GEOINFO
XII Brazilian Symposium on GeoInformatics
November 27-29 2011, Campos do Jordão, SP, Brazil

Computing Polygon Similarity from Raster Signatures

Léo C. R. Antunes[1] and Leonardo G. Azevedo[1,2]

{leo.antunes, azevedo}@uniriotec.br

[1]Graduate Program of Information Systems (PPGI)

Applied Informatics Department (DIA)

Federal University of State of Rio de Janeiro (UNIRIO)

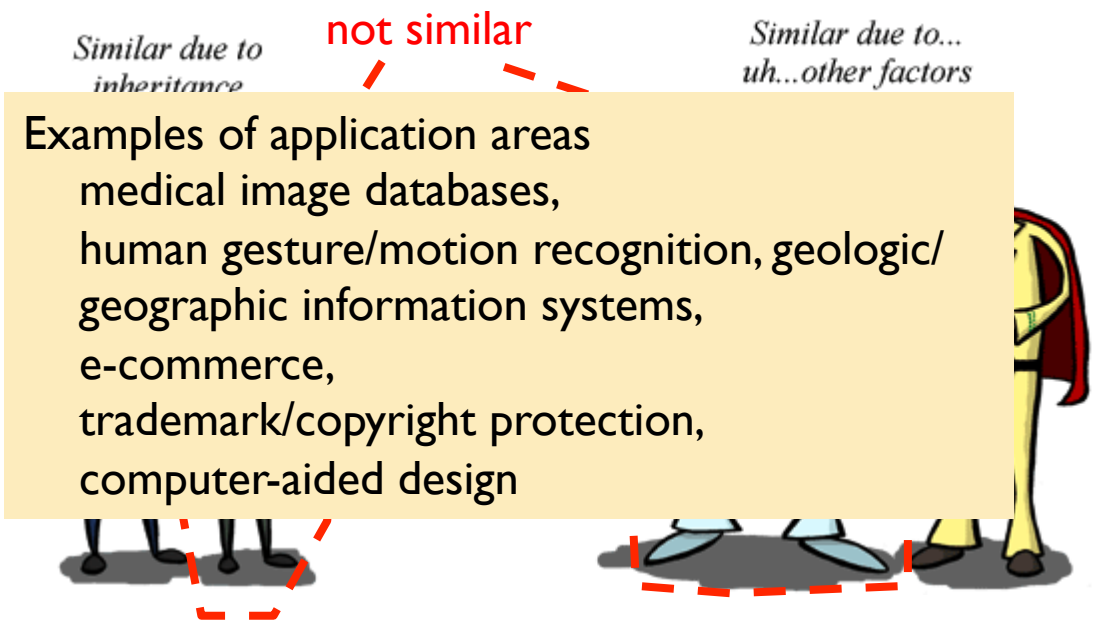
[2]Research and Practice Group in Information Technology (NP2Tec)

Overview

- Similarity definition
- Goal
- Main Concepts
- Raster Similarity and related algorithms
- Experimental tests
- Conclusion and future work

Similarity

- The similarity concept is fundamental for learning, knowledge and thought. [Quine,1969; Cakmakov e Celakoska, 2004]
- Similarity metric is a measure that allows comparison of pair of things.

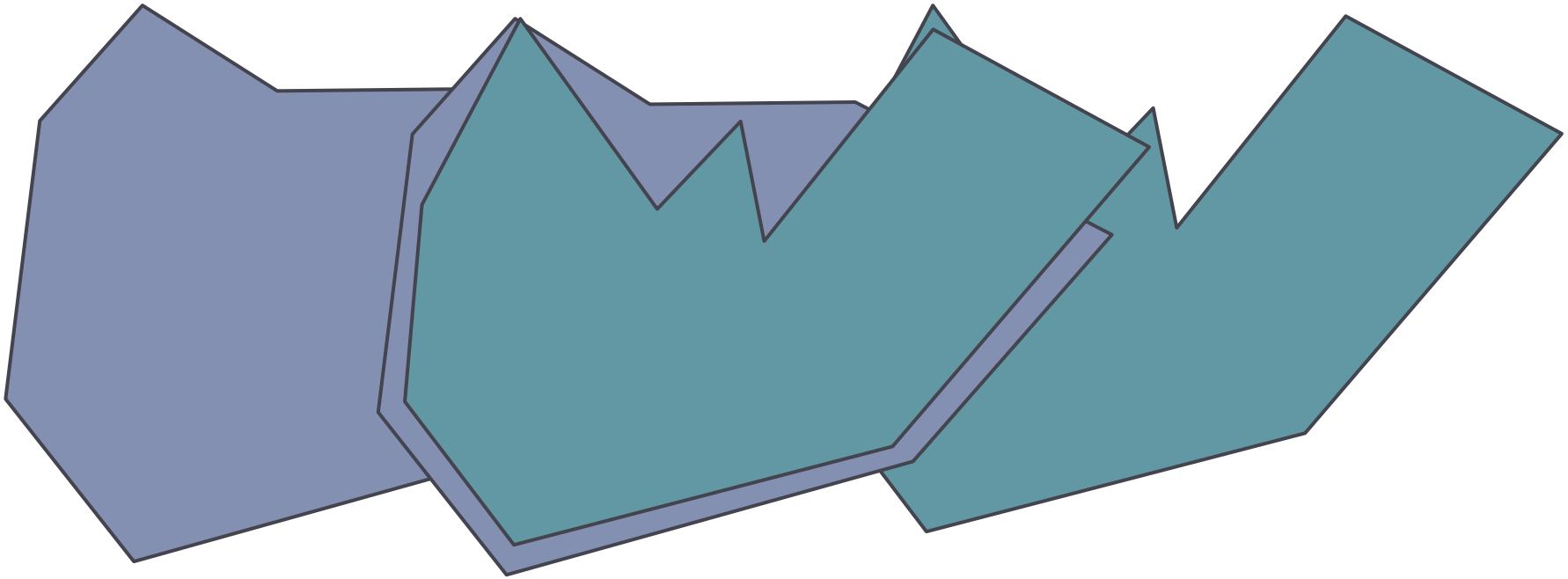


Examples of application areas
- medical image databases,
- human gesture/motion recognition, geologic/ geographic information systems,
- e-commerce,
- trademark/copyright protection,
- computer-aided design

Spatial Similarity

- Spatial similarity is as a subset of similarity
- Entities are compared considering their spatial components.

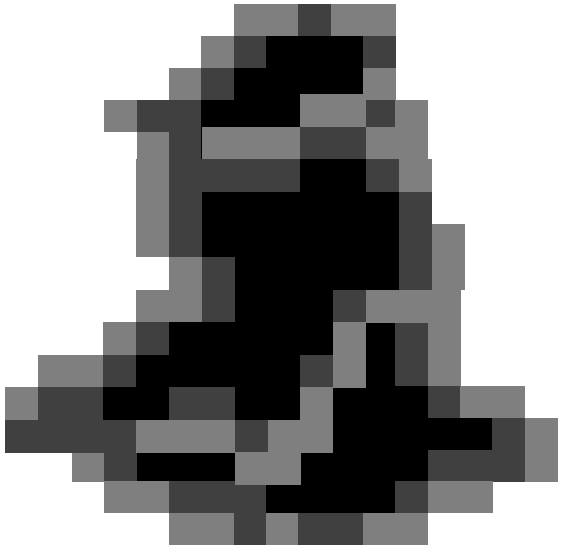
What is the similarity of these two spatial objects?



Goal of this work

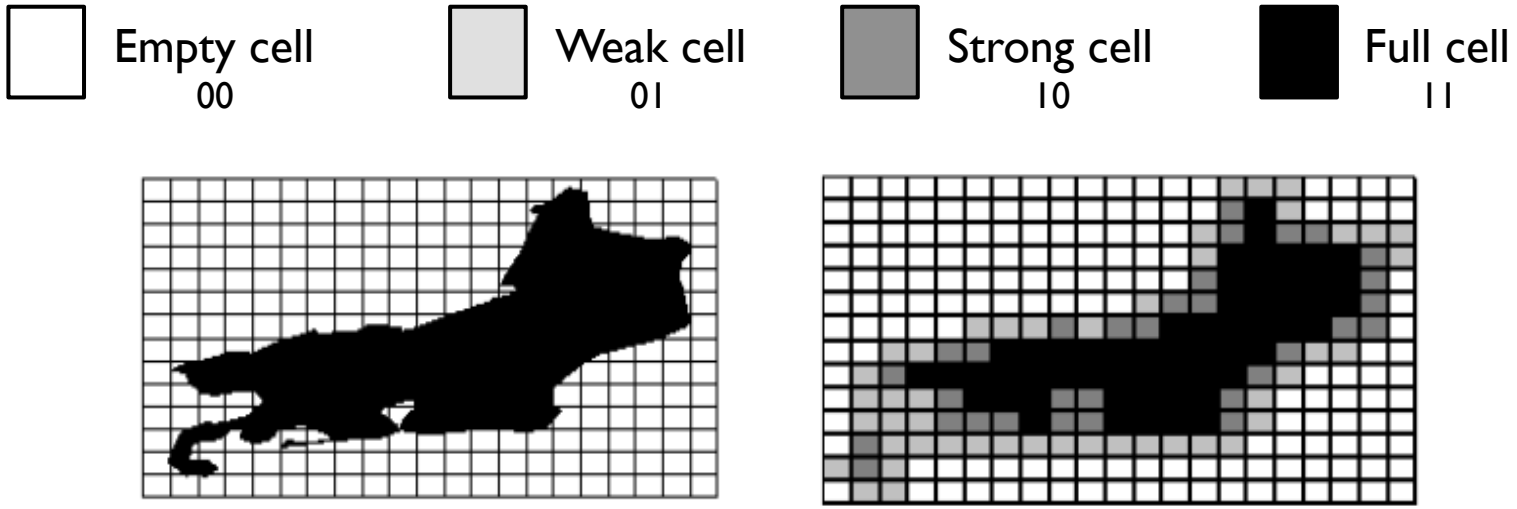
- Compute polygon similarity using polygon approximate representations

4RS



Four-Color Raster Signature (4CRS)

- Stores polygon main features in an approximate and compressed representation (Zimbrao and Souza, 1998)
- Grid of cells that stores relevant information using few bits
  - Fast access and processing



Raster Similarity Function

- Similarity function
  - Returns the similarity of objects considering size, shape, and position in space.
- Similarity of polygon function
  - Considers polygon area

$$S\,(o1,o2) = \frac{A_{\cap}\,(o1,o2)}{A_{\cup}\,(o1,o2)}$$, Jaccard index [Jaccard, 1912] apud Hemert e Baldock [2007]

- Proposal of this work
  - Replace spatial objects by their raster representations

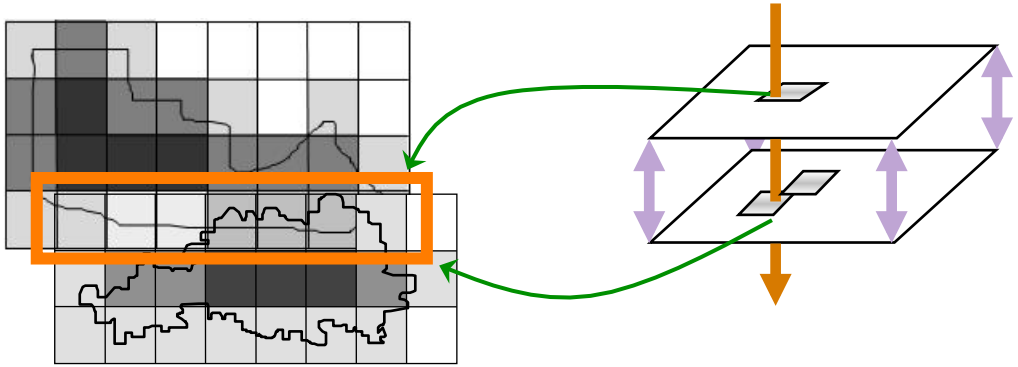
$$S\,(r1,r2) = \frac{A_{\cap}\,(r1,r2)}{A_{\cup}\,(r1,r2)}$$

Algorithm to Compute Raster Similarity

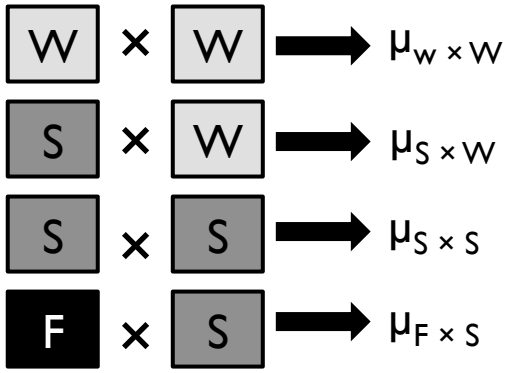
$$S\,(r1,r2) = \frac{A_{\cap}\,(r1,r2)}{A_{\cup}\,(r1,r2)}$$

1) Compute overlapping (intersection) area of polygons

Azevedo *et al.* (2005): Estimating the Overlapping Area of Polygon Join



Premises
-Signatures must have same cell size (Scale change)
-Approximate value for each cell types overlapping



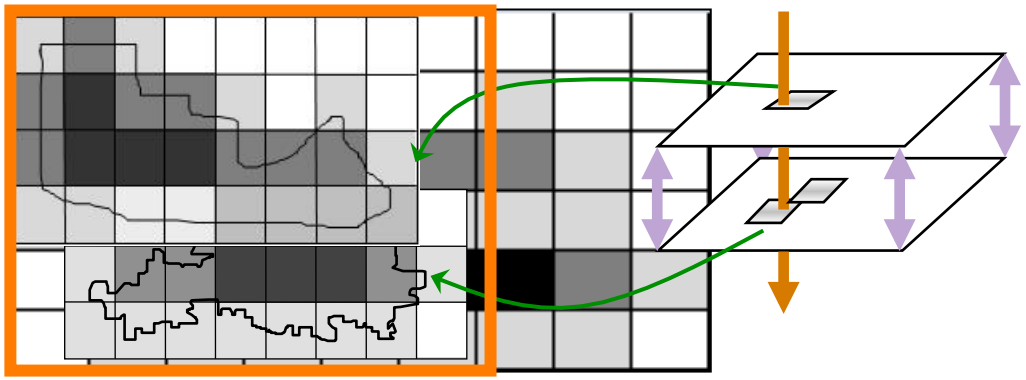
Approximate intersection area =
$(n_{W\times W} \times \mu_{W\times W} + n_{W\times S} \times \mu_{W\times S} + \ldots + n_{F\times F} \times \mu_{F\times F}) \times cellArea$

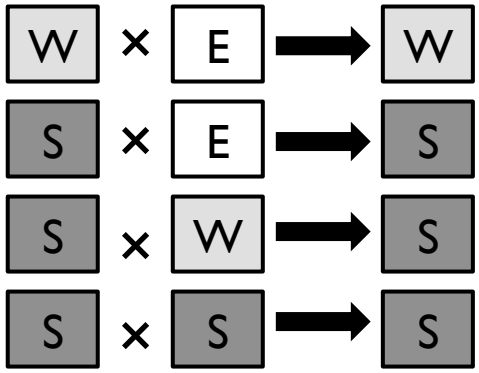
Algorithm to Compute Raster Similarity

$$S(r1,r2) = \frac{A_o(r1,r2)}{A_U(r1,r2)}$$

1) Compute overlapping (intersection) area of polygons
2) Compute area of union of raster signatures
    1) Compute raster signature that represents union of raster signatures
    2) Compute area of this raster signature

Other contribution of this work



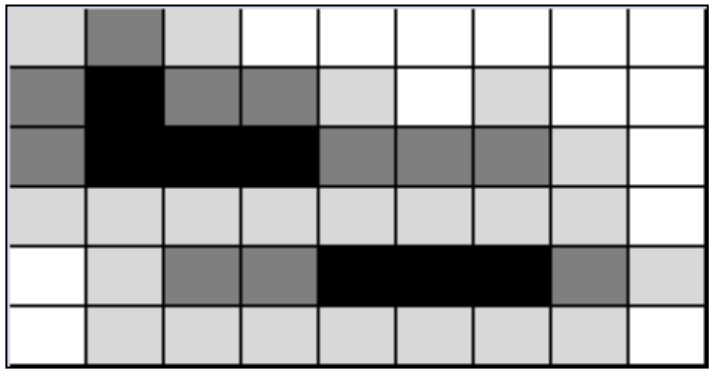


Algorithm to Compute Raster Similarity

$$S(r1,r2) = \frac{A_o(r1,r2)}{A_U(r1,r2)}$$

1) Compute overlapping (intersection) area of polygons
2) Compute area of union of raster signatures
    1) Compute raster signature that represents union of raster signatures
    2) Compute area of this raster signature

Azevedo *et al.* (2004): Approximate Spatial Query Processing Using Raster Signatures



Approximate value for each cell type

E	μ_E
W	μ_W
S	μ_S
F	μ_F

Approximate area =
$(n_E \times \mu_E + n_W \times \mu_W +$
$n_S \times \mu_S + n_F \times \mu_F) \times cellArea$

SECONDO

FernUniversität in Hagen

SECONDO

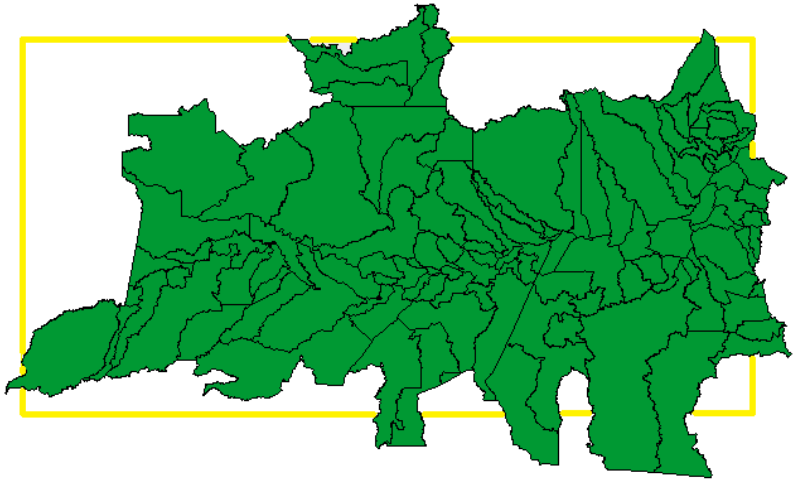
- Generic environment that supports database systems implementation for a large number of data models and query languages.
  - E.g.: Spatial Data
- Research prototype at Fernuniversität in Hagen (Germany) (http://dna.fernuni-hagen.de/Secondo.html/)

All implementations are available on googlecode:
http://code.google.com/p/raster4crs-project/

Experimental Tests

- Goal
  - Evaluate the precision and execution time
  - Similarity computed from Raster × from Polygons
- Spatial data
  - 382 polygons
    - Municipalities from north of Brazil (BRNorth)

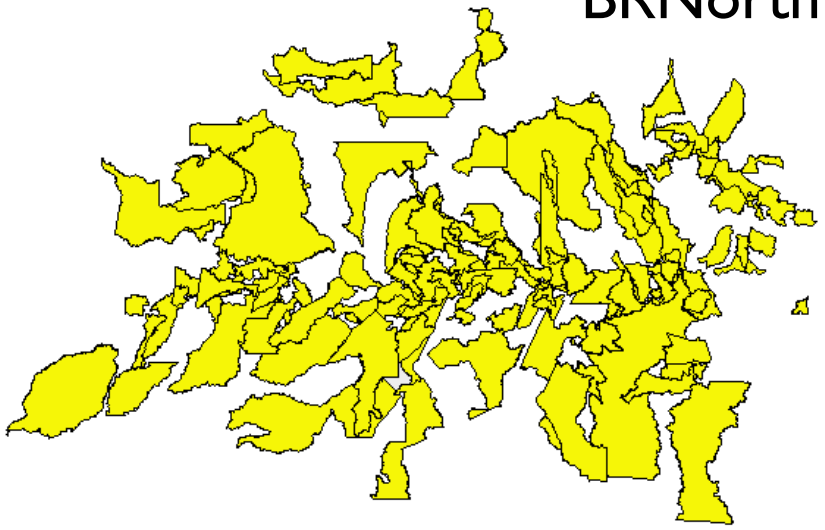
Experimental Tests



BRNorth

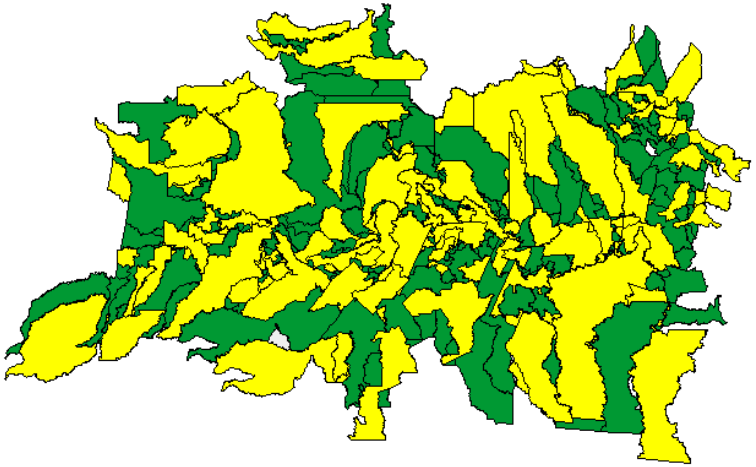
Random shifts
(Brinkhoff *et al.*, 1994)

BRNorthT



Experimental Tests – Execution time

- 585 intersections



Raster similarity is **3 times faster**

- Hot execution time
  - 10 executions of the same query, discarding the first, the highest and the slowest times to avoid outliers.
  - 41.187s to compute similarity from real polygons
  - 14.406s to compute similarity from raster signature

Experimental Tests - Precision

- Remarks
  - There were some outliers
  - Considered results between percentiles 20 and 80
    - Error median value was 10%
- Two datasets to analyze
  - Similarities with error below 10%
  - Similarities with error above 10%
- Characteristics considered in the tests
  - Intersection area
  - Area of union of raster signatures
  - Similarity value

Example of error above error median (10%)

Raster similarity: 15.63%
Polygon similarity: 7.48%
Error: 108.86%

Three main situations contribute to high error:

(i) Small number of overlapping cells

W × W	W × S	W × F	S × S	S × F	F × F	Total
2	7	6	6	5	0	26

(ii) Majority of overlaps involves cell types whose approximation of overlapping area consider the average (*Weak* × *Weak*, *Weak* × *Strong*, *Weak* × *Full*, *Strong* × *Full*, *Strong* × *Strong*)

(iii) Scale change for comparison

Example of error below error median (10%)

Raster similarity: 36.98%
Polygon similarity: 37.55%
Error: 1.51%

Three main situations contribute to small error:

(i) Big number of overlapping cells

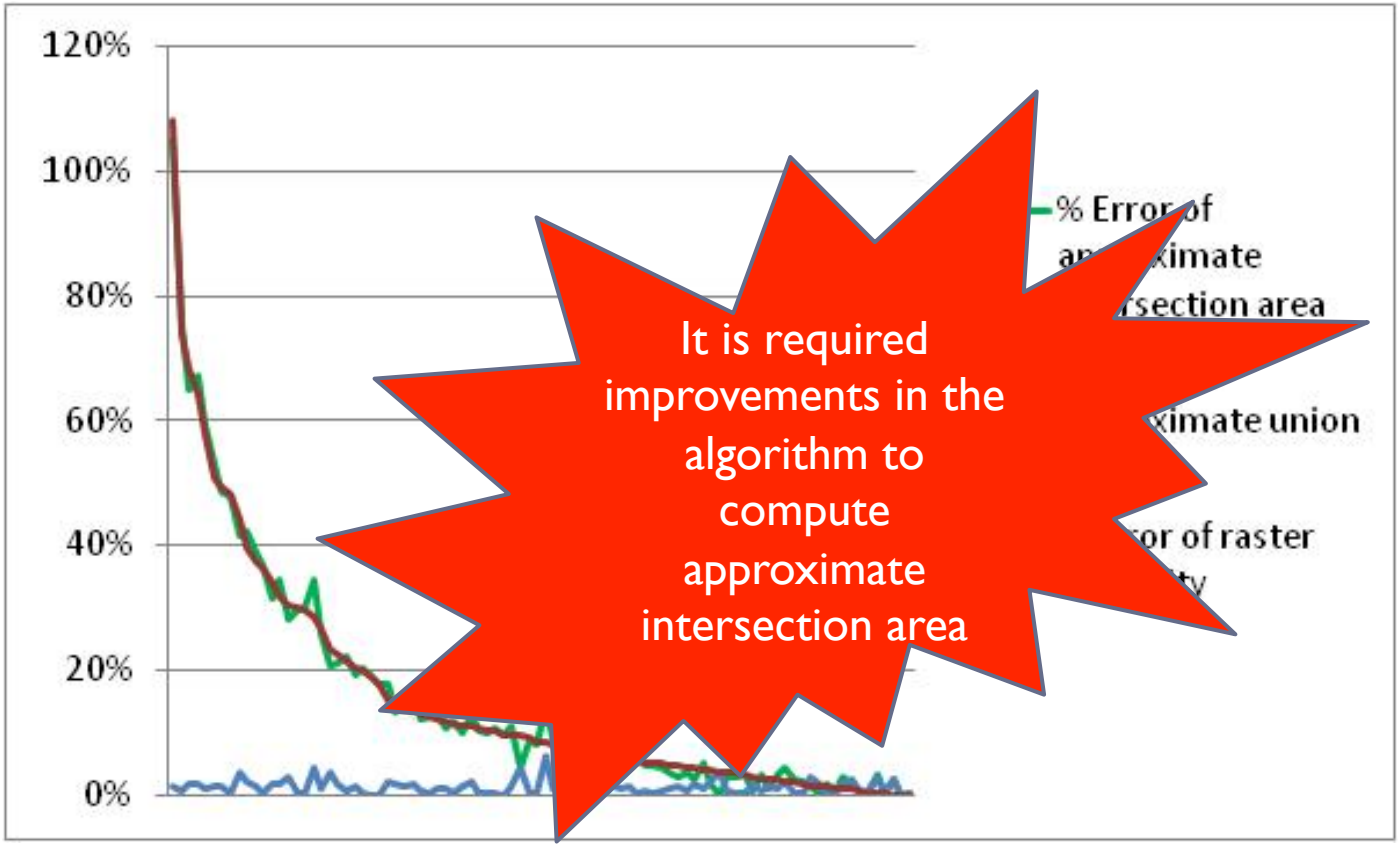
W × W	W × S	W × F	S × S	S × F	F × F	Total
5	5	18	5	16	29	90

(ii) Many overlappings of *Full* × *Full*

(iii) No scale change for comparison

Important remark

- Error of approximate intersection area is one that most contributes to error of raster similarity



Conclusion

- Similarity is a fundamental concept
- There are many applications for similarity function
- Contributions
  - Main contribution of this work
    - Algorithm to compute similarity of polygons from their 4CRS signatures.
  - Other contributions
    - Algorithm to compute union of two 4CRS signatures
    - Implementations of all algorithms in SECONDO
- Experimental tests
  - Executed on real datasets (municipalities from North Region of Brazil)
  - Evaluations considered execution time and precision
- Results
  - Raster similarity algorithm is three times faster than similarity using objects' real representation
  - Precision varies due to
    - Number of overlapping cells
    - Number of overlapping cells of type full (100% of precision)
    - Scale change

Future work

- Improve the algorithm to compute approximate intersection area
- Execute performance evaluations considering others datasets
- Evaluate the use of synthetic data to identify algorithm recommendations
- Improve the algorithm for other scenarios
  - E.g.: Compare objects according to their shape, independent from their size and without executing scale change
- Implement a view for Raster objects in SECONDO

GEOINFO
XII Brazilian Symposium on GeoInformatics
November 27-29 2011, Campos do Jordão, SP, Brazil

Computing Polygon Similarity from Raster Signatures

Léo C. R. Antunes[1] and Leonardo G. Azevedo[1,2]
{leo.antunes, azevedo}@uniriotec.br

[1]Graduate Program of Information Systems (PPGI)
Applied Informatics Department (DIA)
Federal University of State of Rio de Janeiro (UNIRIO)
[2]Research and Practice Group in Information Technology (NP2Tec)

Questions?

Thank you!!!