3D Content-based Search

Petros Daras

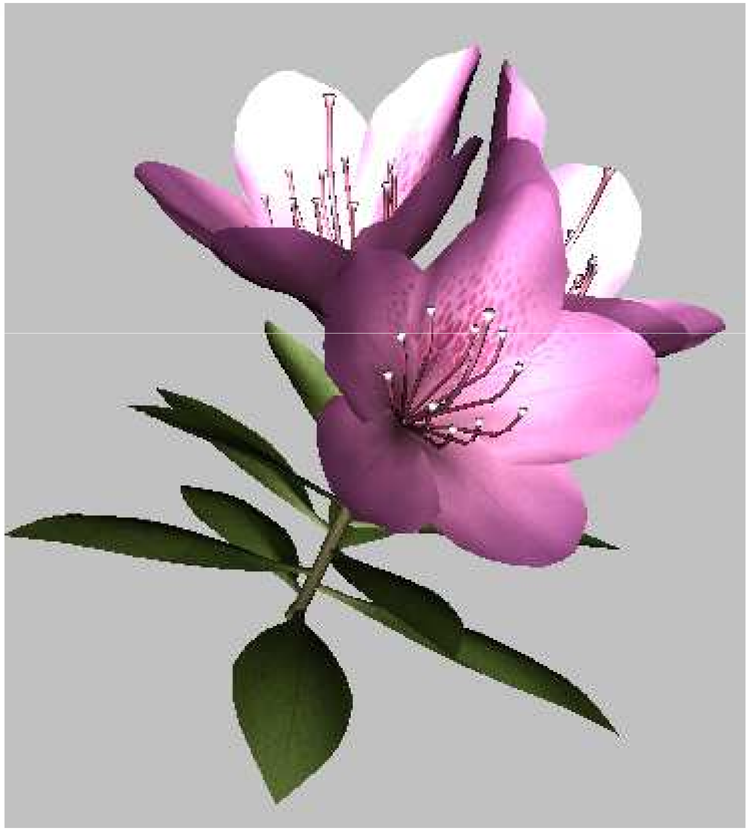
Centre for Research and Technology Hellas

Informatics & Telematics Institute

www.iti.gr

3D content: the king…

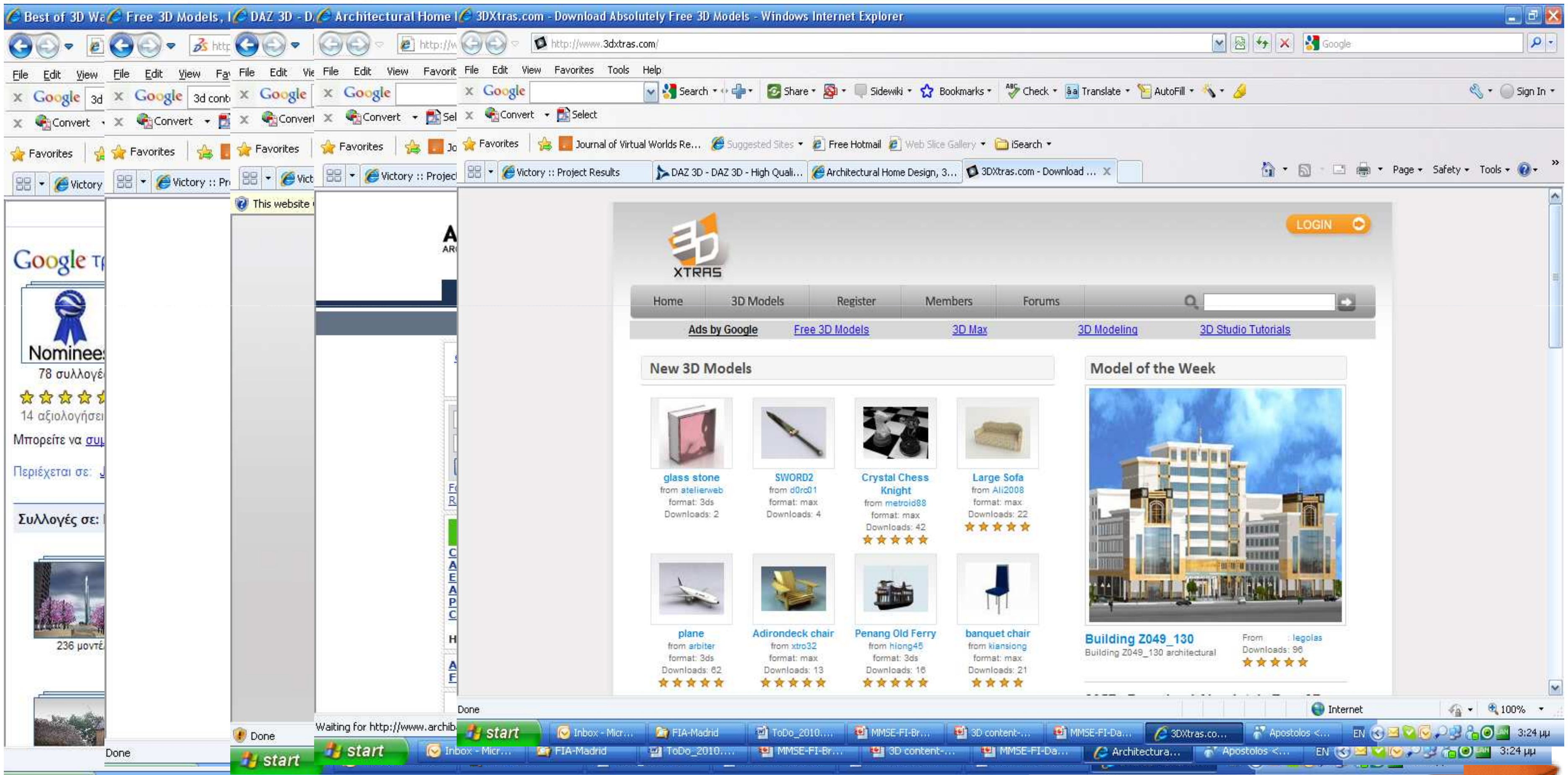
"A picture is worth a thousand words"



3D content: the king…

A 3D Object?

3D Content availability



The challenge...

http://www.victory-eu.org:8080/victory/results/search.html

The problems…

Problems to be faced:

(i) 3D objects' degeneracies (e.g. holes, missing polygons, hidden polygons),

(ii) 3D objects' pose normalization,

(iii) invariance to shape representations,

(iv) invariance to articulation or global deformation,

(v) the trade-off between the time needed for the extraction (which heavily depends on a 3D object's Level-of-Detail) and matching of 3D objects' descriptors

(vi) retrieval accuracy of a method

3D content-based search SoA

Possible solutions:

(i) Triangulation algorithm (e.g. Delaunay triangulation) or a hole filling algorithm

(ii) Pose normalization implies invariance with respect to rotation, scaling and translation of a 3D object.

  - (i) Scaling and translation normalization can easily be resolved
  - (ii) For rotation normalization two widely acceptable solutions have been presented in the literature, natively rotation invariant description of the 3D object and the rotation normalization of the 3D object in a pre-processing step.

(iii) Most of the existing methods work well either with polygonal meshes or polygon soups, while methods which rely on the topology of an object demand certain shape representation (e.g. watertight models)

(iv) The invariance to articulation is a hot research problem which has not been widely addressed so far and requires the extraction of local descriptors. A solution to the latter problem might also lead to more efficient partial matching algorithms.

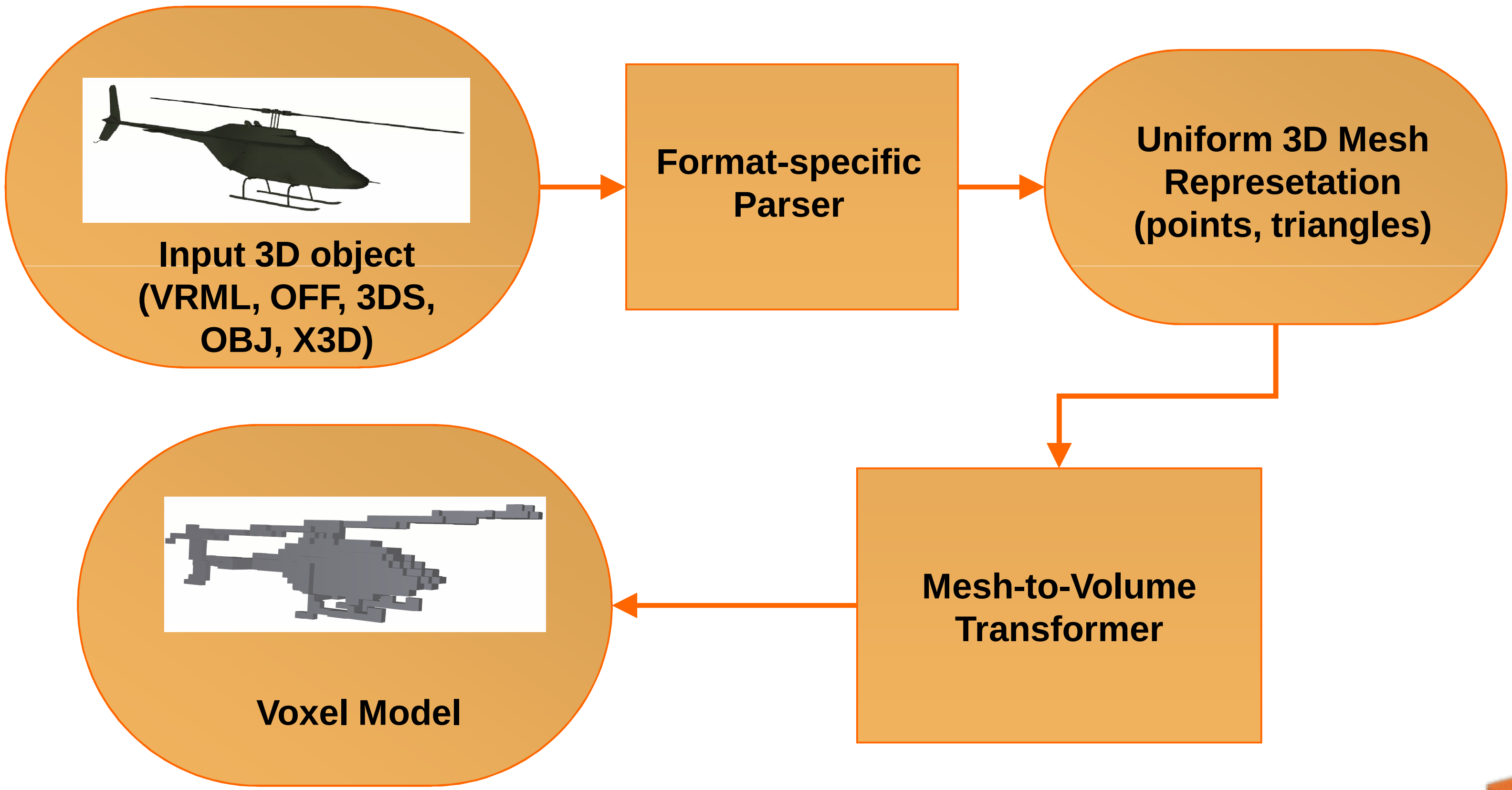
3D content-based search SoA

Category	Comments
Histogram-based	▪ Accumulators of local or global features ▪ Easy to implement ▪ Not sufficiently discriminative
Transform-based	▪ Uses signal processing transforms (Fourier, Spherical harmonics,…) ▪ Descriptor compaction ▪ Pose invariance can be obtained at the expense of shape information
2D view-based	▪ 3D = collection of 2D views ▪ Highly discriminative
Graph-based	▪ Can encode topology ▪ Hard to obtain ▪ Requires graph matching

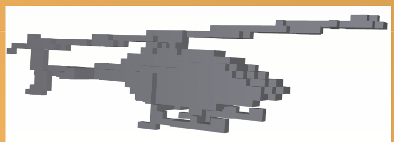
ITI algorithms

- 3D Low-level feature Extraction Algorithms (transform-based)
  - Generalized 3D Radon Transform
  - Spherical Trace Transform
  - 3D Krawtchouk Moments

Pre-processing Sub-module



Descriptor Extraction Sub-module



Voxel Model

Low-level Feature Extraction Algorithms

Generalized 3D Radon Transforms
- 3D Radon Transform
- RIT
- EnRIT
- SIT

Spherical Trace Transform
- "Mutated" RIT
- 1D Fourier Transform
- Radon Transform
- Polar-Fourier Transform
- Hu Moments
- Zernike Moments
- Krawtchouk Moments
- Spherical Harmonics

Weighted 3D Krawtchouk Moments

Geometric Descriptor Vectors

3D Radon – based Descriptors

Spherical Trace – based Descriptors

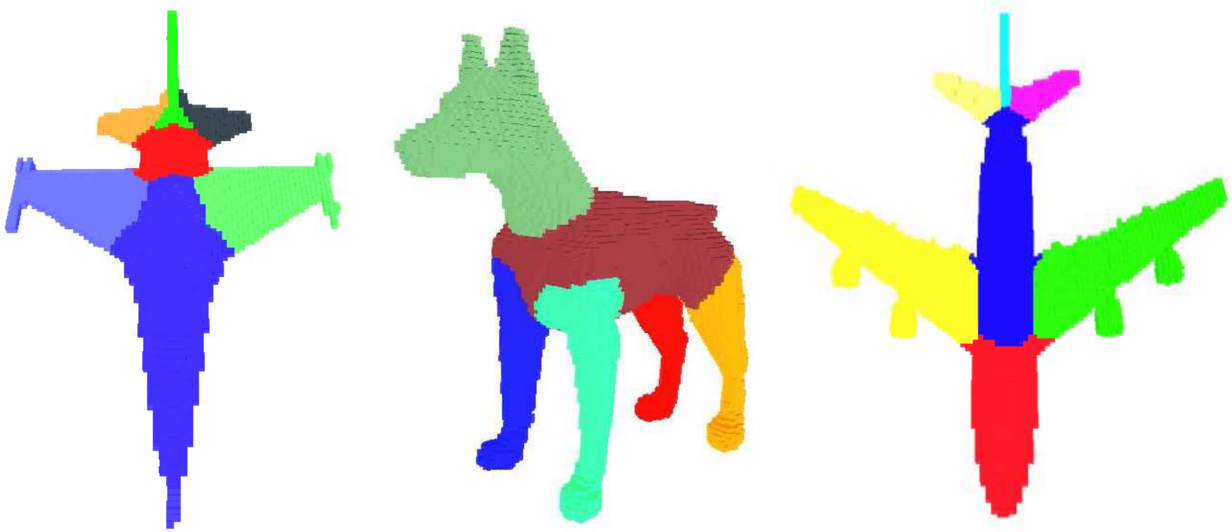
3D Krawtchouk – based Descriptors

ITI - New algorithms

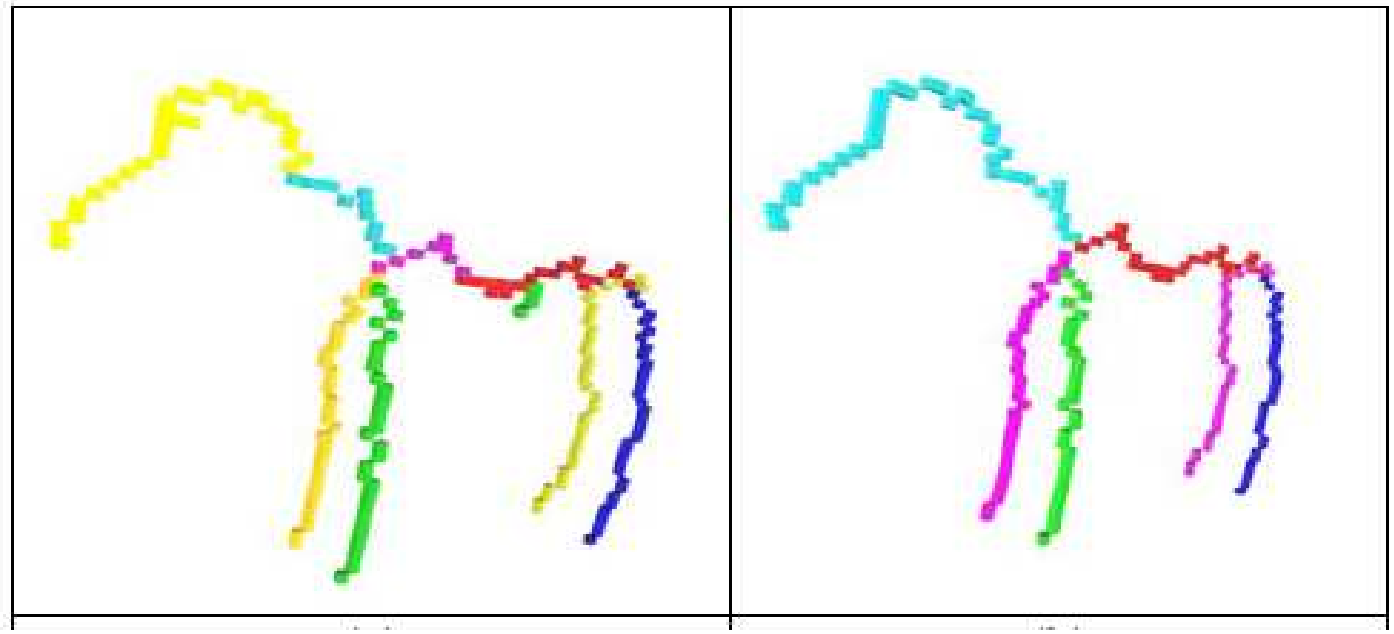
- Feature extraction algorithm based on 3D model segmentation and graph matching
  - Graph & Transform based method
- Feature extraction algorithm based on ellipsoidal harmonics
  - Transform-based method
- Feature extraction algorithm using the 3D shape descriptor
  - Histogram-based method

3D model segmentation and graph matching

- The model is initially segmented into parts based on the segmentation of medial surface
- The model segments are readjusted (correction steps) in order to result in a "meaningful segmentation"



3D model segmentation and graph matching



Correction steps

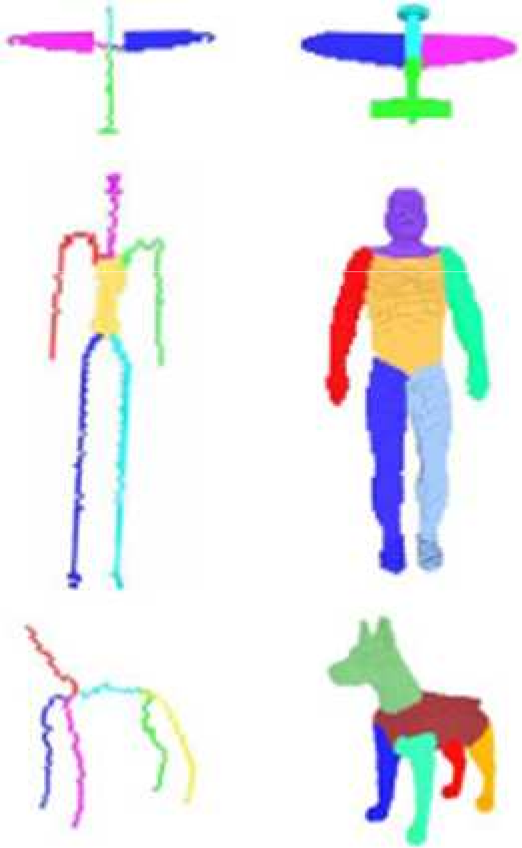
The Elimination Criterion: Segments whose size is considerably small, when compared with the overall medial surface size, are eliminated.

The Merging Criterion: All the adjacent line segments that are connected with nodes are merged into one segment. The degree of a node is the number of edges incident to that node. The Merging and Elimination Criterion result in the elimination of several noisy parts, while the general topology remains the same.

The Correction Criterion: This criterion determines a combined elimination-merging action in order to correct specific parts of the new medial surface. According to the Correction Criterion, a segment which lies between two branch nodes and its size is considerably small, when compared to the overall size, is eliminated and the two branch nodes are merged into one.

The Oversegmentation Criterion: This criterion determines if the object has been oversegmented. If the number of resulted parts is comparable to the number of medial surface voxels, then all the parts are merged in one part. In case of one segment, the method simply does not use the topological information and only the geometry is taken into account.

3D model segmentation and graph matching



3D model segmentation and graph matching

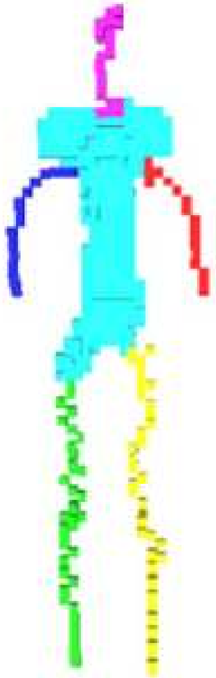
Why medial surface and not medial axis?

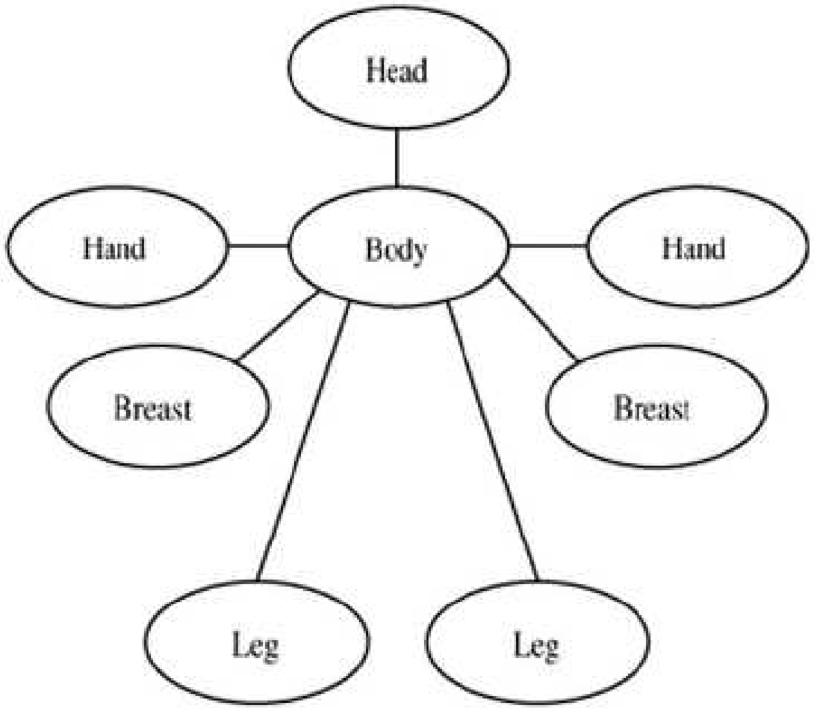
Medial axis based segmentation fails on segmenting 3D objects which contain large flat areas (e.g. tables, chairs) due to ambiguity of medial axis for this kind of 3D objects. In contrast, the medial surface is a well-defined transform for all 3D objects

Boundary voxels are more uniformly distributed along the medial surface, than the medial axis. As a result, the correction step of the medial surface segmentation algorithm produces more stable results.

3D model segmentation and graph matching

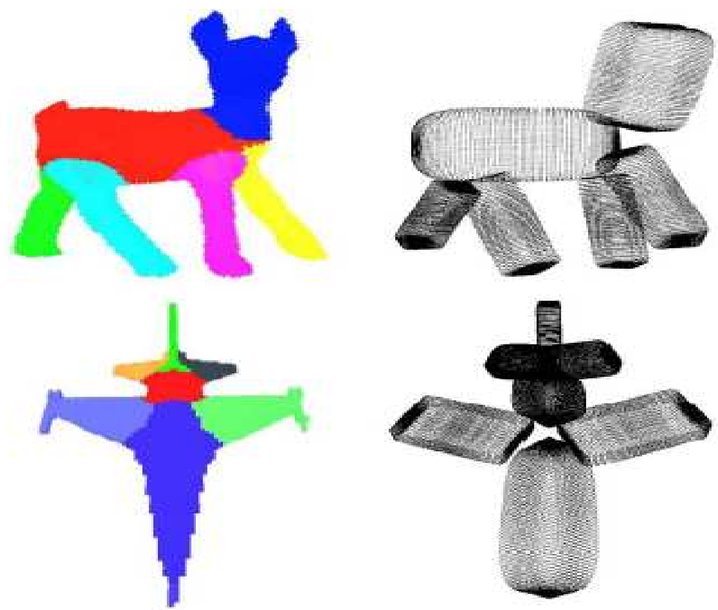
Transforming a medial surface to a graph





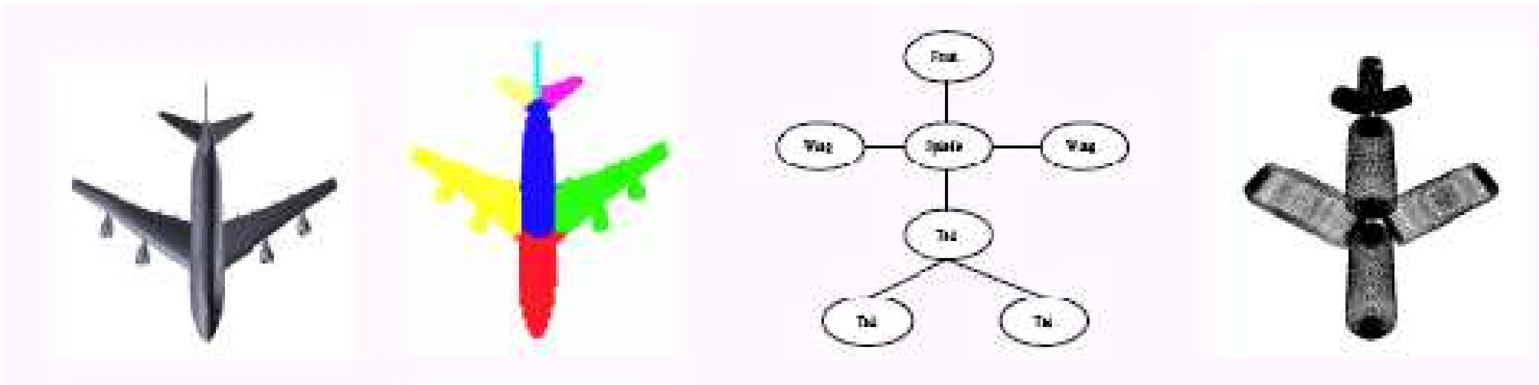
3D model segmentation and graph matching

- Every part is approximated with a Super Quadratic surface and described using the novel 3D Distance Field Descriptor (3D DFD - a descriptor that gives a measure of the difference between the surface of an ellipsoid and the surface of the object)



3D model segmentation and graph matching

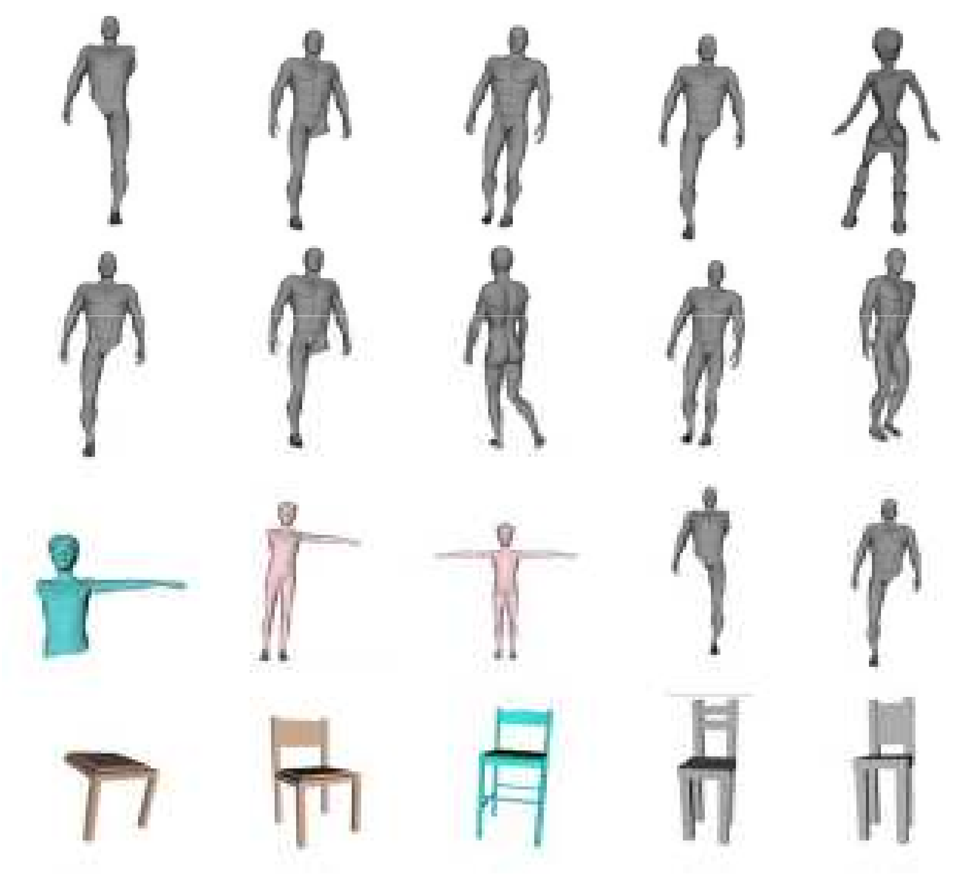
- An attributed graph is formed, where the attributes are the Super Quadric parameters and the 3D DFD
- The matching is based on state-of-art Approximate Attributed Graph Matching (Successive Projection Graph Matching - SPGM).



3D model segmentation and graph matching

Process	Average Time
Voxelization	9.3 sec
Medial Surface Extraction	12.5 sec
Segment Readjustment	1.1 sec
Superellipsoid Approximation (per segment)	4.7 sec
Distance Field Descriptor (per segment)	3.1 sec
Average Processing Time per object	91.5 sec
Matching two 3D objects	0.1 msec

Results – partial matching



Results – global matching (SHREC)

Rank	Approach	ADR
1	LFS	0.549
2	Proposed method	0.525
3	STT	0.524
4	ESD	0.500
5	CRSPD	0.495
6	PDS	0.493
7	C3DHTD	0.492
8	SEFWD	0.306
9	EDA	0.230

Average Dynamic Recall

Rank	Approach	1st-Tier
1	LFS	0.447
2	Proposed method	0.428
3	STT	0.427
4	CRSPD	0.418
5	PDS	0.409
6	C3DHTD	0.392
7	ESD	0.381
8	SEFWD	0.241
9	EDA	0.172

First Tier

Rank	Approach	2nd-Tier
1	LFS	0.2786
2	STT	0.2566
3	PDS	0.2563
4	Proposed method	0.2562
5	CRSPD	0.2561
6	C3DHTD	0.2501
7	ESD	0.2285
8	SEFWD	0.1525
9	EDA	0.1222

Second Tier

Pros & Cons

Advantages:

- The method can be efficiently used for both partial and global 3-D object search and retrieval.
- A novel combination based on the topological features and the highly discriminative geometrical features of the 3-D object is introduced
- The method doesn't require a perfect mesh as opposed to the other SoA methods

Disadvantages:

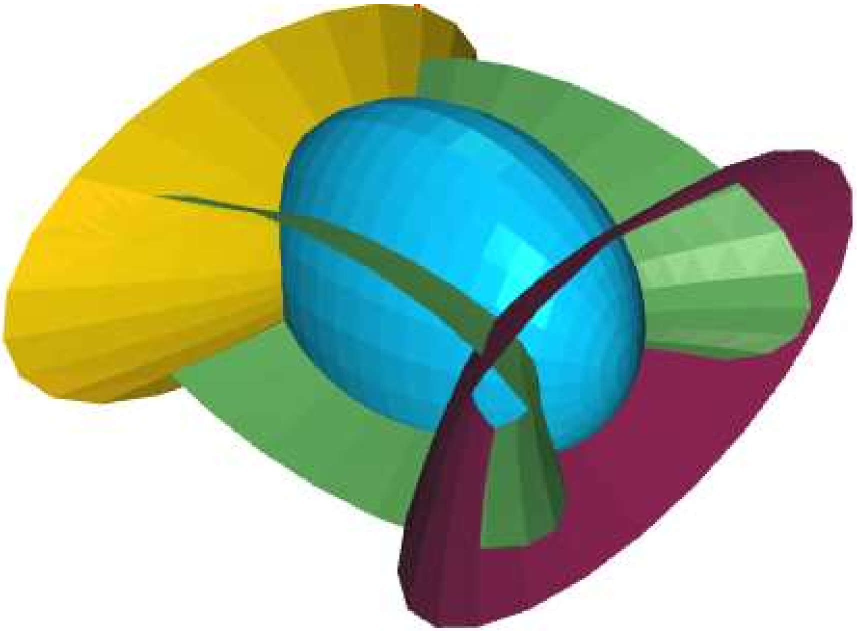
- In case a 3D model produces only one segment or a big number of segments, the topological information is lost! Instead, only the super-ellipsoid approximation and the extraction of descriptors based on the Distance fields are used, thus the retrieval accuracy is not so good…
- Low procedure

3D model segmentation and graph matching

- Relevant publication:
  - **A. Mademlis, P. Daras, A. Axenopoulos, D. Tzovaras, and M.G. Strintzis**: "*Combining Topological and Geometrical Features for Global and Partial 3D Shape Retrieval*", IEEE Transactions on Multimedia, Volume 10, Issue 5, August 2008

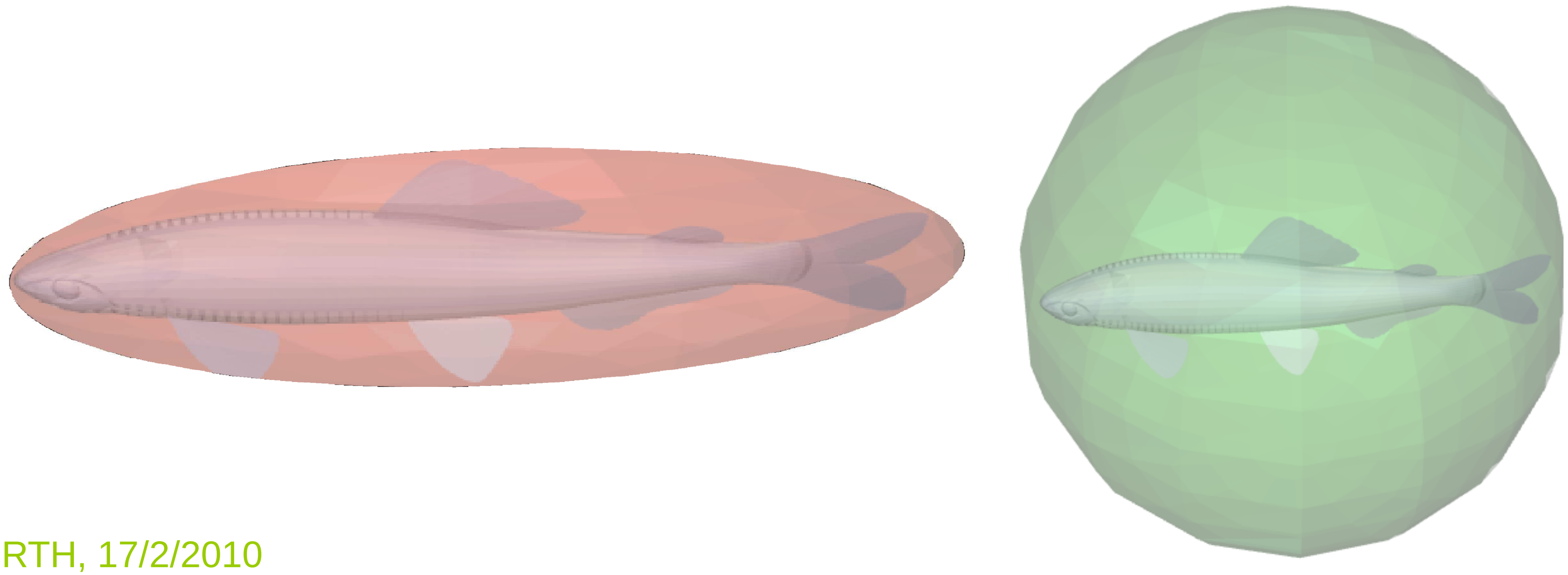
Ellipsoidal Harmonics

- A 3D object is described using Ellipsoidal Harmonics (Ellipsoidal Harmonics are solutions of Laplace’s equations in ellipsoidal coordinates)



Ellipsoidal Harmonics

- Ellipsoidal Harmonics offer:
  - Compact 3D object description
  - Better 3D object approximation



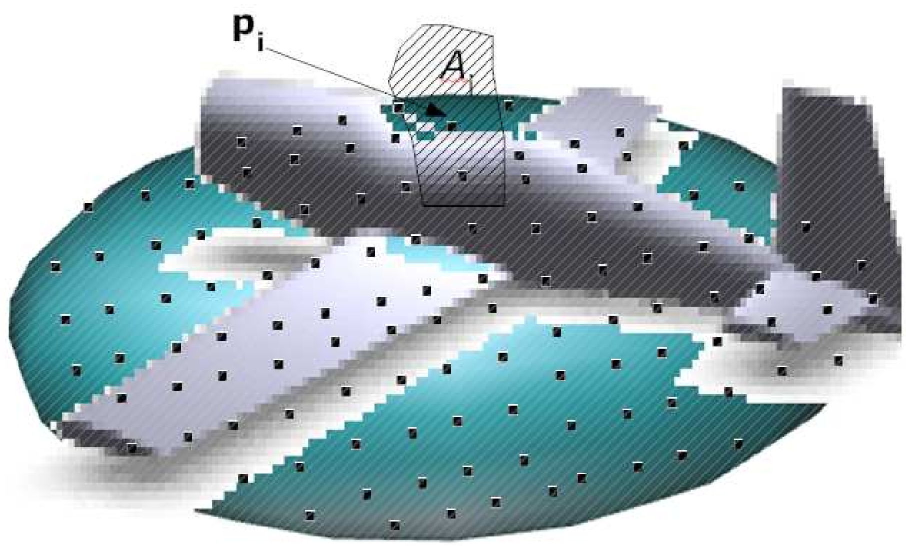
Ellipsoidal Harmonics

Four novel descriptors are introduced:

- The Surface Ellipsoidal Harmonics Descriptor (SEHD), which concerns 3D objects that are described as polygonal surfaces (is applied on the surface),
- The Volumetric Ellipsoidal Harmonics Descriptor (VEHD), which is applicable to volumetric 3D objects (is applied on the volume - voxels),
- The Generalized Ellipsoidal Harmonics Descriptor (GEHD) that is applied to any local 3D object descriptors and,
- The Combined Ellipsoidal-Spherical Harmonics Descriptor (SH-EHD), which leads to a compact and powerful descriptor that inherits the advantages of both approaches: the rotation invariance properties of the Spherical Harmonics and the directional information enclosed in Ellipsoidal Harmonics.

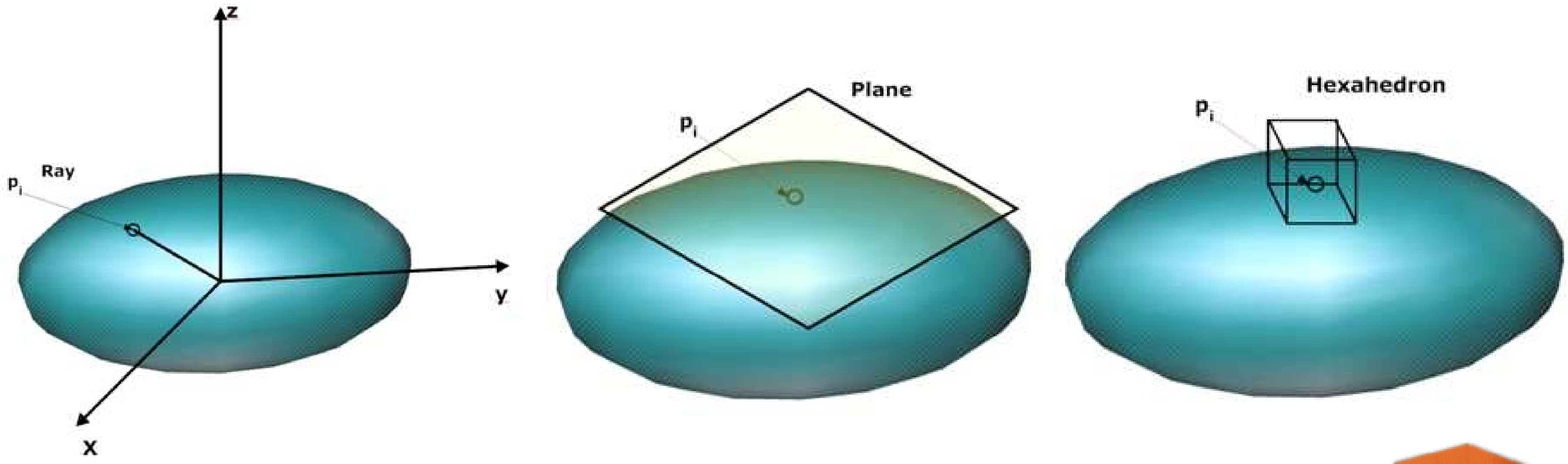
Generalized Ellipsoidal Harmonics

- Generalized Ellipsoidal Harmonics are the extension of Ellipsoidal Harmonic Analysis in order to be applied to local features
- The ellipsoid is sampled, and an area A_i is associated to every sample p_i



Generalized Ellipsoidal Harmonics

- The area A_i can be anything, 1D,2D or 3D. Indicative selections are: a ray, a plane or a hexahedron.



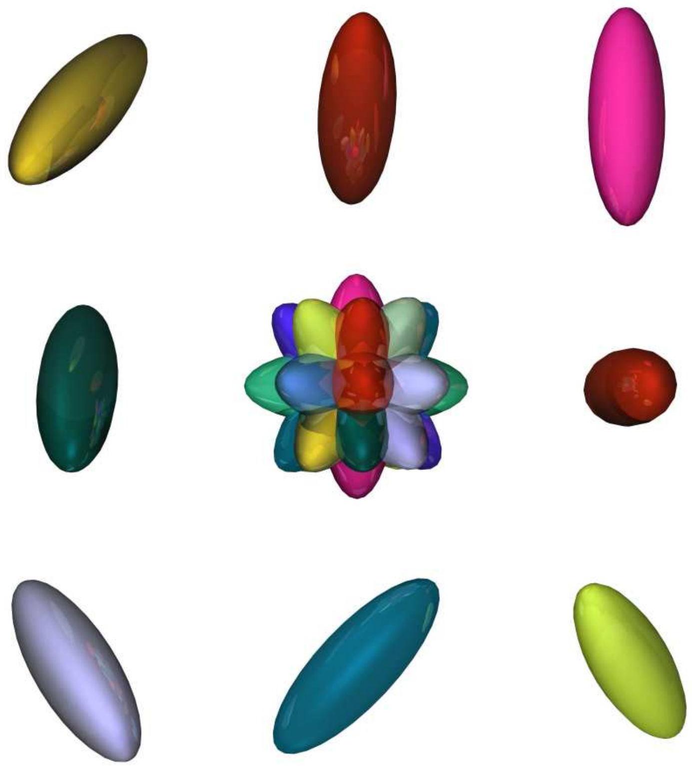
Generalized Ellipsoidal Harmonics

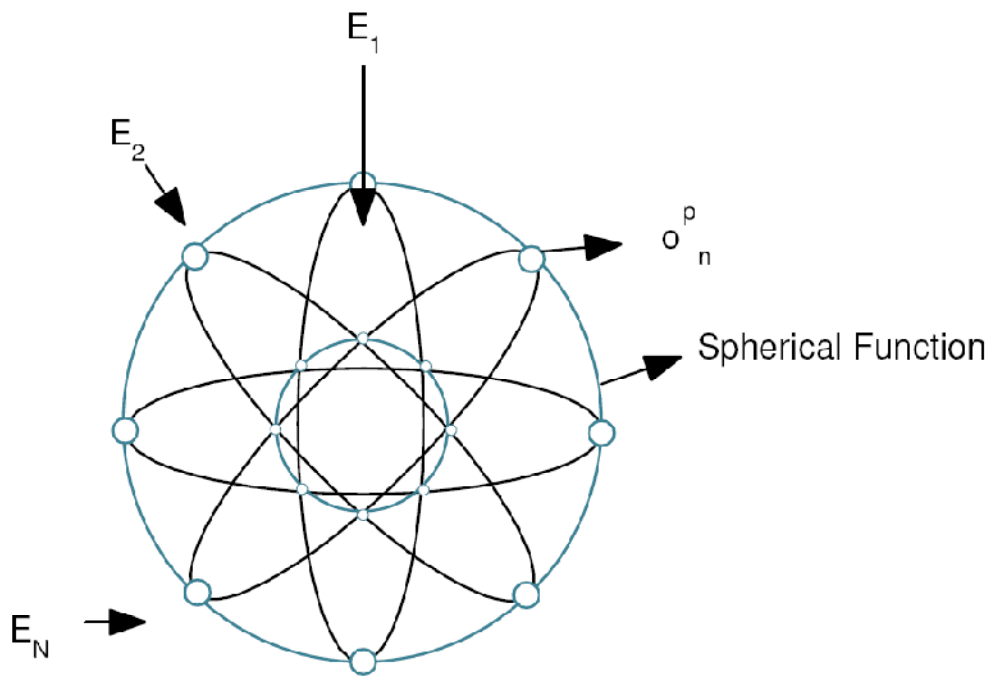
- An appropriate set of local descriptors is computed on every area A_i (e.g. Fourier, Krawtchouk, etc)
- Ellipsoidal harmonic analysis follows on each local descriptor forming the final descriptor.
- A_i now is a 5x5x5 cubic box where the 3D wavelet transform is applied

Combined Ellipsoidal and Spherical Harmonics

- The ellipsoidal harmonics coefficients are combined to spherical harmonic coefficients
- Integration of the directional information enclosed in ellipsoidal harmonics with rotation invariant properties of spherical harmonics

Combined Ellipsoidal and Spherical Harmonics



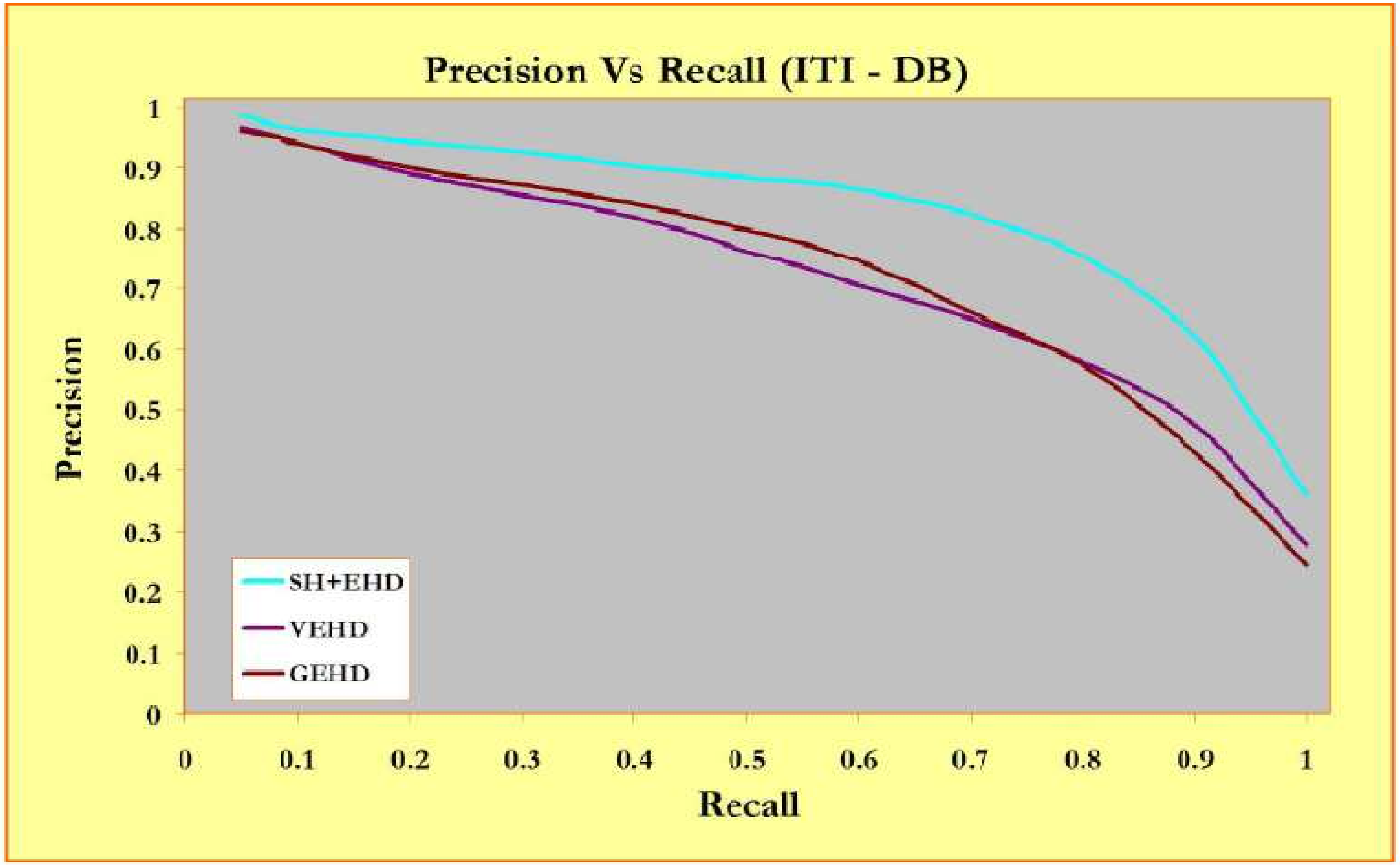


Complexity

SEHD: 225 descriptors (degree of exp =14, $d=(14+1)^2$)
VEHD: 1176 descriptors (d*24(num of ellisps)
GEHD: 2940 descriptors (d*12*5(wavelet coef))
SH-EHD: 6125 (49x5*5*5(SH coef))

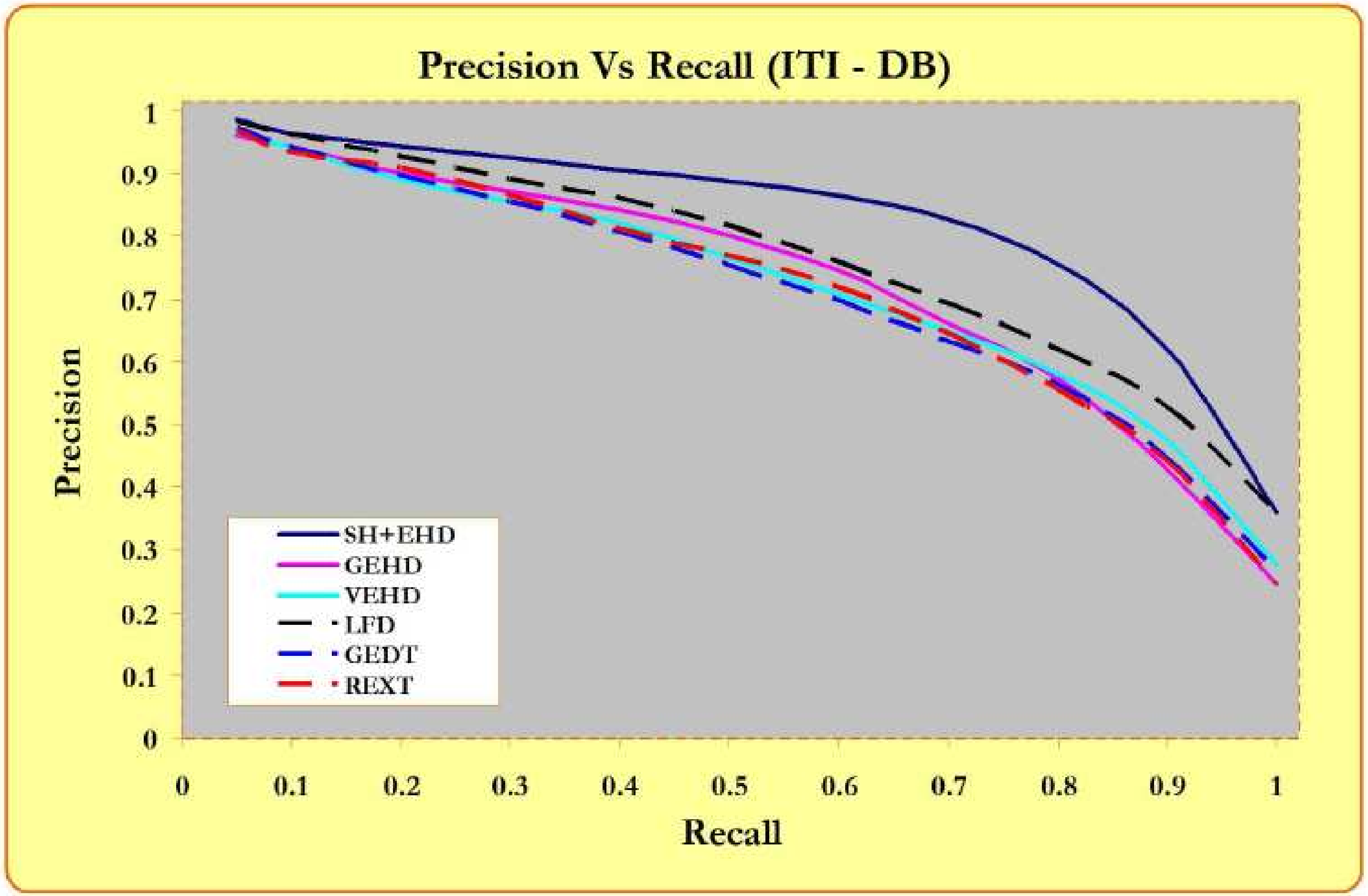
Method	Mean Computation Time
SEHD	297 msecs
VEHD	501 msecs
GEHD	675 msecs
SH+EHD	5.588 secs

Results

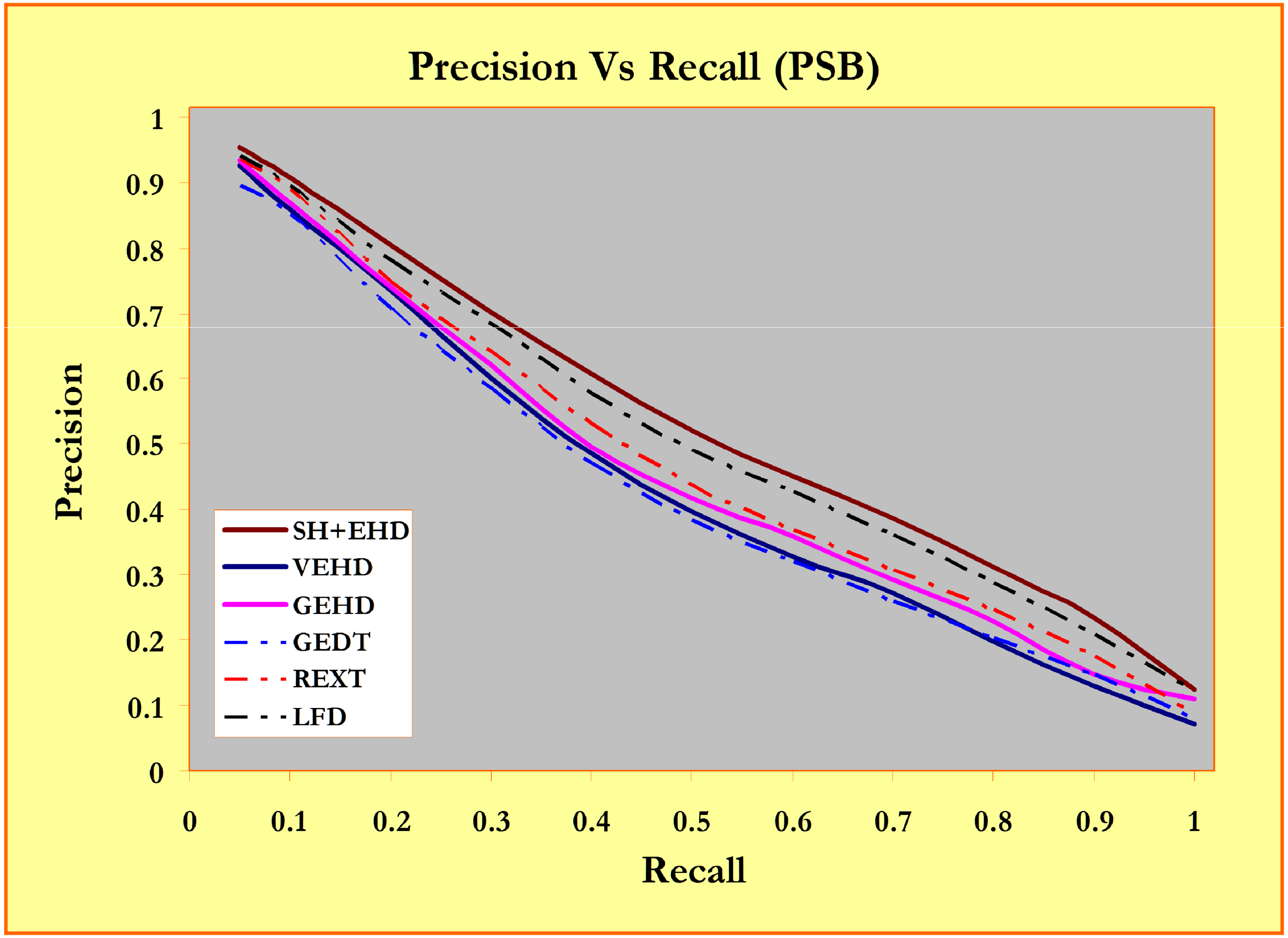


Precision Vs Recall (ITI - DB)
Precision
Recall
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
SH+EHD
VEHD
GEHD

Results

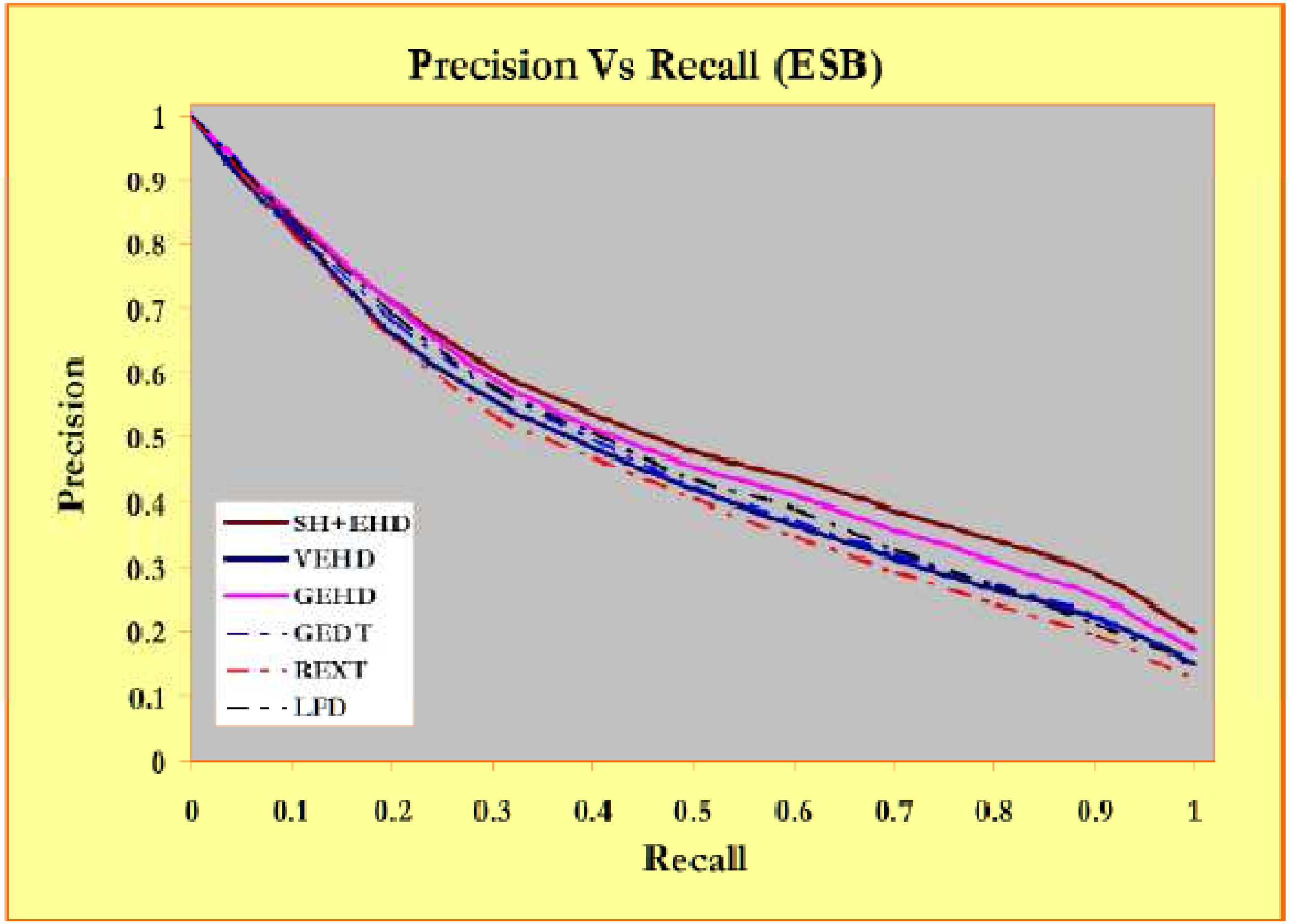

Precision Vs Recall (ITI - DB)
Precision
Recall
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
SH+EHD
GEHD
VEHD
LFD
GEDT
REXT

Results



Precision Vs Recall (PSB)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Precision
SH+EHD
VEHD
GEHD
GEDT
REXT
LFD
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Recall

Results



Precision Vs Recall (ESB)
Precision
Recall
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
SH+EHD
VEHD
GEHD
GEDT
REXT
LFD

Pros & Cons

Advantages:

- Very efficient method

Disadvantages:

- Big descriptor vector
- Time consuming in matching

Ellipsoidal Harmonics

Relevant publications:

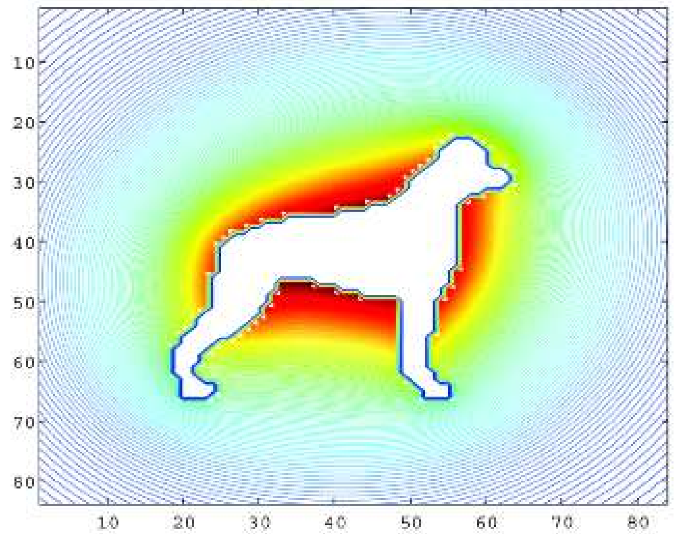
- **A. Mademlis, P. Daras, D. Tzovaras, and M.G. Strintzis**, *"Using Ellipsoidal Harmonics for 3D Shape Representation"* INTERMEDIA Workshop on Hypermedia 3D Internet, 13-14th October 2008, Geneva, Switzerland
- **A. Mademlis, P. Daras, D. Tzovaras and M.G. Strintzis**, *"Ellipsoidal Harmonics for 3D Shape Description and Retrieval"* , IEEE Transactions on Multimedia, Vol11, Issue 8, pp. 1422-1433, December 2009

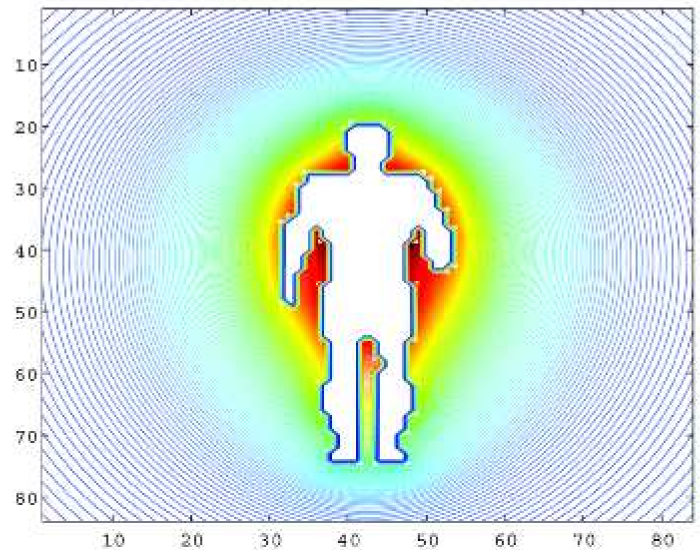
Impact Descriptor

- The key idea of Impact Descriptor is the indirect description of the 3D object's geometry, by computing features that describe the impact of the 3D object in the surrounding space.

Impact Descriptor

- Every object is treated as a distributed mass (volumetric representation) and the gravitational impact is described





Impact Descriptor

- Assuming that the time-space is following the rules of Euclidean geometry, generalizations of Newton’s Laws are utilized in order to compute the Potential and the Density of the surrounding gravitational field
- The values of the field’s potential and density in the surrounding area of the object form histograms, the Newtonian Impact Descriptor (NID)

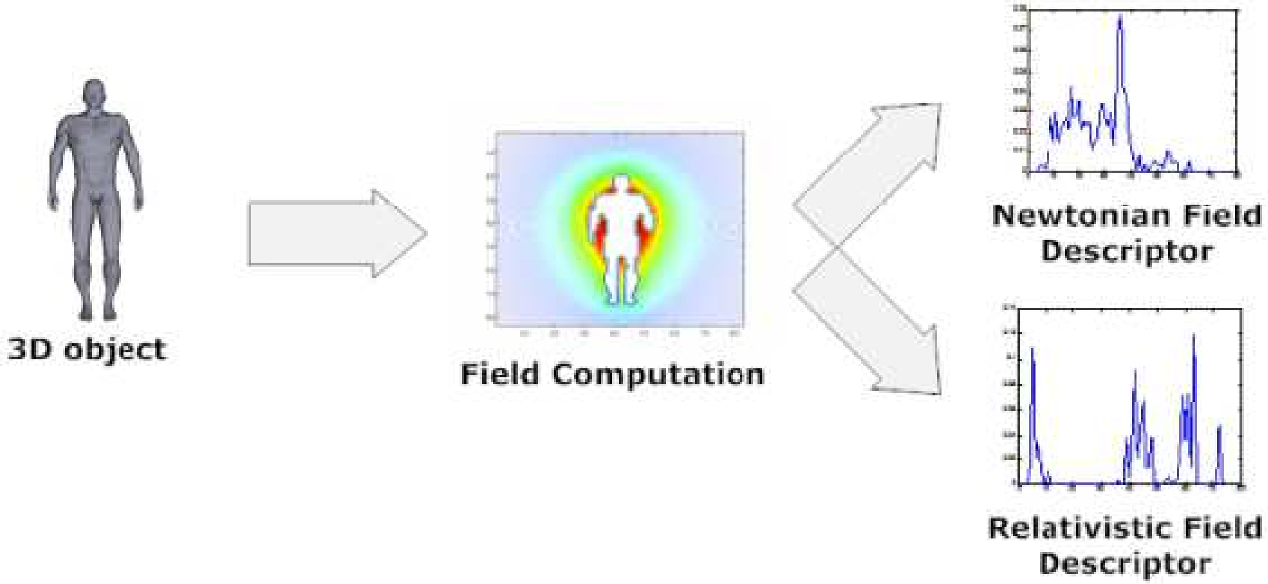
Impact Descriptor

- Assuming that the time-space is following the rules of Riemannian geometry, Einstein's General Relativity Laws are adopted in order to estimate the curvature of the surrounding time-space (or, how the 3D object curves the surrounding time-space).

The relativistic invariants that describe the curvature of the 4D space (time + 3D space) that is caused by the mass of the 3D object (according to the General Relativity laws) form the histograms of the Relativistic Impact Descriptor (RID).

Impact Descriptor

- NID and RID are combined and form the 3D Shape Impact Descriptor (SID)



NID

NID is composed of three major histograms created by:

- The field potential values, computed in points that are equidistant from the object surface
- The field density Euclidean norms, computed in points that are equidistant from the object surface
- The radial component of the field density, computed in points that are equidistant from the object surface

NID

$$\mathbf{E}(\bar{\mathbf{y}}_i) = \sum_{i=0}^{N} \frac{1}{|\bar{\mathbf{y}}_i - \mathbf{y}_i|^{K-1}} (\bar{\mathbf{y}}_i - \mathbf{y}_i)$$

$$\phi(\bar{\mathbf{y}}_i) = \sum_{i=0}^{N} \frac{1}{|\bar{\mathbf{y}}_i - \mathbf{y}_i|^{K-1}}$$

- These values are computed for 1-2 voxels distance
- Number of histograms = 6(K)x3(histograms)x2(voxels)=36
- K=[1,…,6]

RID

- Initially, it is assumed that the surrounding time-space of the object gets curved due to the 3D object's mass. Then, the surrounding space is sampled and at every sample the Einstein's gravity equation is solved
- Two invariants V1, V2 that characterize the time-space curvature are computed again in points that are equidistant from the object surface
- Two histograms are constructed using the values V1 and V2. These histograms capture the curvature of the surrounding time-space, due to the insertion of the object in the time-space.
- Number of histograms = 4

The 3D Shape Impact Descriptor - SID

Combination of NID & RID

Number of histograms=40

Complexity

Extraction time:

NID (36 histograms): 90 secs (mean time)

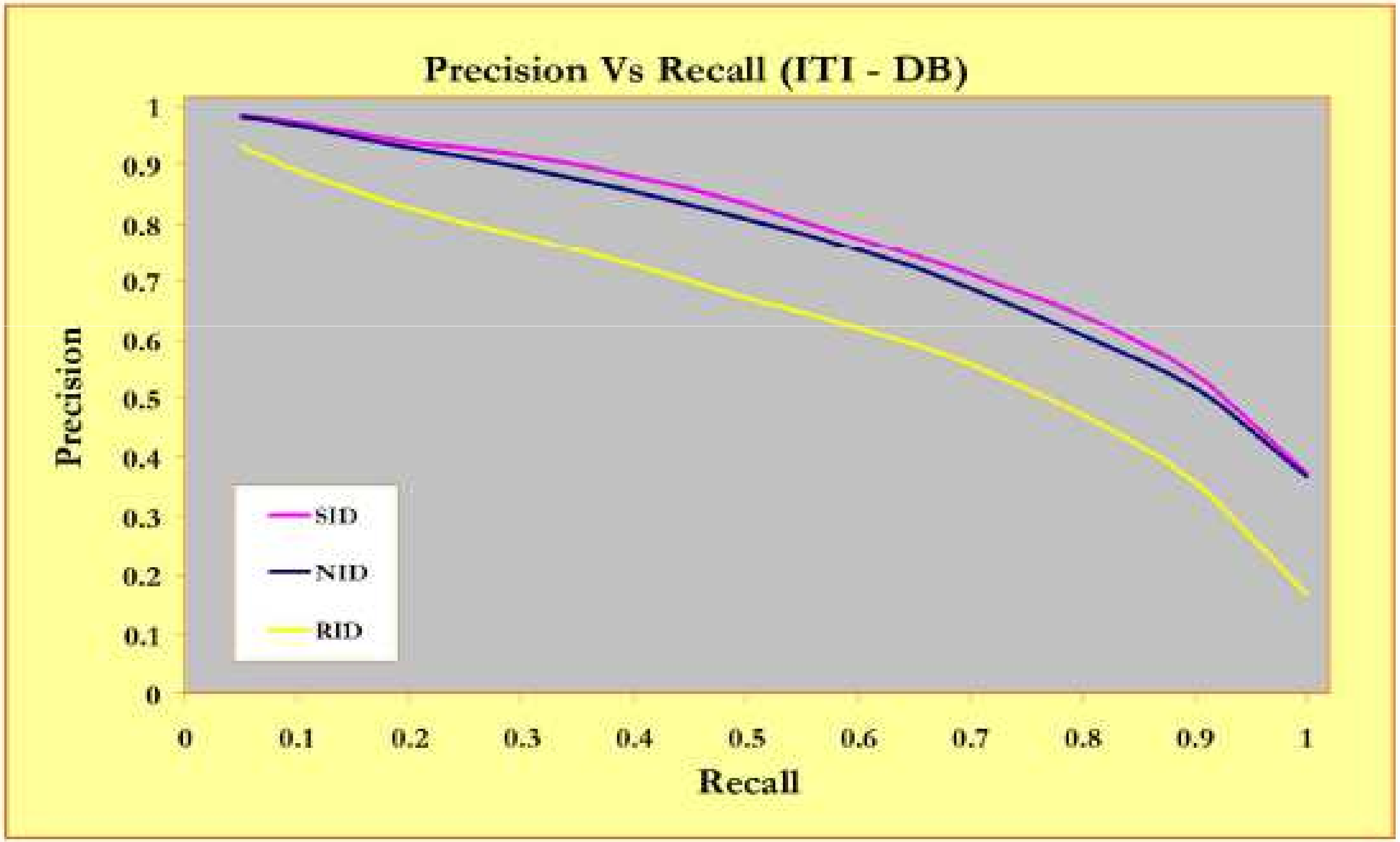
RID (4 histograms + input from NID): 40 secs

SID: 130 secs

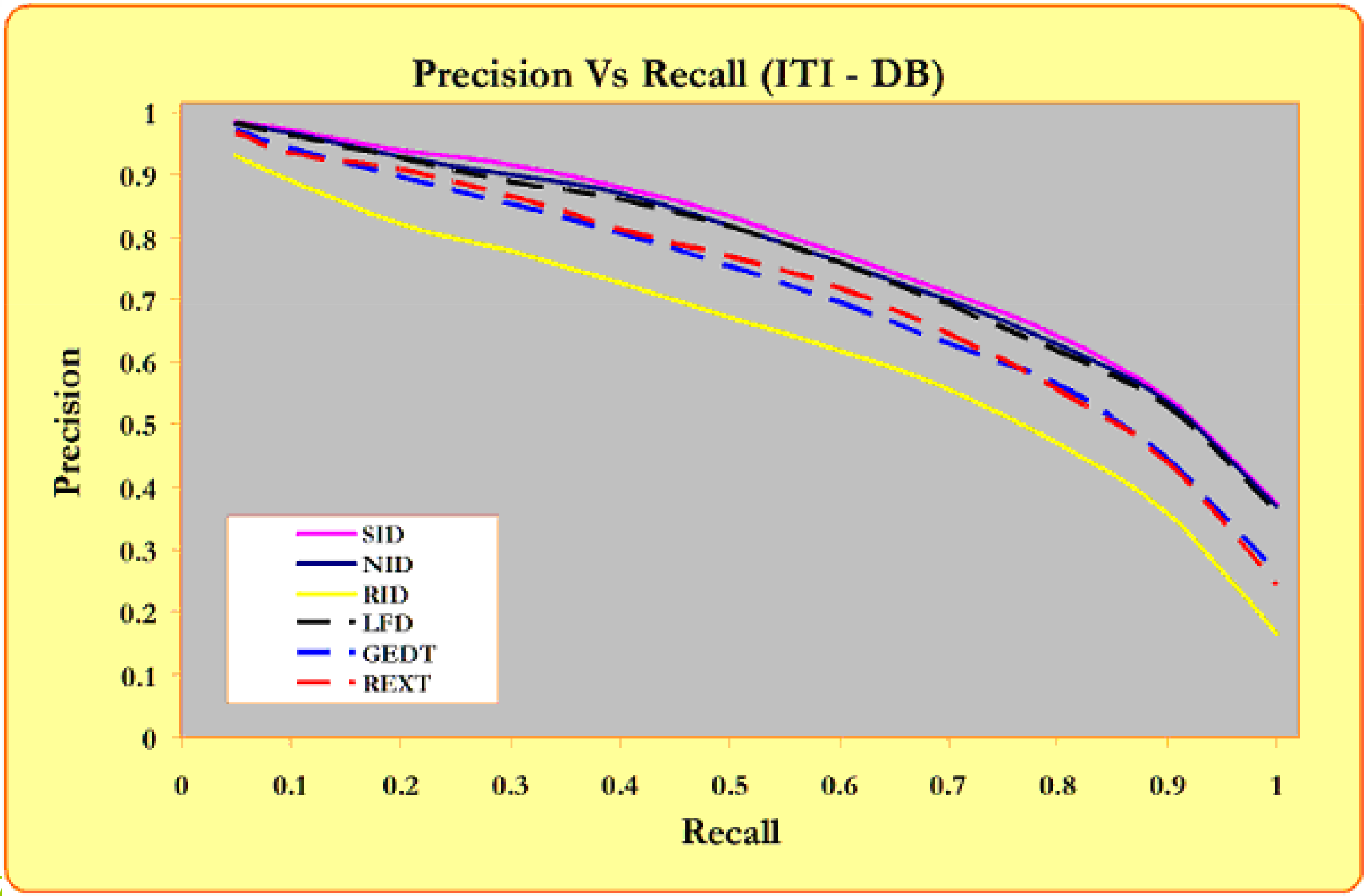
Execution time:

<msec

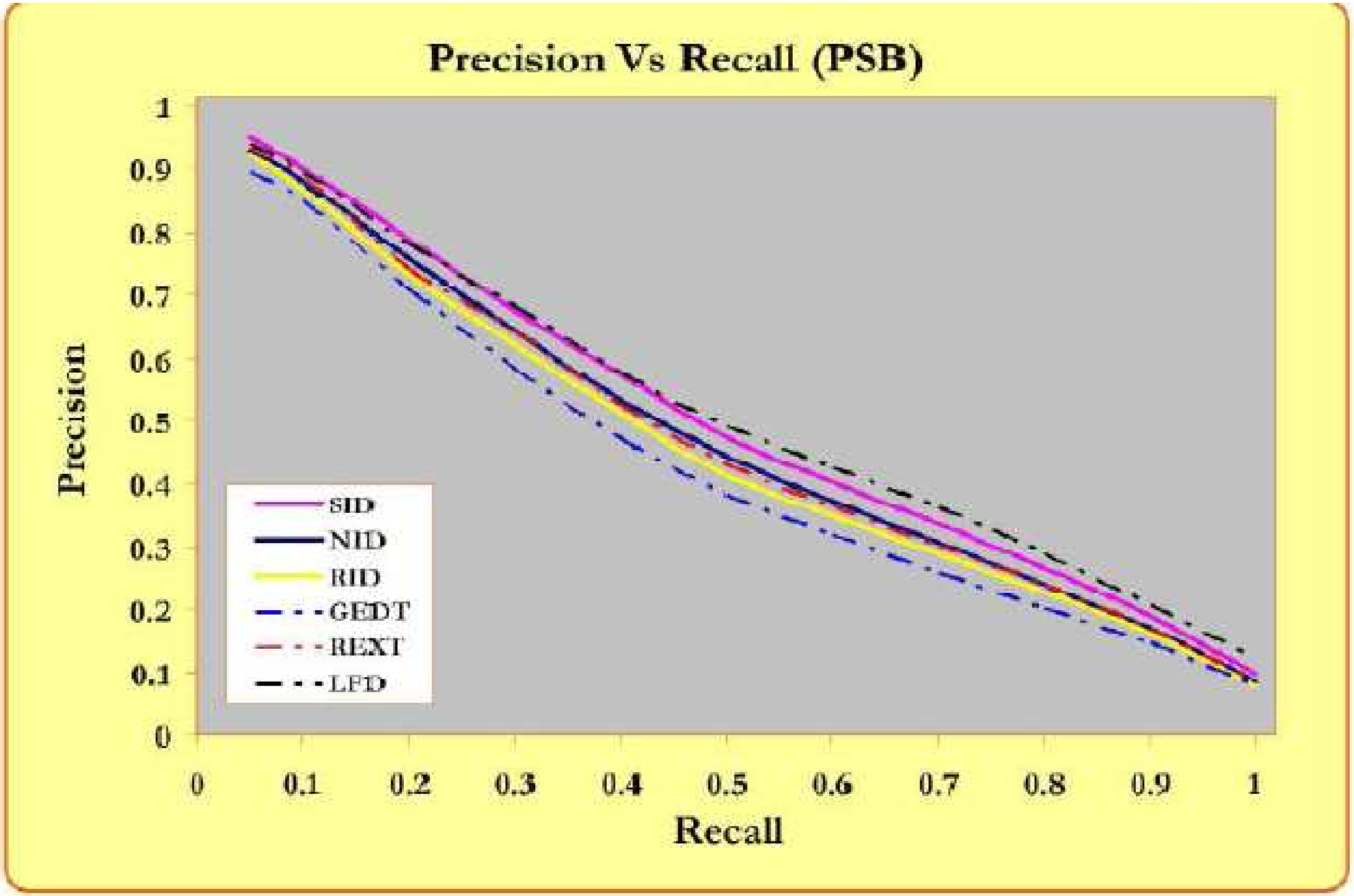
Results


Precision Vs Recall (ITI - DB)
Precision
Recall
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
SID
NID
RID

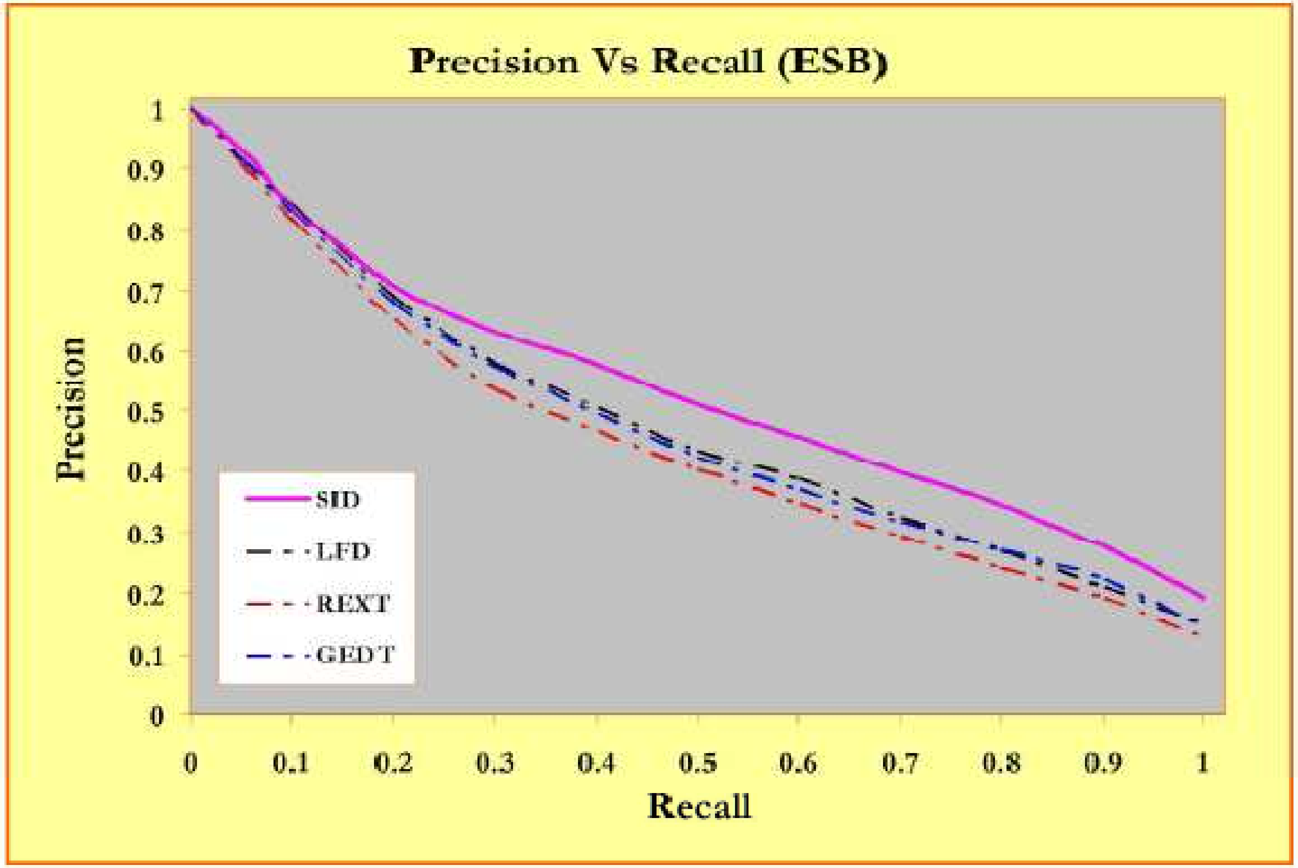
Results


Precision Vs Recall (ITI - DB)
Precision
Recall
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
SID
NID
RID
LFD
GEDT
REXT

Results


Precision Vs Recall (PSB)
Precision
Recall
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
SID
NID
RID
GEDT
REXT
LFD

Results



Precision Vs Recall (ESB)
Precision
Recall
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
SID
LFD
REXT
GEDT

Pros & Cons

Advantages:

- Histogram-based descriptors have been generally considered as descriptors with lower discriminative power, mainly due to their statistical nature. SID is a histogram-based descriptor, which proves that the appropriate selection of the values that construct the histograms is crucial for the discriminative power of the resulting histogram-based descriptor and that histogram-based descriptors can potentially provide highly discrimination.
- Natively rotation invariance
- Theoretically, it captures more information than the view-based methods
- The method can be expanded for moving 3D objects

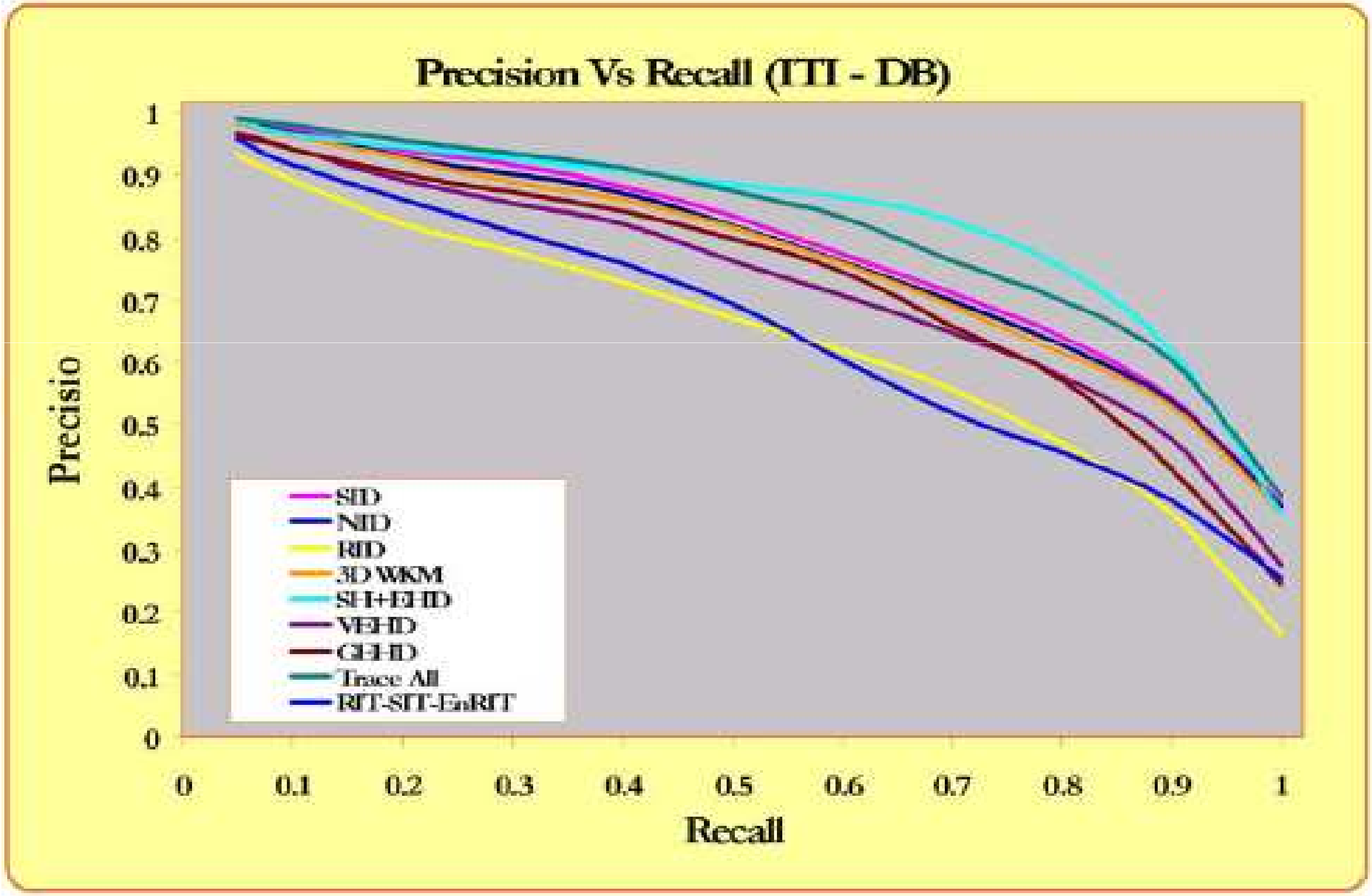
Disadvantages:

- Slow process (for histogram extraction)
- The retrieval accuracy is still not so good …

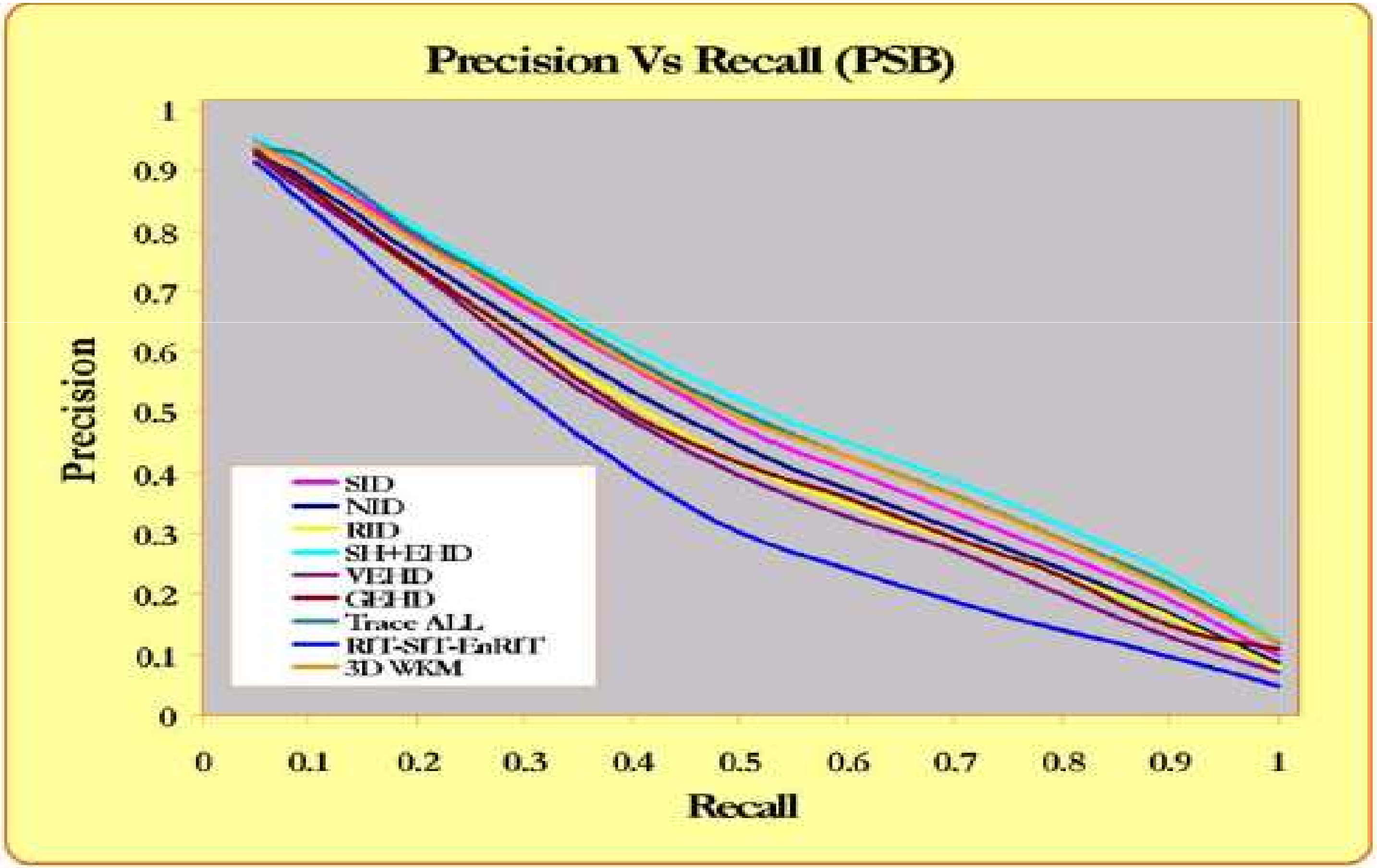
Publications

- Relevant publications
  - **A.Mademlis, P.Daras, D.Tzovaras and M.G.Strintzis**: "*3D Object Retrieval based on Resulting Fields*", 29th International conference on EUROGRAPHICS 2008, workshop on 3D object retrieval, Crete, Greece
  - **A. Mademlis, P. Daras, D. Tzovaras and M.G. Strintzis,** "*3D Object Retrieval using the 3D Shape Impact Descriptor*", ELSEVIER, Pattern Recognition, Volume 42, Issue 11, November 2009, Pages 2447-2459

Comparative Performance Evaluation of all Methods



Comparative Performance Evaluation of all Methods



Combination of View-based and Transform-based methods

- From the aforementioned analysis it is obvious that a method able to work on different shape representations and not perfect meshes (thus it cannot be topology-based), be highly discriminant (thus it cannot be histogram-based) and combine the advantages of the view-based (high discrimination) and transform-based methods (3D information), is of great importance.

View-based method

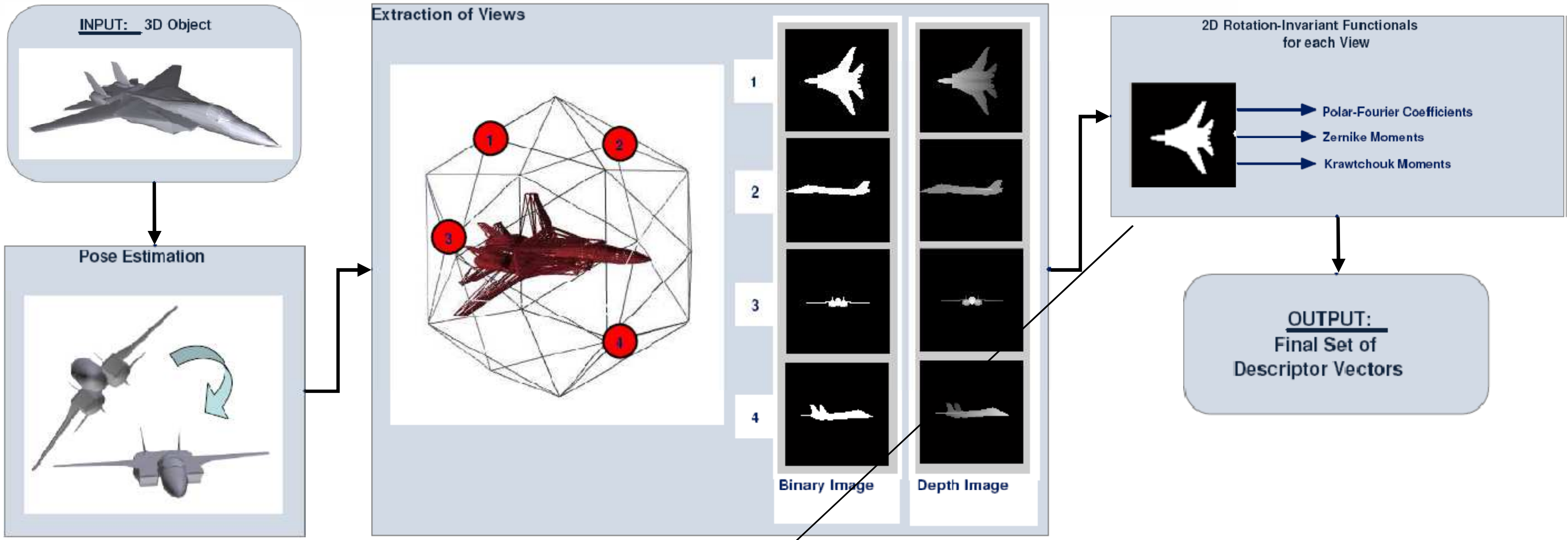
The proposed unified framework for 3D object retrieval:

- Supports multiple types of queries (3D objects, hand-drawn sketches, 2D images).
- Is based on 2D matching of multiple views, taken from a 3D object at uniformly distributed viewpoints.
- Supports both binary (black/white) images and depth images.

Advantages:

- High discriminative power (especially when combined with a transform-based method)
- Robustness to object degeneracies, holes, missing polygons, level of detail (LoD)
- Suitable for partial matching and articulated objects
- Faster descriptor extraction and matching than existing view-based methods
- Use of Depth Images (captures more details in a 3D object)

Descriptor Extraction



- Translation
- Scaling
- Rotation

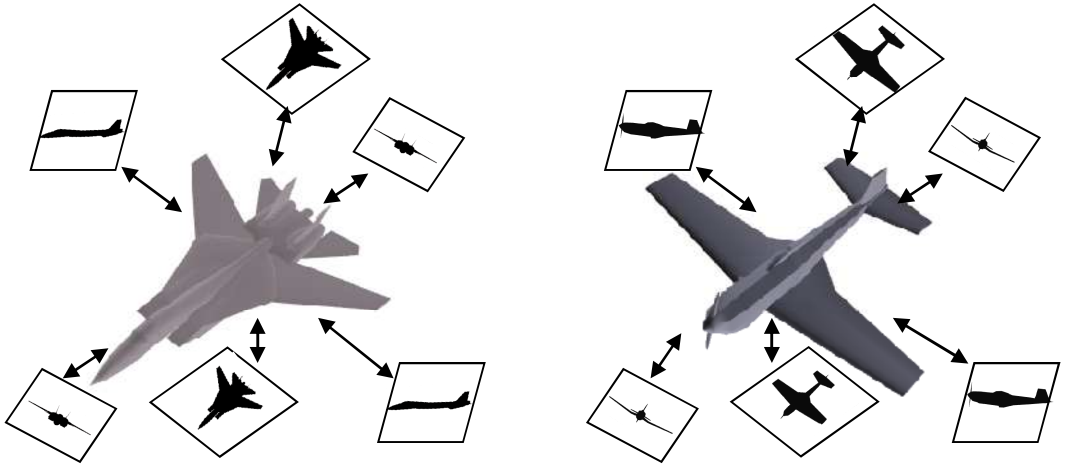
Uniformly distributed views

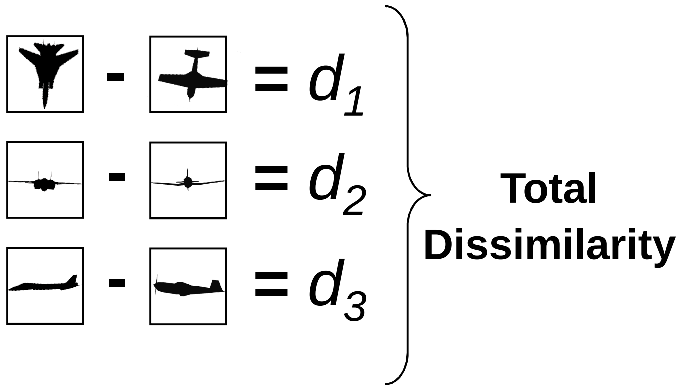
- 18 or 32
- Binary Images/Depth Images

2D Functionals

- Polar-Fourier Coefficients
- Zernike Moments
- Krawtchouk Moments

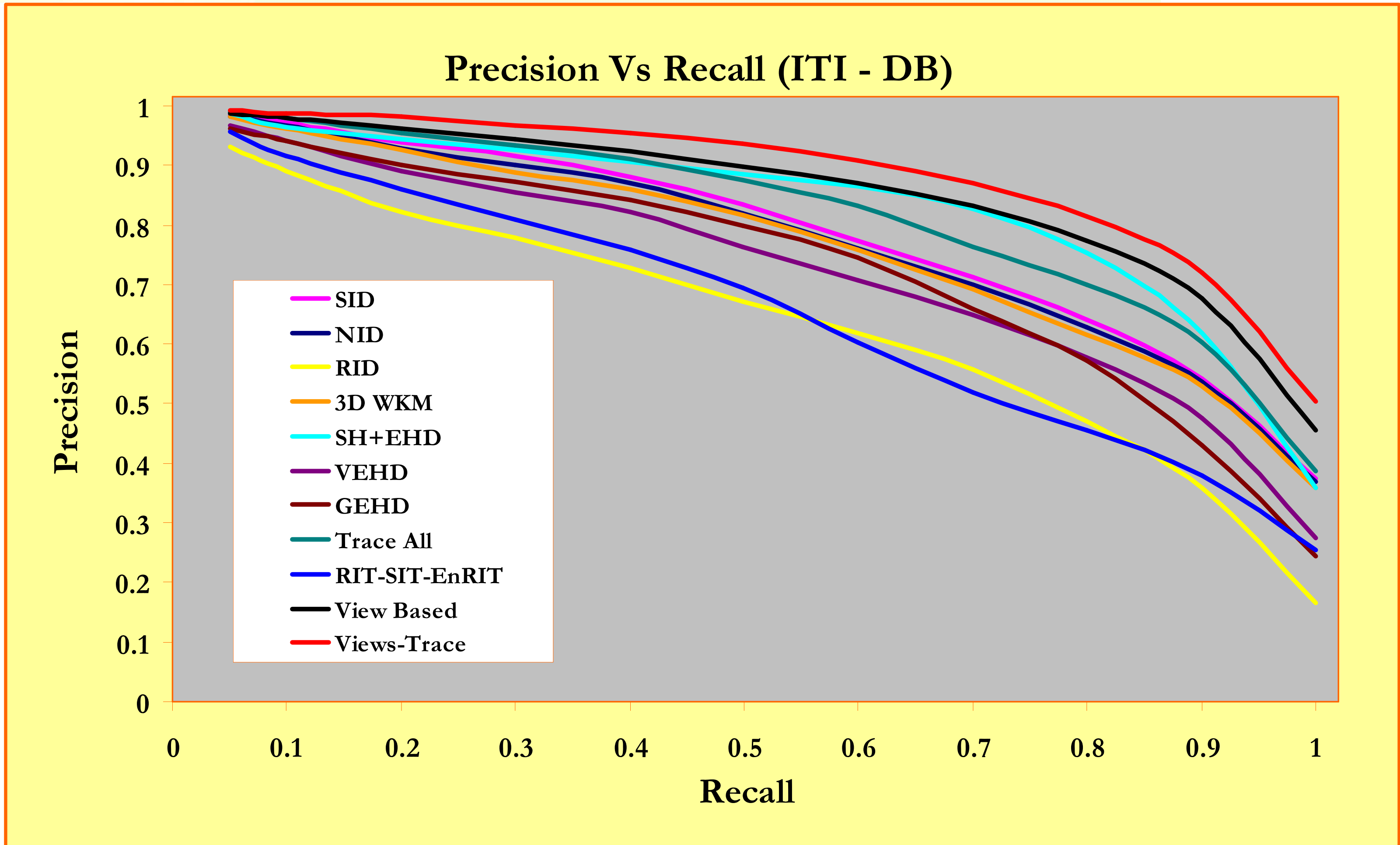
3D/3D Matching



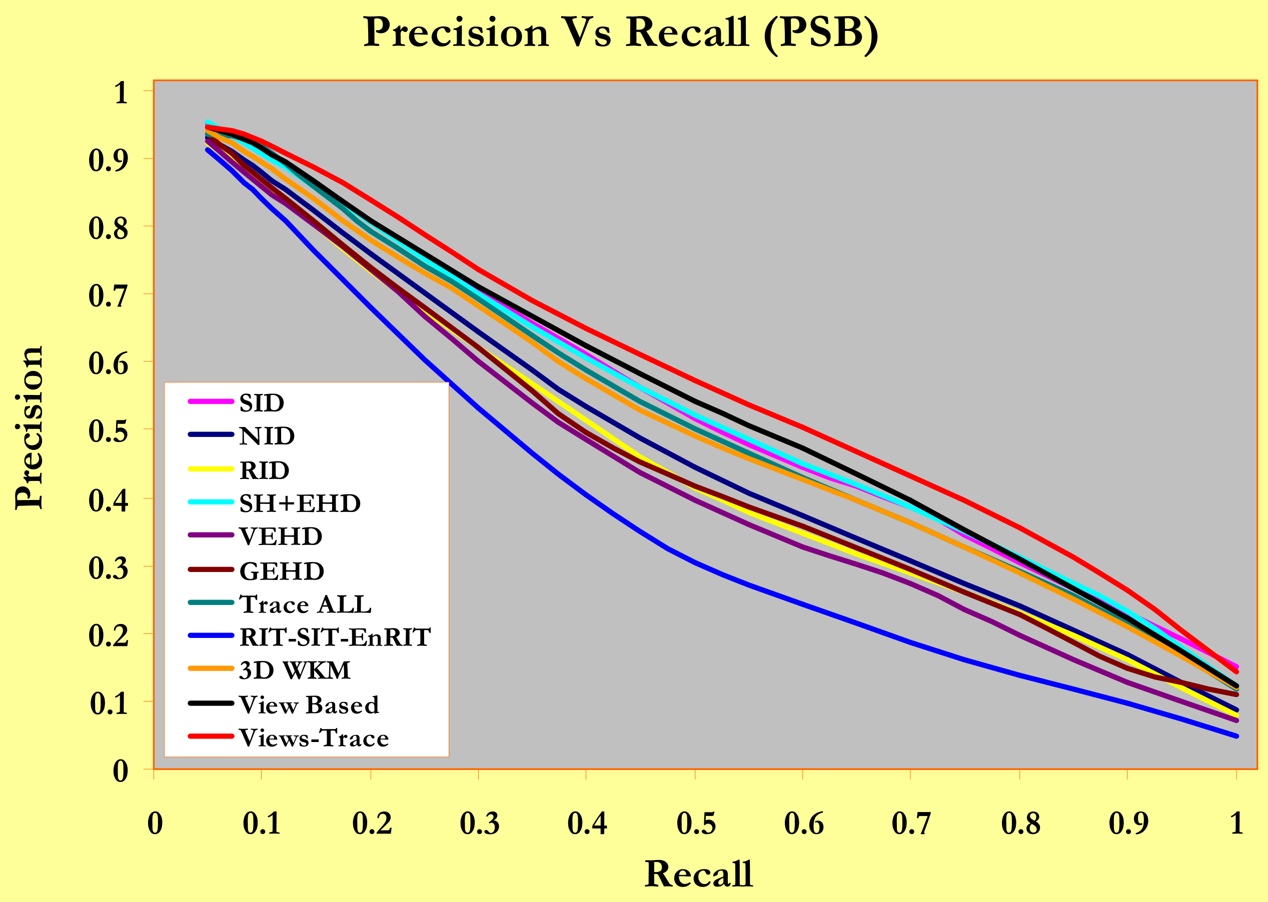


Matching of corresponding views only

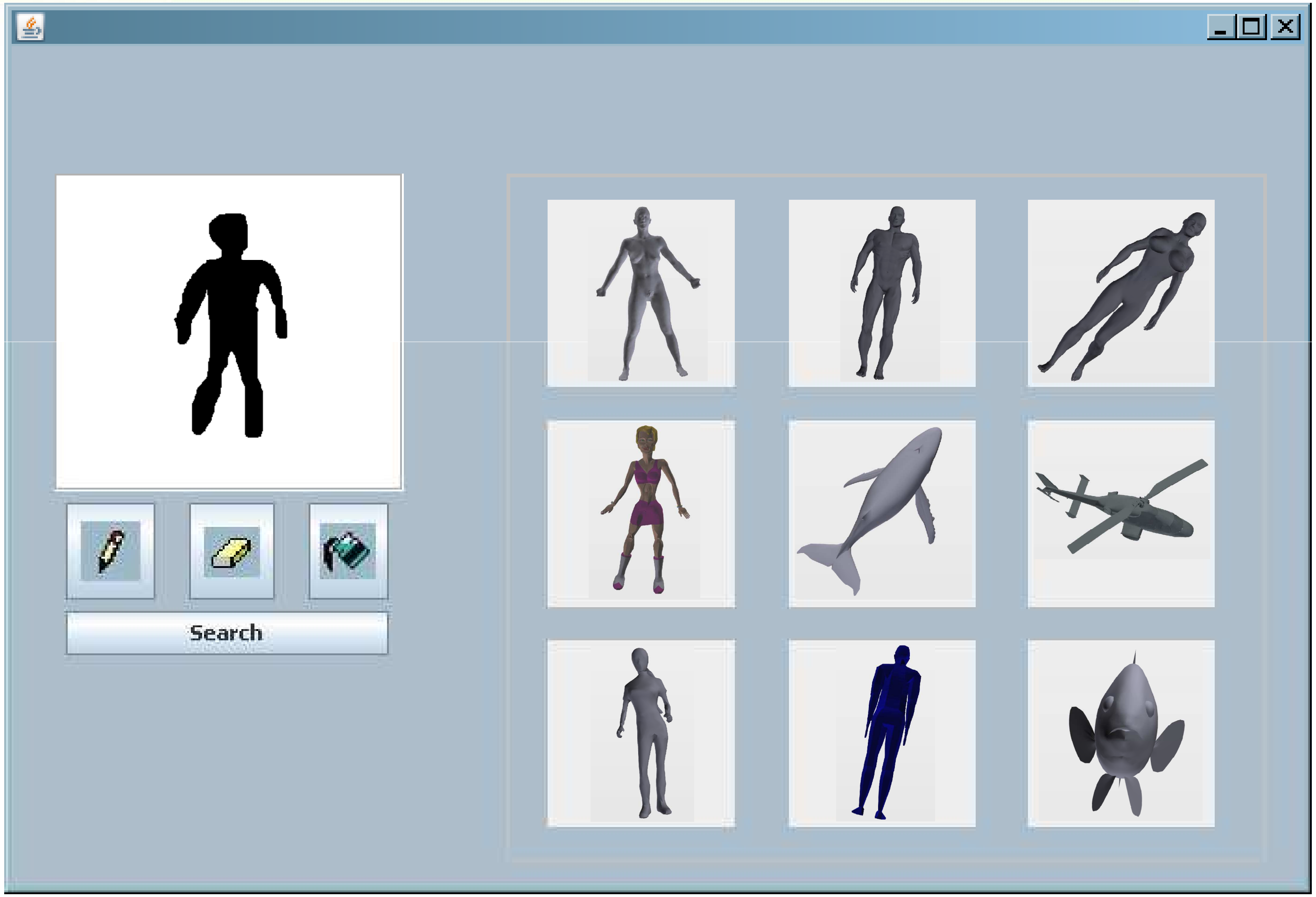
3D/3D Matching Results (ITI Database)



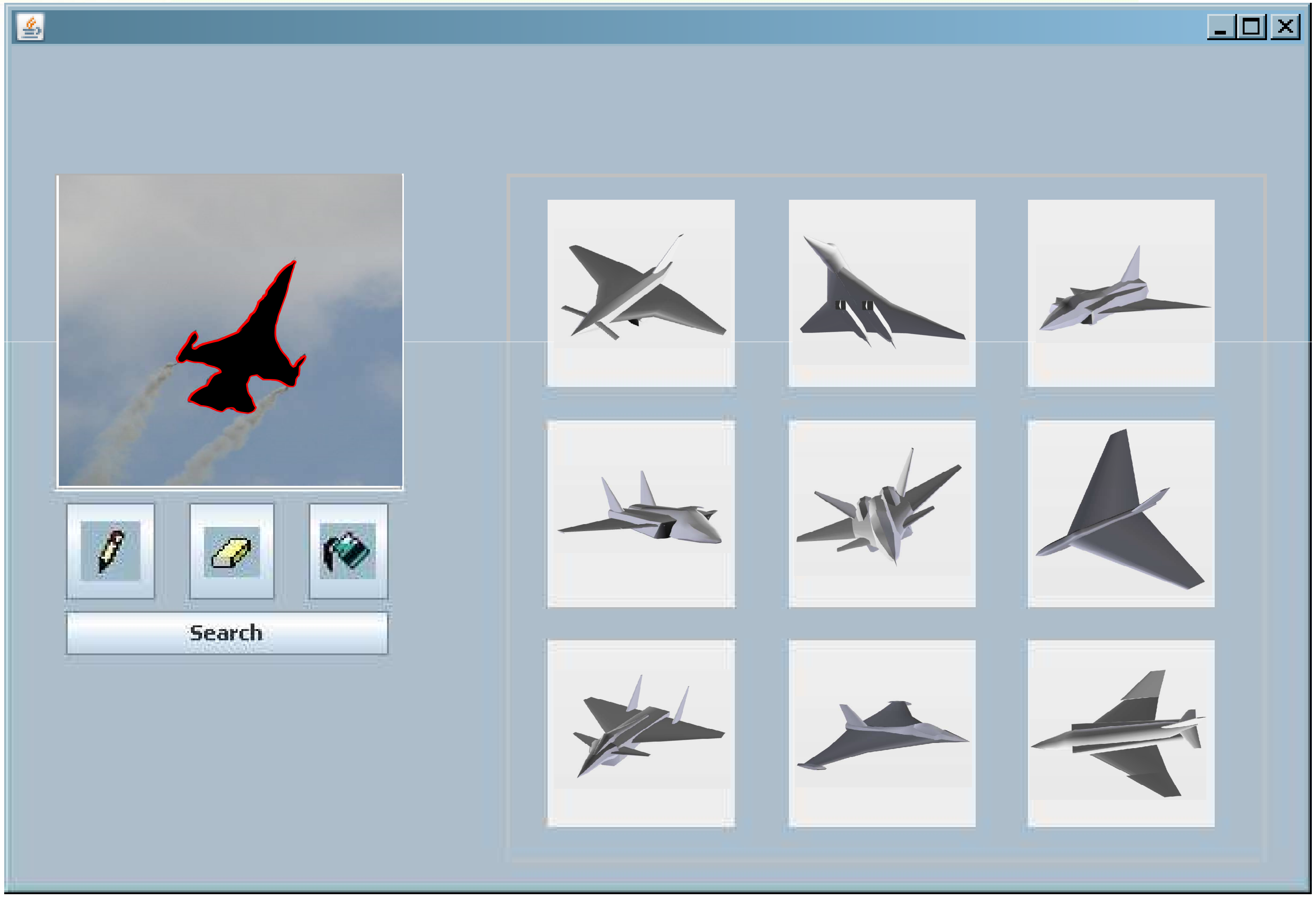
3D/3D Matching Results (PSB Database)



Sketch/3D Matching



Image/3D Matching



Complexity

Polar-Fourier Coefficients: 78 descriptors per view [(N*(N+1)/2), N=12]
Krawtchouk Moments: 78 descriptors per view [(N*(N+1)/2), N=12]
Zernike Moments: 56 descriptors per view (N/2+1)*((N +1)/2+1), N=13

Average Descriptor Extraction Time (msecs)	
Views Generation	2587
Polar-Fourier Descriptors	63
Krawtchouk Descriptors	398
Zernike Descriptors	811
Matching Time between 2 models	
10 msecs	

Publications

- P. Daras and A. Axenopoulos: "A 3D Shape Retrieval Framework Supporting Multimodal Queries", SPRINGER, International Journal of Computer Vision, DOI 10.1007/s11263-009-0277-2, July 2009
- P.Daras and A.Axenopoulos:"*A Compact Multi-View Descriptor for 3D Object Retrieval*", IEEE 7th International Workshop on Content-Based Multimedia Indexing (CBMI 2009), Chania, Greece, June 2009.
- A.Axenopoulos, P Daras, H.Dutagaci, T.Furuya, A.Godil and R.Ohbuchi:"*SHREC 2009 - Shape Retrieval Contest of Partial 3D Models*", 30th International conference on EUROGRAPHICS 2009, Workshop on 3D object retrieval, Munich, Germany, April 2009.
- J.Hartveldt, M.Spagnuolo, A.Axenopoulos, S.Biasotti, P.Daras, H.Dutagaci, T.Furuya, A.Godil, X.Li, A.Mademlis, S.Marini, T.Napoleon, R.Ohbuchi and M.Tezuka:"SHREC 09 Track: *Structural Shape Retrieval on Watertight Models*", 30th International conference on EUROGRAPHICS 2009, Workshop on 3D object retrieval, Munich, Germany, April 2009.
- C.Akgul, A.Axenopoulos, B.Bustos, M.Chaouch, P.Daras, H.Dutagaci, T.Furuya, A.Godil, S.Kreft, Z.Lian, T.Napoleon, A.Mademlis, R.Ohbuchi, P.L.Rosin, B.Sankur, T.Schreck, X.Sun, M.Tezuka, Y.Yemez, A.Verroust-Blondet and M.Walter:"*SHREC 2009 - Generic Shape Retrieval Contest*", 30th International conference on EUROGRAPHICS 2009, Workshop on 3D object retrieval, Munich, Germany, April 2009.

Research challenges

- 3D partial matching
- New ways of information representation including multimedia & things & context & device type & user’s intention.
- Extraction of information from the web on a particular person, place, or thing to auto create a wikipedia entry (e.g., extract and interpret all the graphics, audio, video on a topic such as crime, disease, art).
- Extraction (in real time) and interpretation of multimedia and multiparty communication (speech, posture, gesture).
- Automatic, content-based large scale multimedia indexing.
- Unification of multimedia content search combining real world information and social descriptions into one single descriptor (mixed multimedia search).

Research challenges

- Mobile multimedia content search using cameras, and other kind of available real world information (place, date info, etc).
- New algorithms for multimedia search in virtual worlds, MMORPGs, MM social networks.
- New recommender Systems (web, TV, mobile).
- New search services allowing for discovery of the best available algorithms and give it to all ("white box" approach).
- Creation of search benchmarking systems for the distributed media delivery, which would cover all critical aspects, including the user perceived quality.

Application fields

- Bioinformatics
- Car industry
- Furniture industry
- Medical imaging