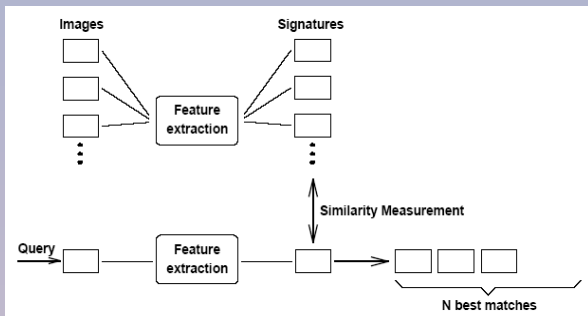


Objectives & NOAA Relevance

- Extract feature vectors of DMSP Satellite Imagery
- Develop an accurate and fast image retrieval based on feature vectors
- Relevance:
 - search for similar weather patterns in imagery (hurricane, aurora, glare, etc.)
 - Detect and predict weather related phenomena
 - Explore the extended urban regions by city lights

Overall Query Scheme

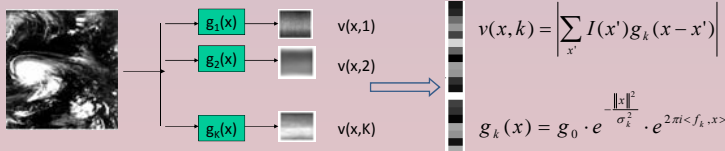


Challenges

- The curse of dimensionality of the images
 - ✓ Image size : 363 x 293 pixels
 - ✓ Image fusion: different image types
- Volume of the Database
 - ✓ 17K/month, approx. 200K/yr
 - ✓ Total, about 3.8M images
- Computer Memory
 - 2GB/month, approx. 24 GB/ yr
 - Total, about 480 GB
- Time complexity of the Similarity Search Engine
 - How fast? Meet real-time requirements?

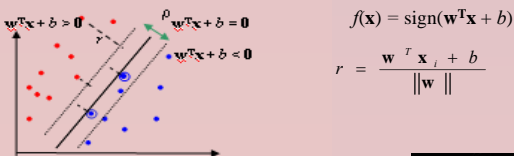
Feature Extraction

- Shape extraction for feature vector
- Global Image Spatial Transform description: the images are segmented by a 4 by 4 grid for which orientation histograms are extracted.
- Feature vector dimension is 320



Classification

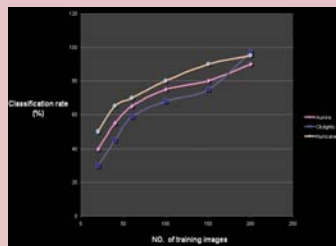
Examples closest to the hyperplane are **support vectors**.
Margin of the separator is the distance between support vectors.



Results

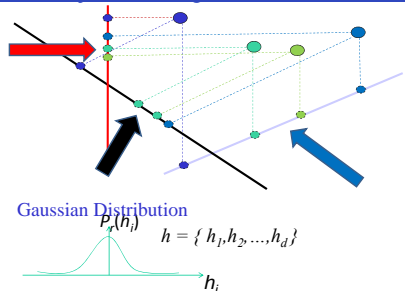
	Aurora	City lights	Hurricane
Aurora	475	34	28
City lights	42	542	35
Hurricane	22	32	580

Overall correct classification rate = 92%



Locality Sensitive Hashing in L₂-Space For Finding Nearest Neighbors

Scalar Projections using Gaussian Distribution



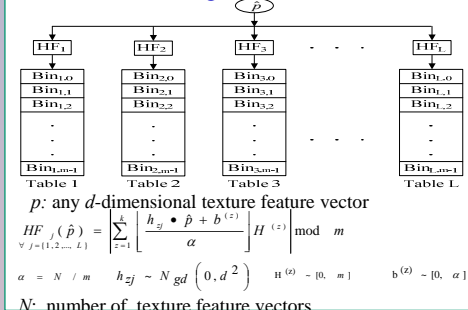
2-stable Distribution Theorem

$$\sum_{i=1}^d h_i p_i \Rightarrow \left(\sum_{i=1}^d |p_i|^2 \right)^{\frac{1}{2}} h$$

Note : $\sum_{i=1}^d h_i p_i = a \cdot p, \quad h, p \in \mathbb{R}^d$

$$h \cdot p - h \cdot q \Rightarrow \left(\sum_{i=1}^d |p_i - q_i|^2 \right)^{\frac{1}{2}} h$$

Creating Hash Tables



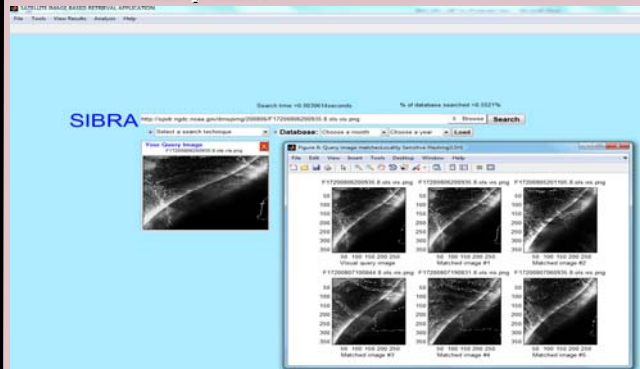
Results

Query Index	NNCR	Search Time(sec)		LSH Search Gain	PDBS (%)
		LS	LSH-I ₂		
5555	10	0.0949	0.0190	4.99	4.64
64113	10	0.0912	0.0048	18.93	2.27
99045	10	0.0926	0.0016	57.86	0.27
101241	10	0.0918	0.0019	48.88	0.55
135426	10	0.0921	0.0031	29.58	1.36
165198	10	0.0933	0.0029	31.92	1.33
264676	10	0.0919	0.0025	37.01	0.93
276028	10	0.0911	0.0025	35.88	0.99
339715	10	0.0934	0.0019	49.06	0.57
429762	10	0.0929	0.0046	20.32	2.13
Average	10	0.0925	0.0045	33.44	1.51

Conclusion

- Gist feature well represent shape features
- SVM classification helps Label classes to use for search by keyword than using LSH only
- pre-processing classification speeds up retrieval process

GUI - Snap Shot



Future work

- Annotate Images
- Search by sketch
- Combine different satellite Imagery sets

