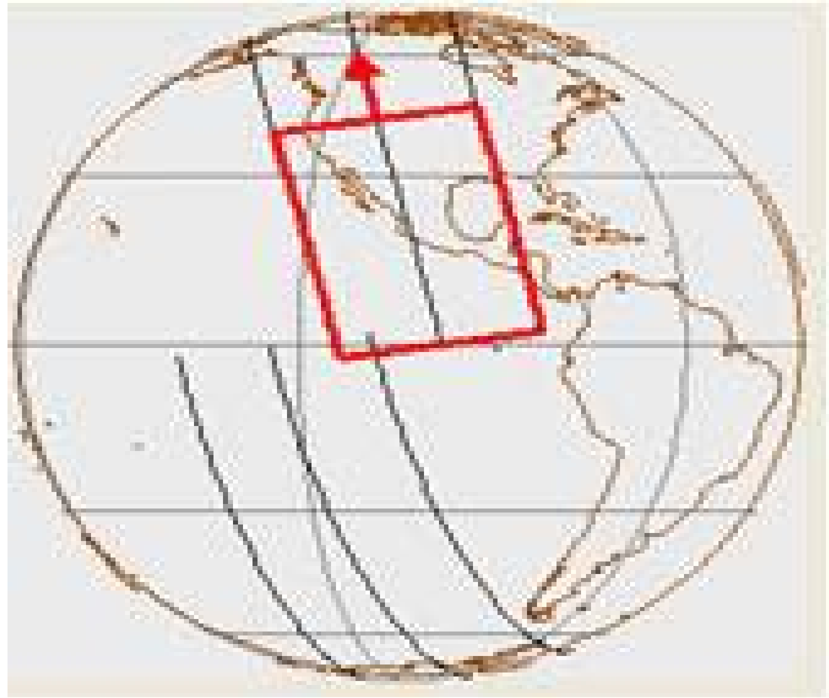
Similarity quantification of climatic images and tropical cyclone tracking and intensity estimation

Abdollah Homaifar (PhD)
Duke Energy Eminent Professor
Department of Electrical and Computer Engineering
North Carolina A&T State University
Contributors: Lacewell, Gebril, Buaba, Fetanat (PhD Students)
Knapp, Khin (NOAA),Obeidat (NCAT)

Fact: Data is growing rapidly!!!

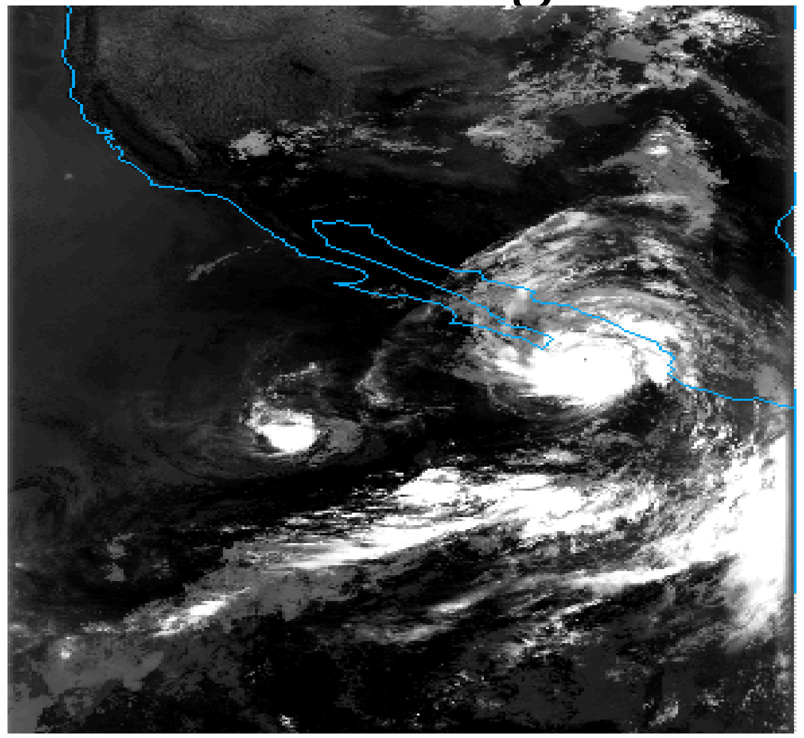
Example: DMSP Satellite images

Location



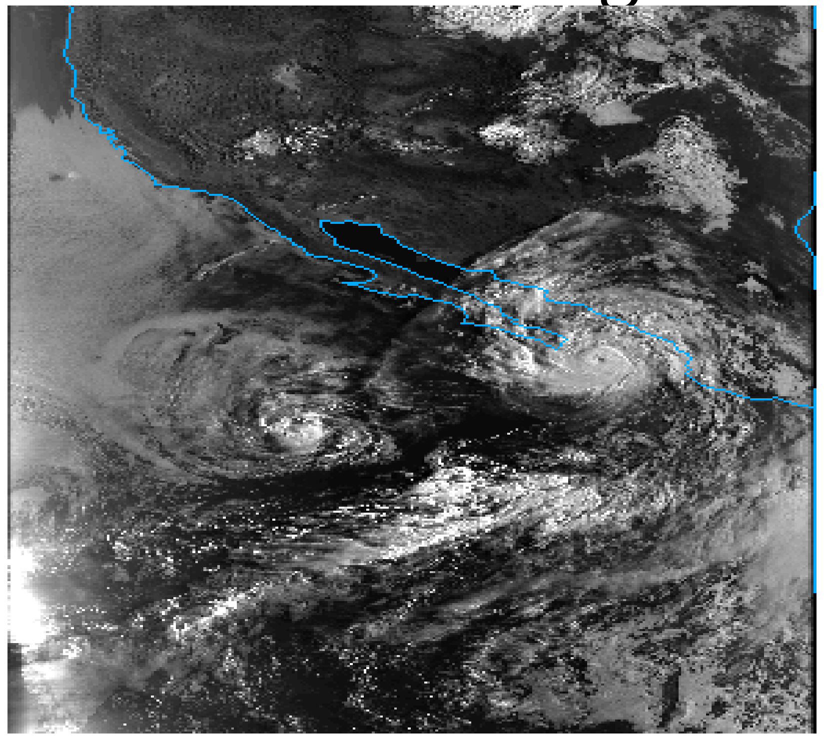
2.7 km ground sample coverage

Visual Image



363 x 293 pixels

Thermal Image



363 x 293 pixels

- “Curse of dimensionality”
- 17k/month, approx. 200k/yr , (24 GB/yr)
- About 3.8 million images available (about 0.5 TB)

Problem Definition

- Extract the dominant features that can uniquely represent satellite images
- Use the extracted features to:
  - track the origin of tropical storm (TS) Debby 2006 MCSs,
  - estimate the intensities of tropical cyclones (TCs), and
  - develop an approximate nearest neighbor (NN) algorithm to find similar images quickly.

Outline

- Feature Extraction
- Estimation of Tropical Cyclone Intensity
- Similarity Estimation of Climatic Images
- Conclusion
- Future work
- Published Papers
- References

Feature Extraction

- Texture is the spatial arrangement of the gray levels of the pixels
- Capturing visual content for indexing & retrieval
  - Basic image features: texture, shape, and color
- Most common techniques for describing texture:
  - Gray-Level Co-occurrence Matrix (GLCM)
  - Tamura
  - Gabor filtering
  - Pyramidal wavelet decomposition
- Texture features are suitable for:
  - Satellite images
  - Images of documents

Feature Extraction…

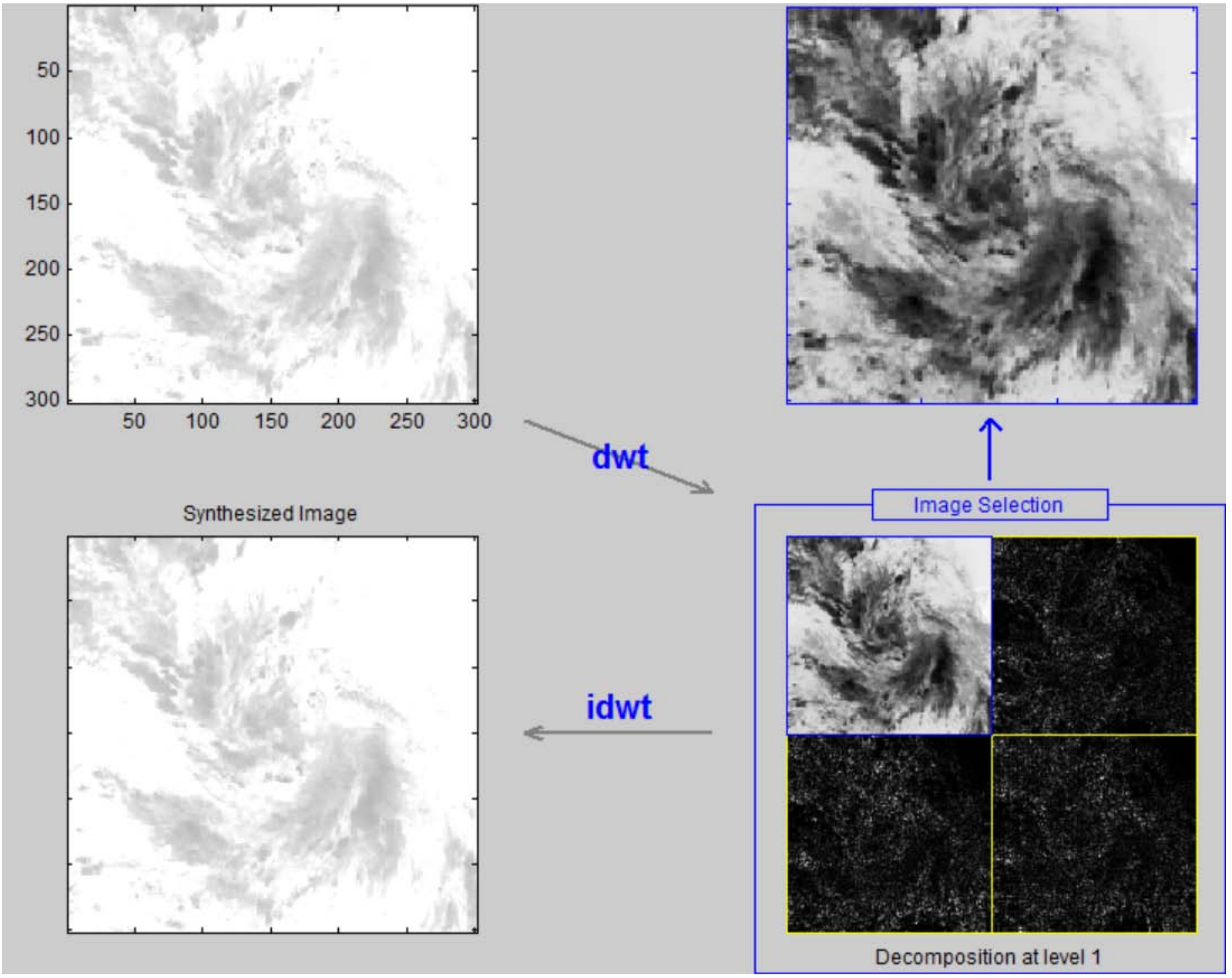
- GLCM :Statistical measurement of texture
  - Pros: Simple, creates a good outline of the object
  - Cones: Doesn’t extract the object, must have a known background
- Tamura texture feature
  - Method for approximating intuitive texture features:
    - ✓ Coarseness: coarse vs fine
    - ✓ Directionality: directional vs non-directional
    - ✓ Contrast: high contrast vs low contrast
    - ✓ Regularity: regular vs irregular (periodicity, randomness)
    - ✓ Roughness: rough vs smooth
  - Pros: Simple, simulates psychological measurement of directionality
  - Cones: Does not represent dependencies between pixels

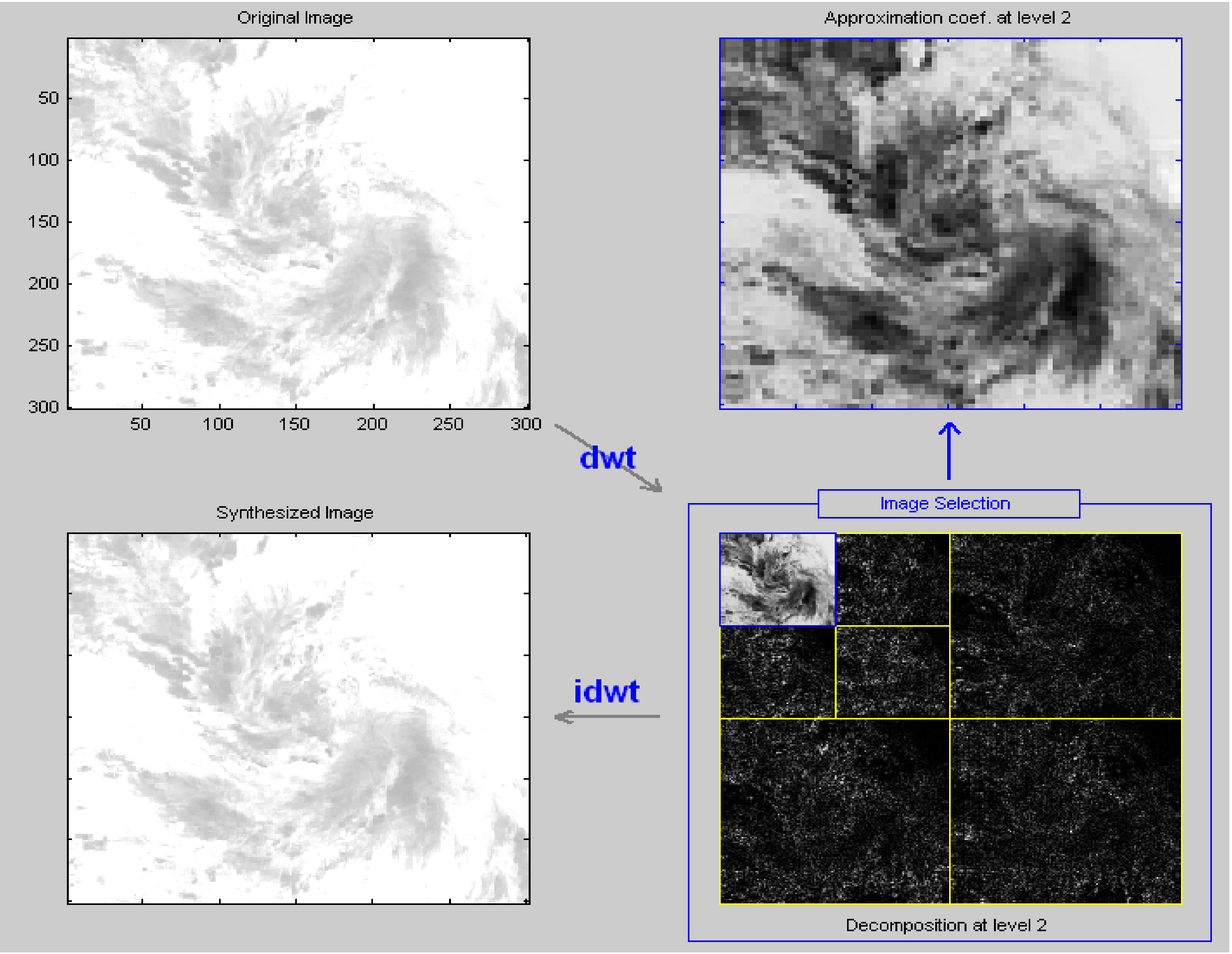
Feature Extraction…

- Wavelet Transforms
  - Mathematical transformations are applied to signals to obtain a further information from that signal that is not readily available in the raw signal.
  - Wavelet-based processing algorithms
    - ✓ Ability to discriminate different frequencies and to preserve signal details at different resolutions.
    - ✓ ‘zoom in’ and ‘zoom out’ capability of the wavelet filters which can translate themselves to a location of a signal that is of interest
  - Pros:
    - ✓ Multi-scale
    - ✓ Complete reconstruction
    - ✓ Edge types
    - ✓ Resist noise
  - Cons:
    - ✓ Time complexity

Feature Extraction…

Example: DWT

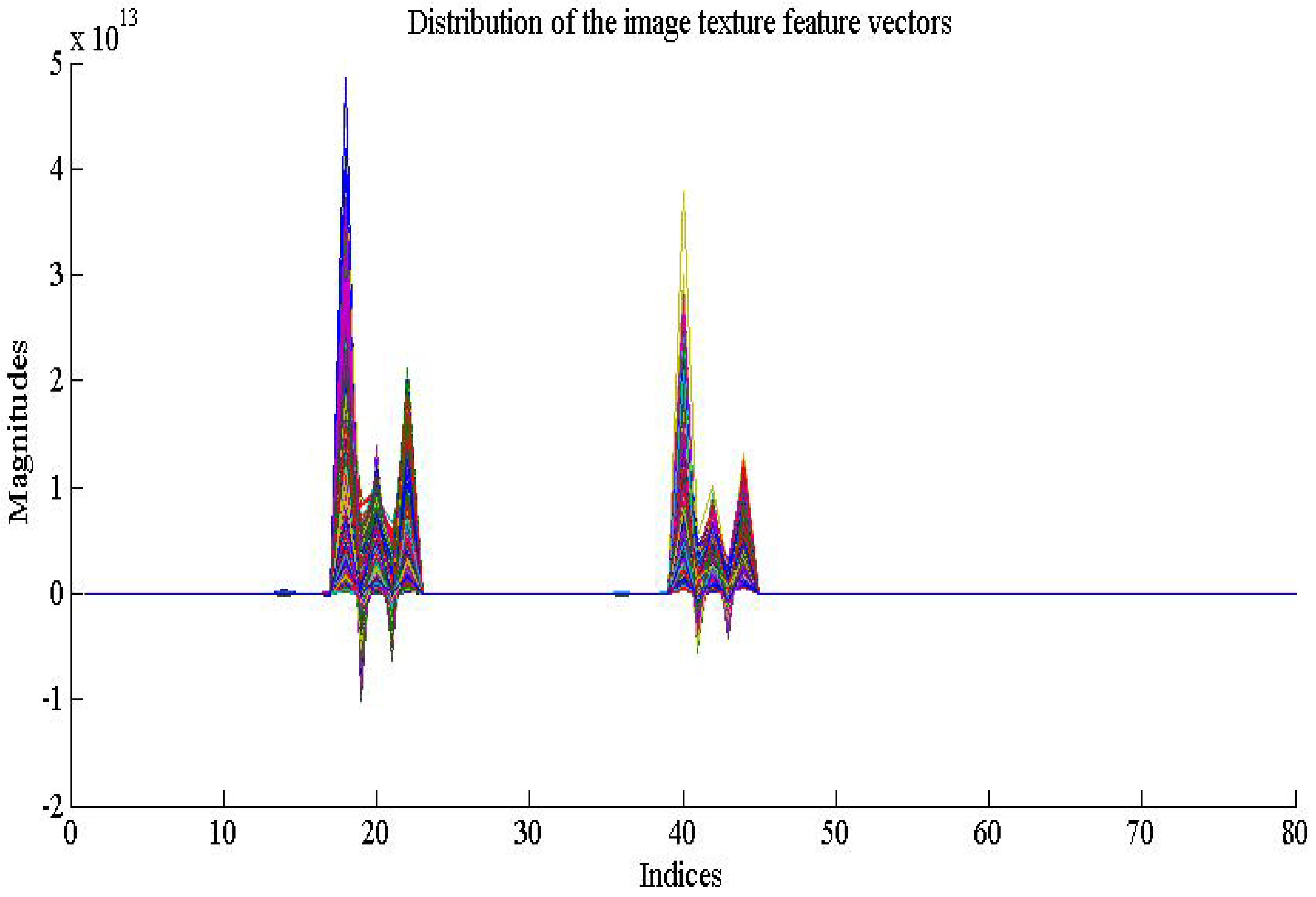




Sample Discrete Wavelet Transform Decomposition

Feature Extraction…

DWT applied to DMSP images

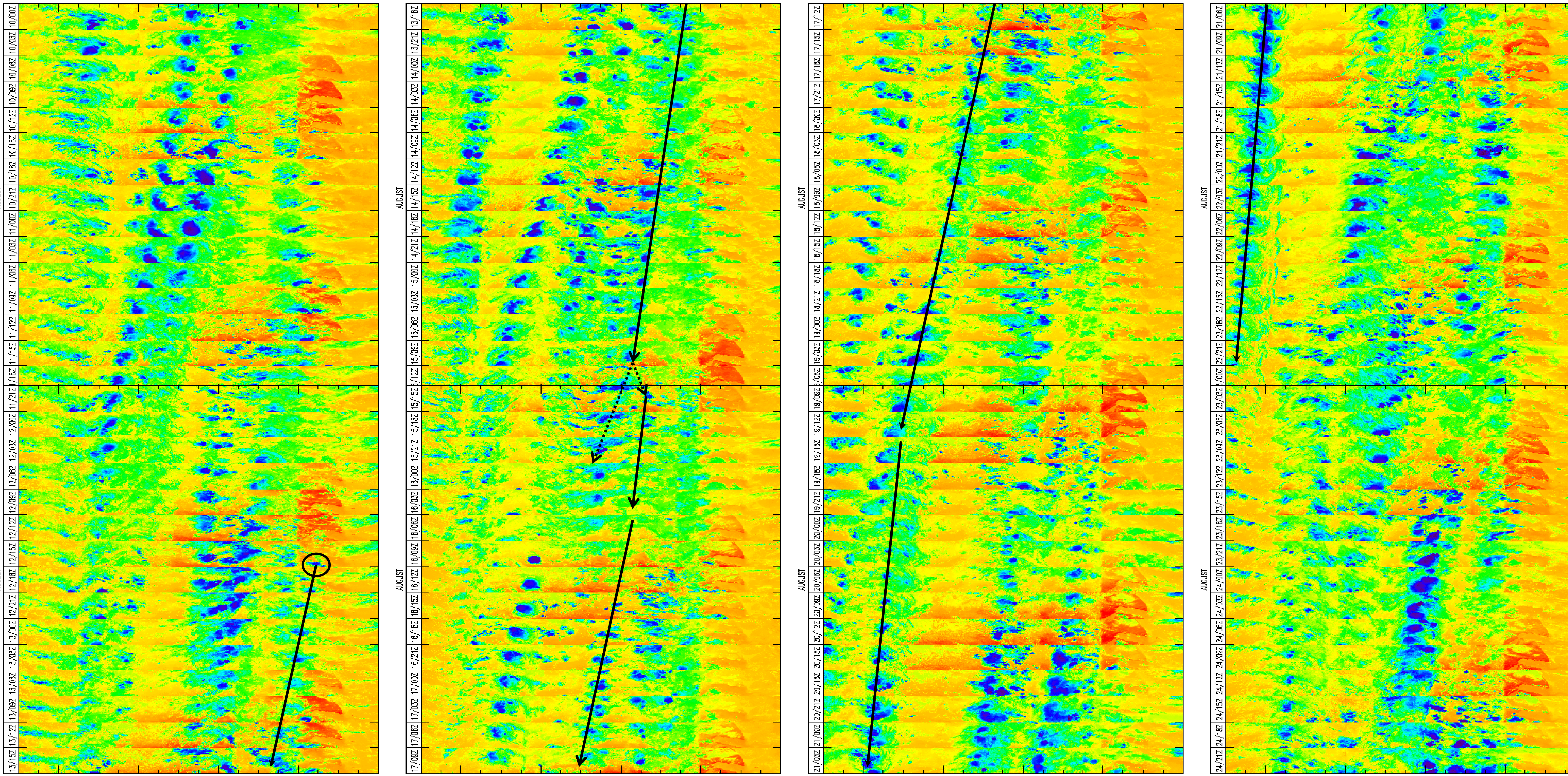


- Both visual and thermal images have been reduced from **363x293** to **1x40** pixels.
- Further reduced to 1x5 through feature selection
- Texture feature vector per an image set is 1x10
- Benefit
  - Reduce memory from 2GB/month to 1.5 MB/ month

Tracking of Debby (2006)

Manual:

Region of interest: 30^{o}W – 60^{o}E and 5^o – 15^{o}N

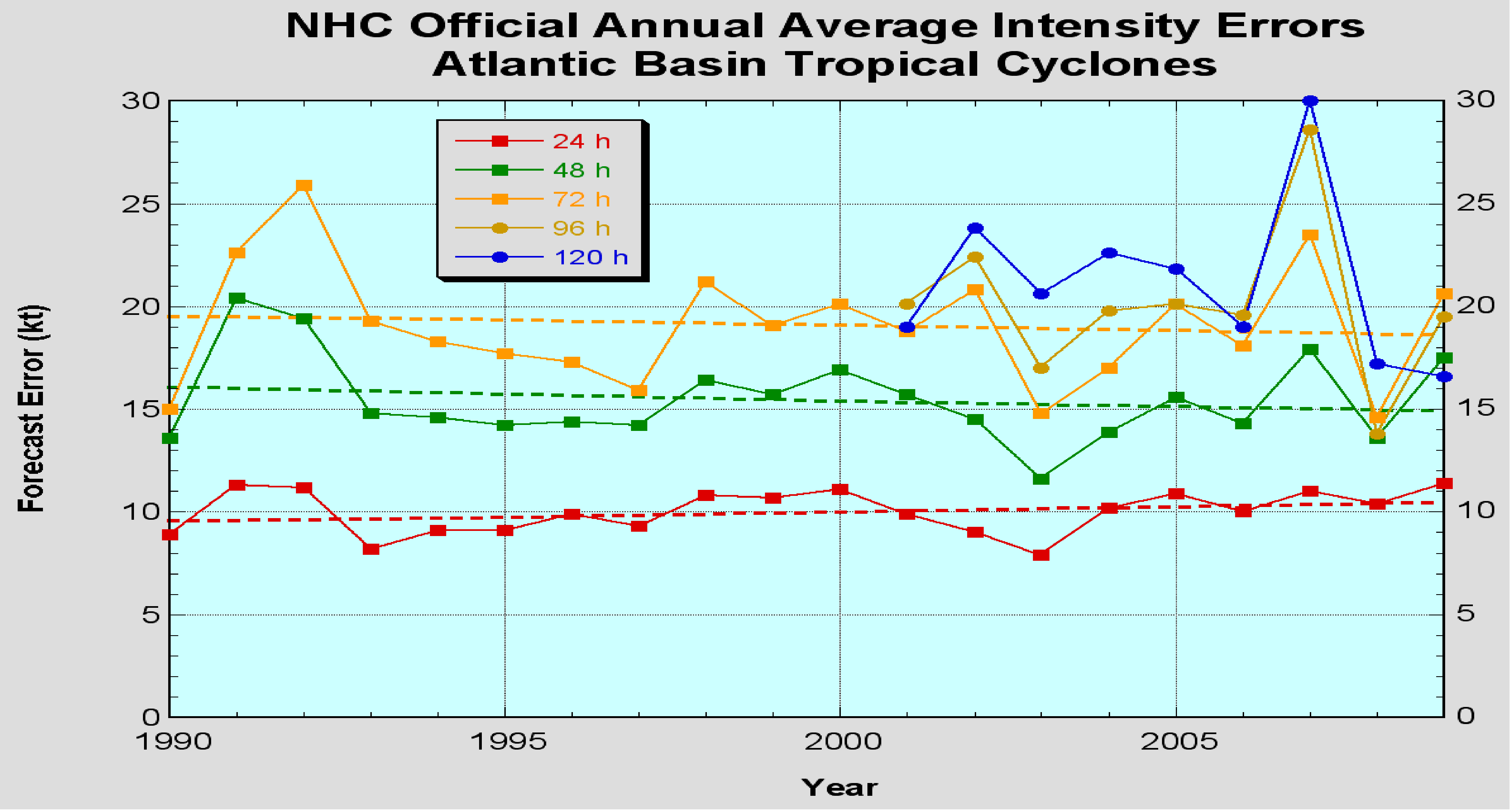


Tracking of Debby (2006)

Automatic:

- Data fusion
  - Image fusion technique (Lacewell et al. 2010)
    - ✓ Hybrid technique
      - ✓ Discrete wavelet transform (DWT)
      - ✓ Genetic Algorithm (GA)
- Pattern Recognition
  - Fuzzy C-Means Clustering (unsupervised technique)
  - Feature Extraction and Tracking
    - ✓ Technique used: Scale and Orientation Adaptive Mean Shift Tracking (SOAMST) proposed by Ning et al. (2010)
    - ✓ Solves problems of estimating the scale and orientation changes of an object
    - ✓ Based on the area of the object and the second order moments (variance)

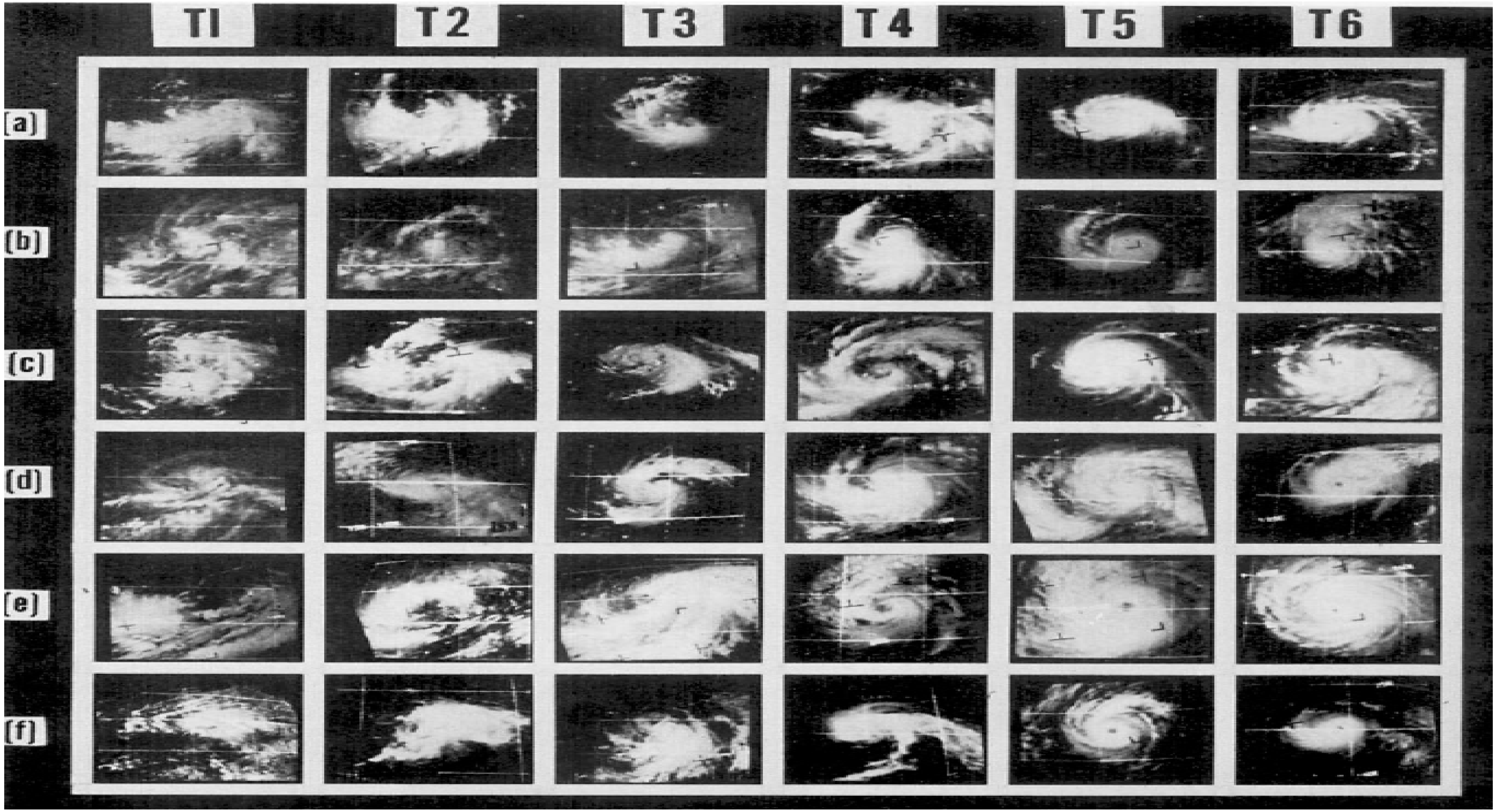
TC Intensity Estimation & Prediction



No progress in intensity prediction in the last 2 decades!

Dvorak Technique

Three decades using the Dvorak TC intensity estimation inferring manually from cloud patterns and features



Examples of tropical cyclone patterns and their associated T-numbers (adapted from Dvorak, 1975)

Data

Product	HURSAT-B1	HURSAT-Basin	HURSAT-GEO	HURSAT-AVHRR
Period of record	1983-2005 *1979-current*	2004-2005 *1979-current*	Part of 2006 *????-current*	None *1979-current*
Spatial span	10.5° from storm center for all global TCs	Basin wide ~100° long ~50° lat.	10.5° from storm center for GOES-observed TCs	10.5° from storm center for all global TCs
Temporal resolution	3 hourly	3 hourly	Varying (~30min)	Varying (~6 hourly)
Spatial resolution	8km	8km	4km	4km
Based on …	ISCCP B1	ISCCP B1	GOES full resolution data	AVHRR GAC
Channels available	IRWIN (11μm) IRWVP (6.7μm) VIS (0.65μm)	IRWIN (11μm) IRWVP (6.7μm) VIS (0.65μm)	All GOES channels	All AVHRR channels
Calibration[1]	Clim. - IRWIN. ISCCP – IRWVP, VIS *Clim. - IRWVP[2]*	Clim. - IRWIN. ISCCP – IRWVP, VIS *Clim. – IRWVP[2]*	Operational calibration.	Climate calibrated.
Yearly volume (GB) Uncompressed (Comp)	<6.5 (<2)	<20 (<4.6)[3]	Unknown[4]	Unknown[5]
Format	NetCDF	NetCDF	NetCDF	NetCDF
Current Version	1 *(version 2)*	Beta	Beta	None
Movies	Available	Planned	None	None

HURSAT Data

Northern Atlantic TAR (301.2MB)	Eastern Pacific TAR (553.4MB)	Western Pacific TAR (962.4MB)	Southern Hemisphere TAR (642.7MB)	Indian Ocean TAR (129.1MB)
ALBERTO (28.1MB) Jun-10 to 19	ALETTA (15.4MB) May-27 to 31	060 (29.2MB) Mar-01 to 07	CLARE (23.8MB) Jan-06 to 10	012 (35.2MB) Jan-12 to 19
198 (8.2MB) Jul-16 to 19	BUD (23.7MB) Jul-11 to 17	CHANCHU (62.1MB) May-07 to 19	TAM (11.8MB) Jan-11 to 14	MALA (28.3MB) Apr-24 to 29
BERYL (9.6MB) Jul-18 to 22	CARLOTTA (31.0MB) Jul-12 to 20	JELAWAT (37.3MB) Jun-22 to 29	URMIL (6.7MB) Jan-13 to 15	182 (14.6MB) Jun-30 to Jul-03
CHRIS (20.0MB) Aug-01 to 06	DANIEL (41.5MB) Jul-16 to 28	EWINIAR (56.5MB) Jun-29 to Jul-10	DARYL (28.9MB) Jan-18 to 23	MUKDA (23.9MB) Sep-19 to 26
DEBBY (28.7MB) Aug-21 to 28	EMILIA (39.8MB) Jul-21 to 31	BILIS (38.4MB) Jul-07 to 14	BOLOETSE (104.2MB) Jan-22 to Feb-06	271 (10.6MB) Sep-28 to 30
ERNESTO (32.5MB) Aug-24 to Sep-04	GILMA (16.8MB) Aug-01 to 05	KAEMI (46.6MB) Jul-17 to 26	JIM (29.4MB) Jan-27 to Feb-02	300 (16.3MB) Oct-27 to 30
FLORENCE (58.1MB) Sep-03 to 19	FABIO (17.0MB) Jul-31 to Aug-05	PRAPIROON (45.3MB) Jul-28 to Aug-05	VAIANU (24.4MB) Feb-10 to 17	
GORDON (50.5MB) Sep-10 to 23	IOKE (87.3MB) Aug-16 to Sep-06	SAOMAI (40.5MB) Aug-04 to 12	049 (37.7MB) Feb-18 to 23	
HELENE (57.0MB) Sep-12 to 27	HECTOR (30.5MB) Aug-15 to 24	MARIA (27.2MB) Aug-03 to 10	KATE (8.0MB) Feb-22 to 24	
ISAAC (8.0MB) Sep-27 to Oct-03	ILEANA (29.7MB) Aug-21 to 29	BOPHA (31.5MB) Aug-05 to 11	CARINA (39.9MB) Feb-22 to Mar-02	
	JOHN (30.4MB) Aug-28 to Sep-04	WUKONG (31.8MB) Aug-12 to 20	EMMA (10.0MB) Feb-26 to 28	
	KRISTY (39.1MB) Aug-30 to Sep-09	SONAMU (15.5MB) Aug-13 to 16	DIWA (53.0MB) Mar-01 to 09	
	LANE (15.6MB) Sep-13 to 17	234 (18.5MB) Aug-22 to 25	LARRY (23.1MB) Mar-16 to 21	

Sample 2006 TC classified by Location

TC Ground Truth Data

Saffir–Simpson Hurricane Scale

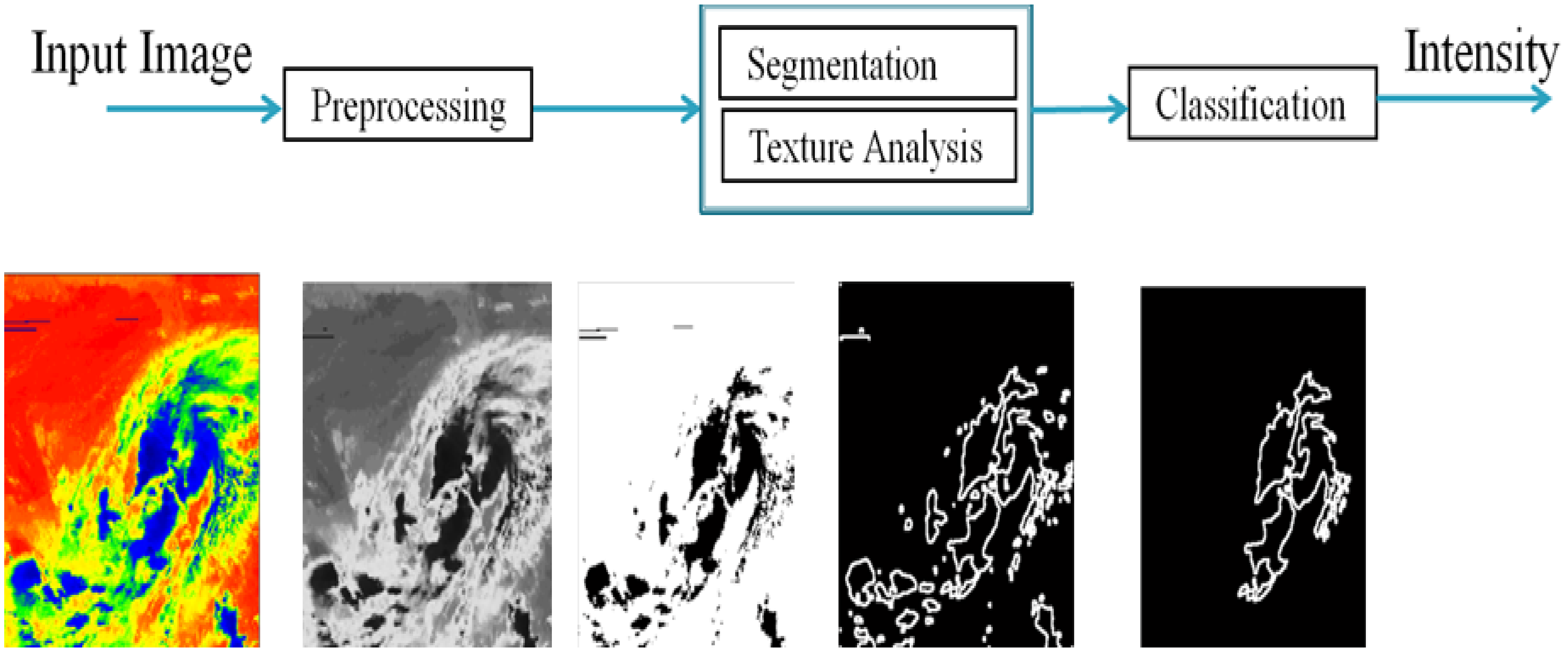
Category	Wind speed mph (km/h)	Storm surge ft (m)
Five	≥ 156 (≥ 250)	> 18 (> 5.5)
Four	131–155 (210–249)	13–18 (4.0–5.5)
Three	111–130 (178–209)	9–12 (2.7–3.7)
Two	96–110 (154–177)	6–8 (1.8–2.4)
One	74–95 (119–153)	4–5 (1.2–1.5)

Saffir–Simpson Hurricane Scale - Wikipedia, the free encycloped
http://en.wikipedia.org/wiki/Saffir%E2%80%93Simpson_Hurrica
Screen clipping taken: 7/8/2011 10:47 AM

Year Class	2006	2005	2004	2003
0	891	2993	1521	1413
1	223	500	204	139
2	55	109	162	89
3	32	103	155	119
4	0	115	150	85
5	0	48	51	50
Total	1201	3868	2243	1895

Sample Classification Results (North Atlantic)

TC Intensity Classification



Results using *K*-mean classification

Number of Training and testing images

Class	0	1	2	3	4	5
Training data	800	800	336	291	244	60
Testing data	200	200	84	73	61	15
Total	1000	1000	420	364	305	75

Northern Atlantic storms 1995 – 2005

Confusion Matrix

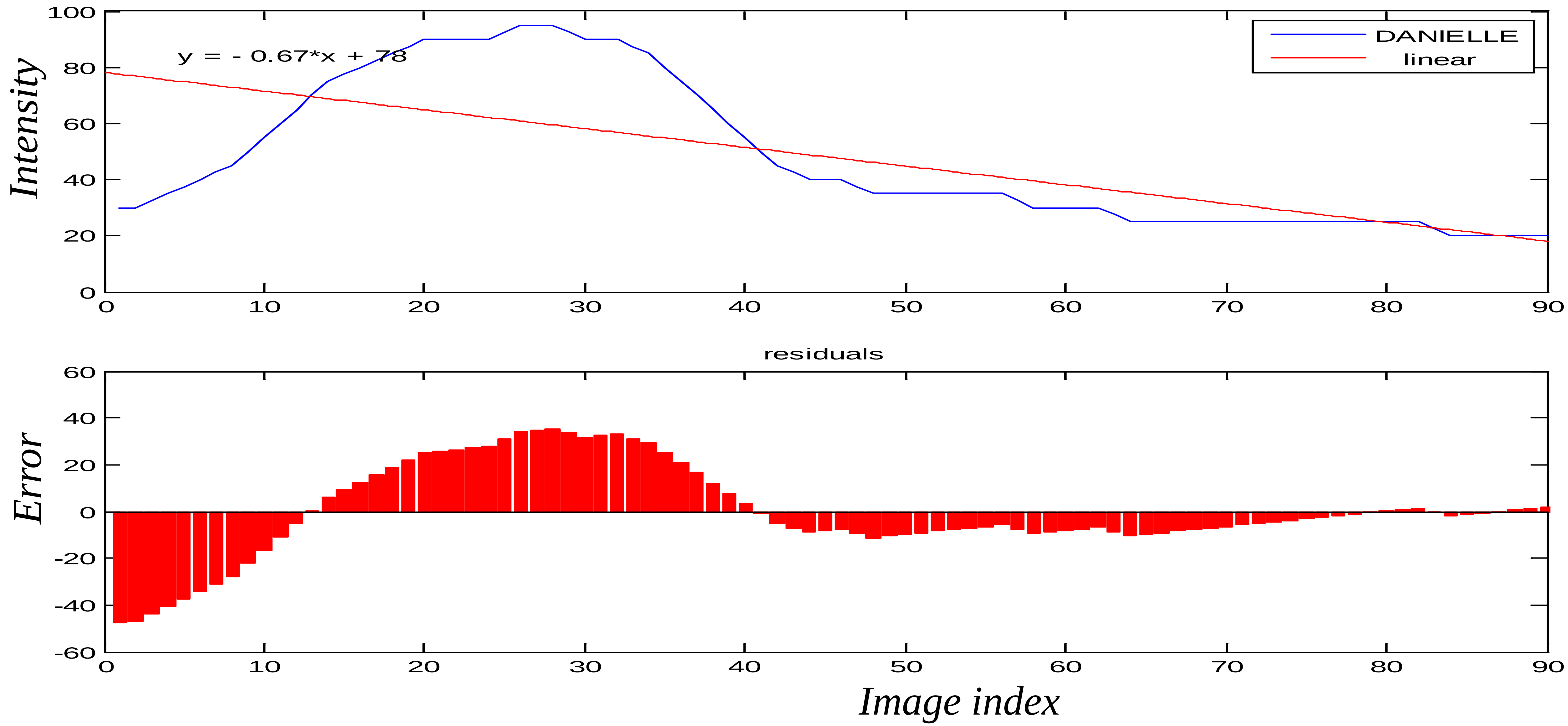
Class	0	1	2	3	4	5
0	**142**	15	2	0	0	0
1	35	**155**	17	5	2	0
2	5	17	**50**	10	0	0
3	5	3	10	**47**	6	0
4	11	10	5	11	**51**	5
5	2	0	0	0	2	**10**
Total	200	200	84	73	61	15

Average Classification Accuracy: 71%

Is the intensity of TC stationary?

Consider: "Danielle storm"

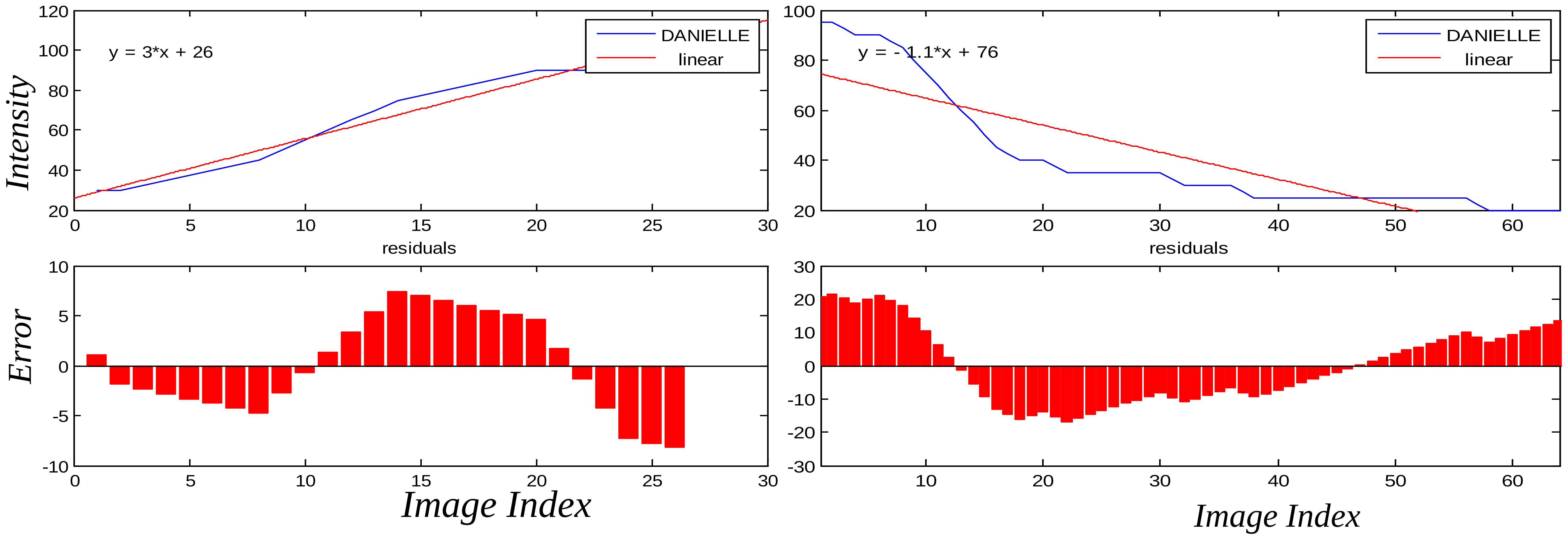
Linear regression modeling



- *Error:* $e[n] = x[n]-y[n]$
- *Average absolute error = 14.1472*

Danielle storm…

Piece-wise linear regression modeling



- *Average Absolute Error: 4.2485*
- *Average Absolute Error :10.1504*

Similarity Estimation

- **In 1D**
  - Build a data structure by sorting the points in ascending order -ideal
    - ✓ Complexity: *O(NlogN)*
    - ✓ Space: *O(N)*
  - Answer query using Binary search
    - ✓ Complexity: $O(log_2 N)$
- **In 2D**
  - Build a Veronoi diagram
    - ✓ Complexity: $O\left(N^{\lceil 0.5d \rceil}\right) \approx O(N)$
    - ✓ Space: $O\left(N^{\lceil 0.5d \rceil + \delta}\right)$
  - Answer query using any planar point location algorithm, e.g. slab decomposition [Dokin, Lipton, 1976] ,triangulation refinement , etc
    - ✓ Complexity: *O(logN)*
- **In *d*>2**
  - Complexity, Space and Query Time grow exponentially!

Bottleneck

- Building a Veronoi Diagram for data in *3D –space* containing *N* items:

  Complexity: $O(N^2)$
  Space : $O(N^2)$
  Search Complexity: $O(\log N)$

- In practice we deal with higher dimensions *(say d≥10) and N ≥ 1million:*

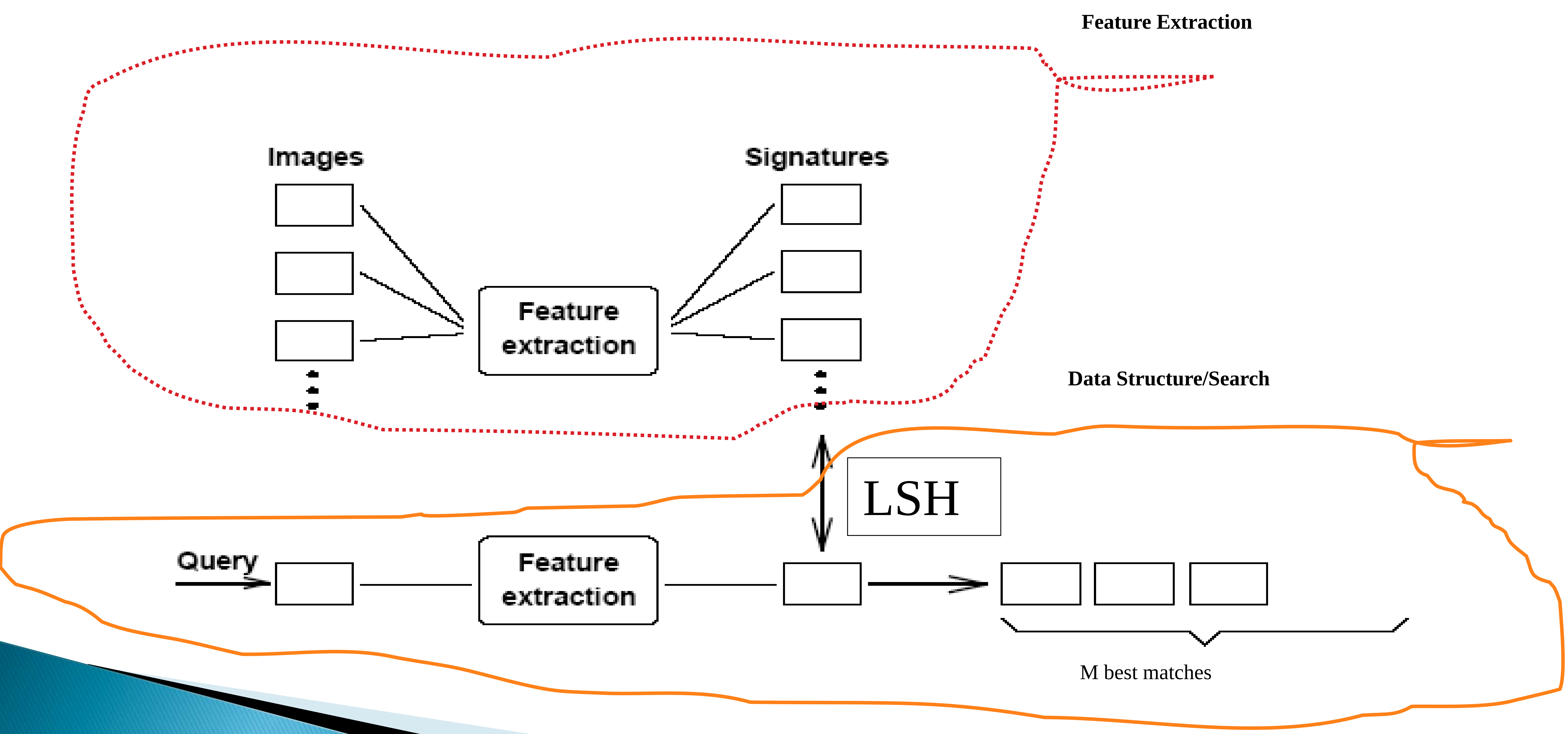
  Complexity: $\geq O(N^5)$ *-Ugh this program is taking too much time*
  Space : $\geq O(N^5)$ *- Out of memory.*

  Search Complexity: $\geq O(\log N)$

**Problem: Space is polynomial in *N*

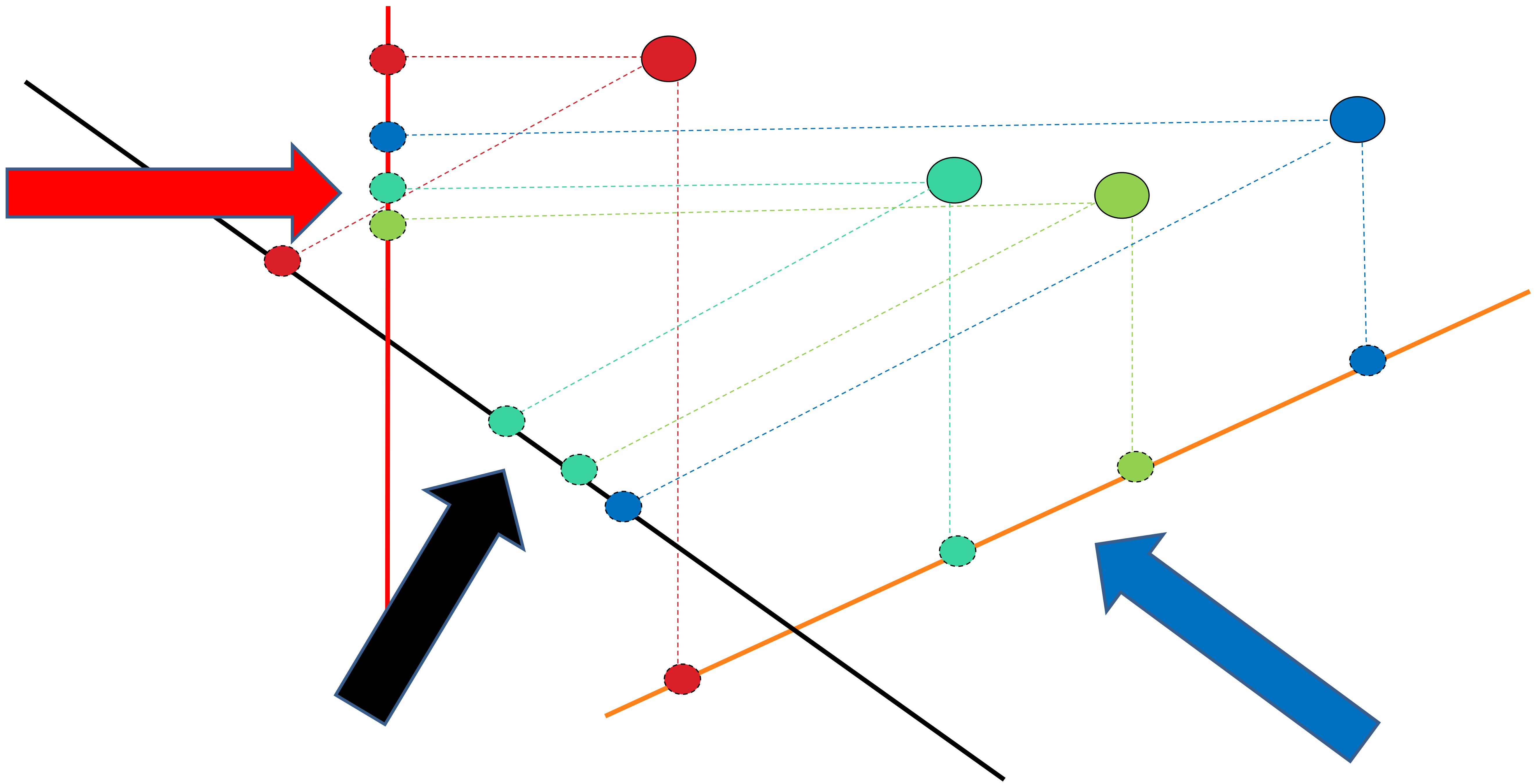
- **Is it possible to build a data structure to simultaneously achieve roughly linear space and logarithmic query time?**
  - kd-trees (Friedman, Bentley, and Finkel, 1977) :
    Space: $O(N)$
    Search Complexity: $O(\log N)$
    - ✓ In practice, kd- trees **break down** for higher dimensions
  - LSH (Motwani, 1998) :
    Space: $O(dN)^{O(1)}$
    Search Complexity: $O(d\log^{O(1)}(dN))$

General Methodology



Locality Sensitive Hashing (LSH)

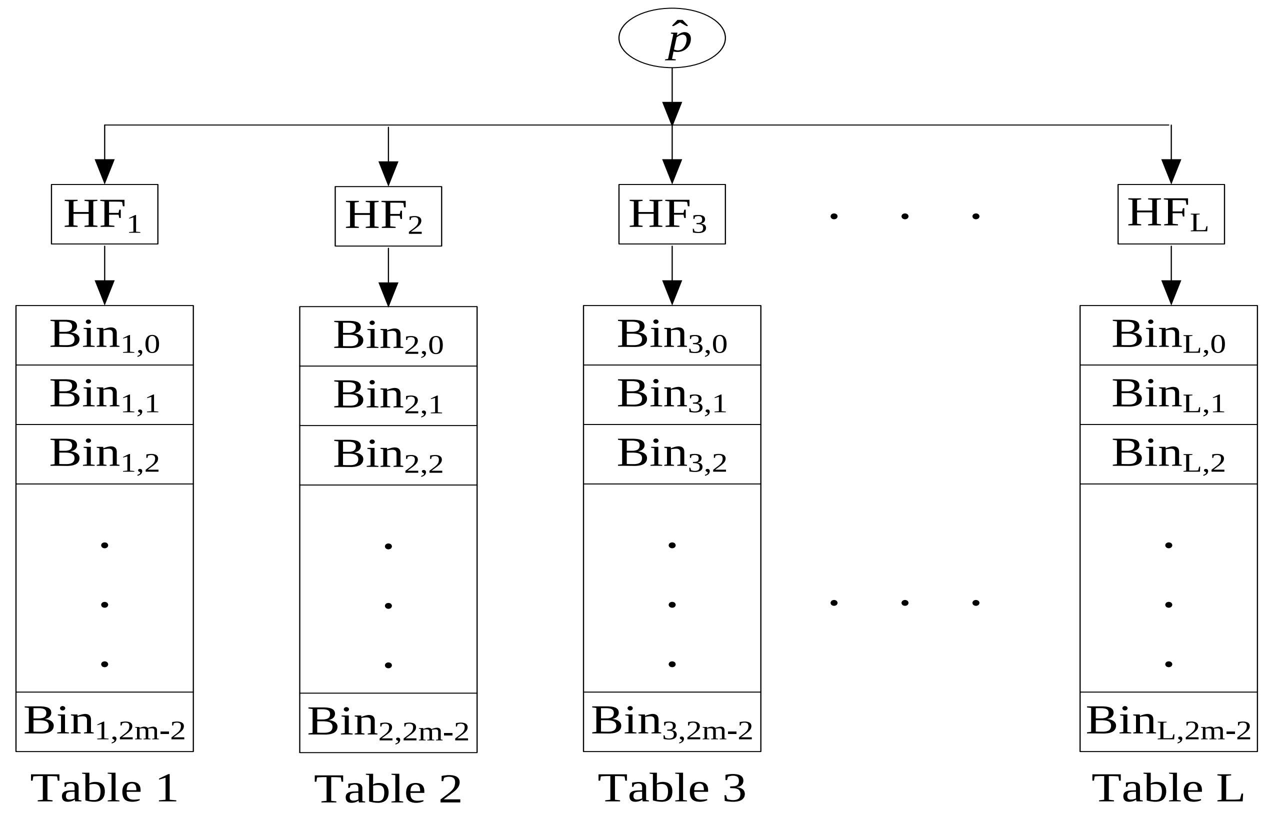
Basic Idea of LSH-> scalar projections



Hyper-planes => Gaussian normal distributions

modified Locality Sensitive Hashing (mLSH)

Creating Hash Tables



p: any *d*-dimensional texture feature vector

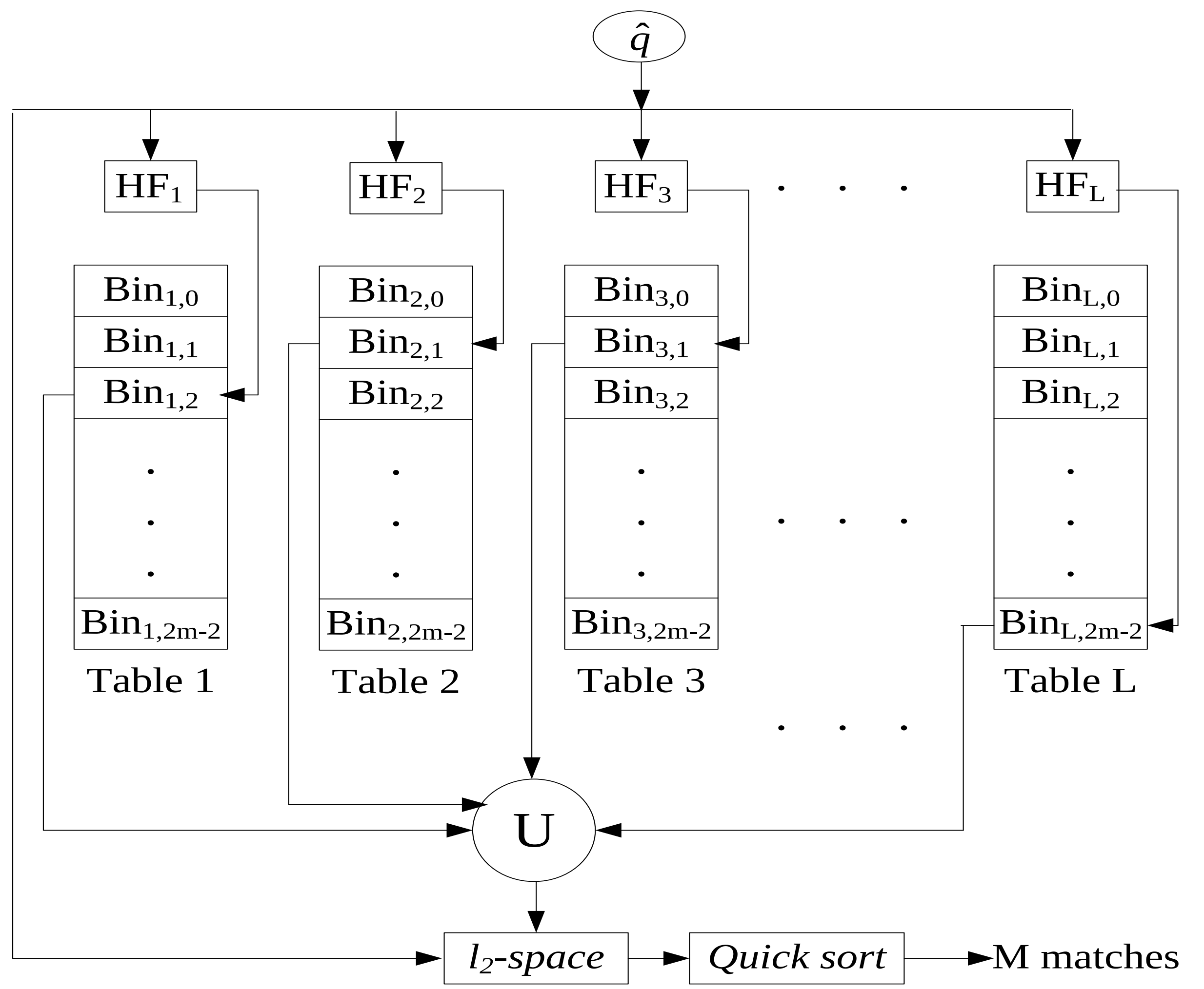
$$\underset{\forall\, j=\{1,2,\ldots,L\}}{HF_j(\hat{p})} = \left(\sum_{z=1}^{k} \left\lfloor \frac{h_{zj} \bullet \hat{p} + b^{(z)}}{\alpha} \right\rfloor H^{(z)} \right) \bmod m + m - 1$$

$h_{zj} \sim N_{gd}\left(0, d^2\right) \qquad \alpha = N/m \qquad b^{(z)} \sim [0,\ \alpha]$

$H^{(z)} \sim [0,\ m]$

N: number of texture feature vectors

Running A Query



q: query texture feature vector

M: number of nearest neighbors desired

Computational Complexity

mLSH

Creating The Hash Tables

Action	# Operations
Hashing p into a bin	2dk +k+2
Hashing p onto L tables	(2dk+k+2)L
Hashing all samples	(2dk+k+2)LN

Running A Query

Action	# Operations
Match set	(2dk +k+2)L
l_2-distance	(3d-1)αL
Quick-sort	O(αL log αL)

Total: (2dk+3dα+k+2-α)L + O(αL log αL)

LS

Running A Query

Action	# Operations
l_2-distance	(3d-1)N
Quick-sort	O(N log N)

Total: (3d-1)N + O(N log N)

Note: (i) αL << N
(ii) PDS = (αL/N)*100%

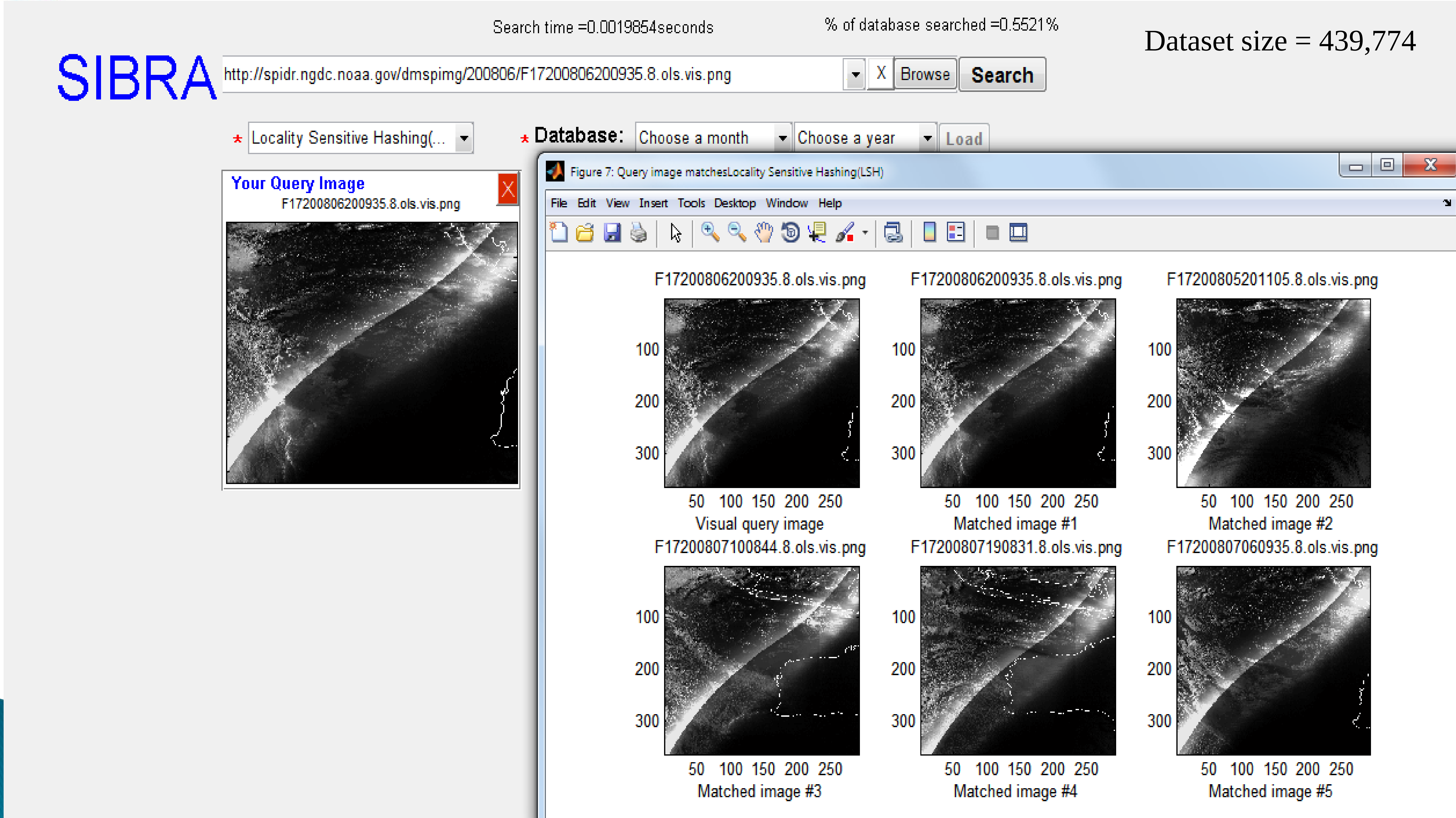
$$\text{Gain} = \frac{(3d-1)N + O(N \log N)}{(2dk+3d\alpha+k+2-\alpha)L + O(\alpha L \log \alpha L)}$$

Gain>>1

$$\underset{\forall\, j=\{1,2,\ldots,L\}}{HF_j(\hat{p})} = \left(\sum_{z=1}^{k} \left\lfloor \frac{h_{zj} \bullet \hat{p} + b^{(z)}}{\alpha} \right\rfloor H^{(z)} \right) \bmod m + m - 1$$

Implementation Platform

Dataset size = 439,774



Results

Note: M = 50, PDS: percentage of database searched (100% for LS)

Query Index	NNs out of M	Search Time(ms)		mLSH Gain Over LS	PDS (%)
	mLSH	LS	mLSH		mLSH
5555	50	154.5	17.6	8.79	4.64
64113	50	153.3	6.9	22.22	2.27
99045	45	163.2	1.6	102.95	0.27
101241	41	148.7	2.4	63.26	0.55
135426	50	148.4	9.3	15.89	3.26
165198	50	148.9	4.7	31.37	1.33
264676	50	149.5	3.4	43.39	0.93
276028	49	146.9	3.5	42.16	0.99
339715	50	147.0	11.6	12.69	4.00
429762	48	149.6	6.7	22.45	2.13
Average	**48**	**151.0**	**6.7**	**36.52**	**2.04**

Conclusions

- Pre-TS Debby (2006) originated over the peak of the Ethiopian Highlands
  - (10^{o}N, 40^{o}E) on 8/12/15Z
- The K-mean classification performs fairly well for the TC intensity classification
- TCs are non-stationary processes
- mLSH is highly scalable and about 36 times faster than the LS
- Results will be improved and generalized once sufficient data becomes available

Future Work

- Investigate the applicability of other feature extraction techniques (e.g. Shape)
- Implement a program to determine the probability of a cloud cluster developing into a tropical cyclone (TC)
- Develop a statistical technique using histogram and feature attributes to accurately detect origin of TCs
- Preprocess all tropical storms (TSs) to eliminate erroneous data
- Investigate temporal behavior of the extracted features as a step toward prediction of future dynamics of a TC
- Use mLSH to index the TCs for faster retrieval
- Validate the stationary or non-stationary nature of all the other TSs
- Improve upon the mLSH by using an optimal hyper-plane for projection

Published Papers

- G. Fetanat, R. Buaba, A. Homaifar,"A New ECG Arrhythmias classification technique based on Locality Sensitive Hashing" Submitted to Digital signal processing, Elsevier, Dec. 2010.
- R. Buaba, M. Gebril, A. Homaifar, E. Kihn, and M. Zhizhin,"Satellite image retrieval using low memory locality sensitive hashing in Euclidean space," Earth Science Informatics 2011, DOI 10.1007/s12145-010-0076-x, vol. 4, pp. 17–28.
- R. Buaba, A. Homaifar, M. Gebril, E. Kihn ,"Satellite Image Retrieval Application Using Locality Sensitive Hashing in L_2-Space," 32^{nd} IEEE Aerospace Conference, 2011.
- M. Gebril, R. Buaba, A. Homaifar, E. Kihn "Structural Indexing of Satellite Images Using Automatic Classification," 32^{nd} IEEE Aerospace Conference, 2011.
- R. Buaba, M. Gebril, A. Homaifar, E. Kihn, and M. Zhizhin, "Locality Sensitive Hashing for Satellite Images using Texture Feature Vectors," 31^{st} IEEE Aerospace Conference, 2010.
- M. Gebril, R. Buaba, A. Homaifar, E. Kihn, and M. Zhizhin, "Structural Indexing of Satellite Images Using Texture Feature Extraction Retrieval,"31^{st} IEEE Aerospace Conference, 2010.
- C. W. Lacewell, M. Gebril, R. Buaba, and A. Homaifar, "Optimization of Image Fusion Using Genetic Algorithms and Discrete Wavelet Transform," NAECON, 2010.